Raport Badawczy
Research Report

RB/59/2013

Midpoint evaluation for CMA-ES

K. Opara

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. dr hab. inż. Olgierd Hryniewicz

Warszawa 2013

Midpoint evaluation for CMA-ES

Karol Opara

January 2, 2014

Abstract

CMA-ES is one of the state-of-the art evolutionary algorithms. It consists of sampling from multivariate normal distribution, whose covariance matrix is claimed to approximate the inverse hessian of the objective function. The midpoint of this distribution should be therefore the best linear unbiased estimator of the optimum. This hypothesis was tested on the BBOB 2013 benchmark set using the standard CMA-ES implementation. Evaluation of the objective function in the midpoint neither improves nor deteriorates the performance of the algorithm. Moreover, it turns out that the standard implementation of CMA-ES is competitive but not as good as the best CMA-ES variants, which took parts in the BBOB 2009 competition.

1 Introduction

Covariance Matrix Adaptation Evolution Strategy (CMA-ES), introduced in [7] is one of the state-of-the-art evolutionary algorithms [5]. The method consists of iterative sampling from a multivariate normal distribution $\mathcal{N}(\boldsymbol{m}^{(g)}, \sigma^{(g)}\boldsymbol{C}^{(g)})$. Its parameters, mean $\boldsymbol{m}^{(g)}$, covariance matrix $\boldsymbol{C}^{(g)}$ and the scaling factor $\sigma^{(g)}$ are updated based on the values of the objective function to obtain a better-adapted multivariate normal distribution in iteration $g+1$.

The series of mean values in consecutive iterations $\boldsymbol{m}^{(g)}, \boldsymbol{m}^{(g+1)}, \boldsymbol{m}^{(g+2)}, \ldots$ is not directly used in the CMA-ES algorithm. The best of the sampled points is treated as an estimate of the optimum. The authors of CMA-ES claim there is "strong empirical evidence" that the covariance matrix in this algorithm approximates the inverse hessian [2]. In such case, for locally spherical functions location of the population mean would be the best linear unbiased estimator of the optimum (according to the Gauss-Markov theorem).

Therefore it might be beneficial to compute the value of the fitness function in the midpoint. Similar approach proved effective in a study of Differential Evolution by Arabas. He suggests that the midpoint should not be added to the population due to the risk of premature convergence but only used to update the estimate of the best point.

To verify the hypothesis of usefulness of midpoint evaluation for CMA-ES an experiment was performed. The standard implementation of CMA-ES [3] (version 3.62 beta, retrieved in December 2013) was compared with a variant, which was evaluating the midpoint in every tenth iteration ($p_e = 10$). Comparison was performed for computational budget of $D \cdot 10^4$ function evaluations, where D is the dimensionality of the search space. Evaluation was based on the BBOB 2013 benchmark.

2 Results

Results from experiments according to [4] on the benchmark functions given in [1, 6] are presented in Figures 3, 4, 5, 2, 2, and Tables 1 and 2. The **expected running time** (ERT), used in the figures and table, depends on a given target function value, $f_t = f_{opt} + \Delta f$, 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 [4, 8]. **Statistical significance** is tested with the rank-sum test for a given target Δf_t 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 if available.

3 Conclusions

Results presented in section 2 show that there is hardly any difference between the two investigated variants of the CMA-ES. This means, that evaluation of the midpoint neither improves nor deteriorates the performance of this algorithm. Hence, it may be skipped to avoid unnecessary complications without bringing any difference in performance.

The empirical runtime cumulative distribution function plot Fig. 3 presents the performance of the reference implementation of CMA-ES [3] (thick, red line) and all algorithms, which took part in the BBOB 2009 contest (thin beige lines). The higher the area under each curve the better the performance of an algorithm. The reference CMA-ES implementation is quite competitive but not as good as the best algorithms from 2009 contest, which also included variants of CMA-ES [5].

Acknowledgements Study was supported by research fellowship within "Information technologies: research and their interdisciplinary applications" agreement number POKL.04.01.01-00-051/10-00.

References

[1] S. Finck, N. Hansen, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions.

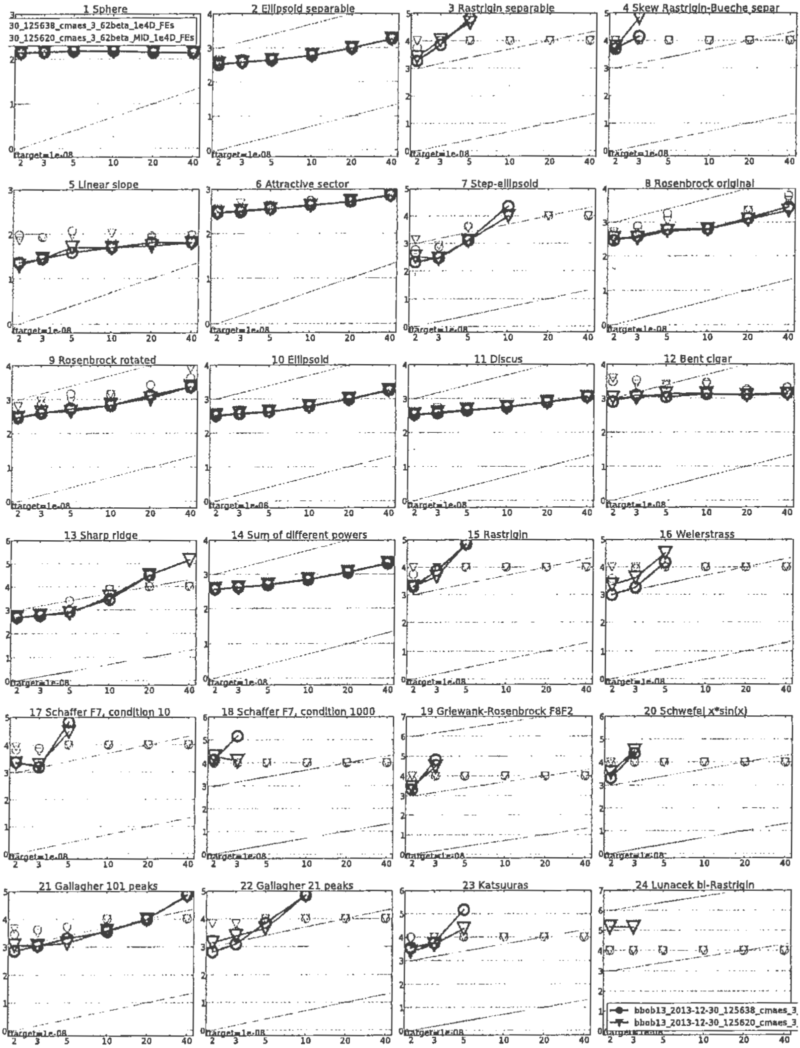
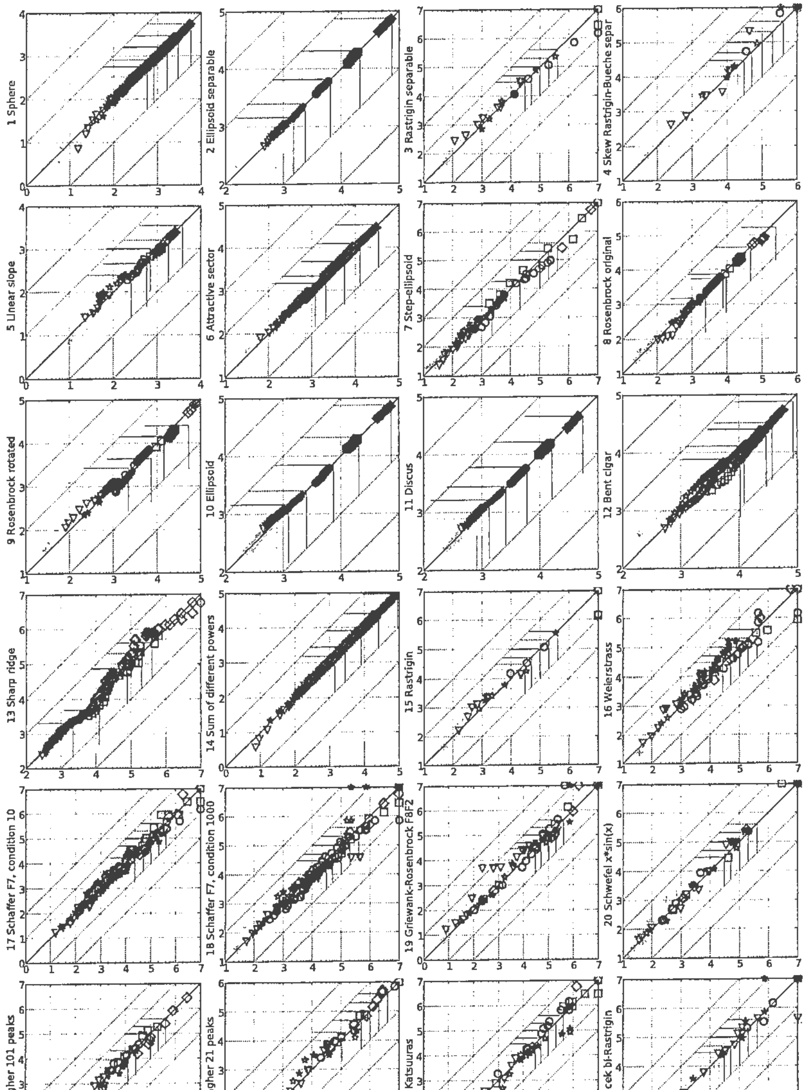


Figure 1: Expected running time (ERT in number of f-evaluations) divided by dimension for target function value 10^{-8} as $\log_{10}$ values versus dimension. Different symbols correspond to different algorithms given in the legend of f_1 and f_{24}. Light symbols give the maximum number of function evaluations from the longest trial divided by dimension. Horizontal lines give linear scaling, slanted dotted lines give quadratic scaling. Black stars indicate statistically better result compared to all other algorithms with $p < 0.01$ and Bonferroni correction number of dimensions (six). Legend: ∘:CMA-ES, ▽:CMA-ES MID.



Expected running time (ERT in $\log_{10}$ of number of function evaluations) of CMA-ES (x-axis) versus CMA-ES MID (y-axis) for 46 target values $\Delta f \in [10^{-8}, 10]$ in each dimension on functions f_1–f_{24}. Markers on the upper or right edge indicate that the target value was never reached. Markers represent dimension: 2:+, 3:$\triangledown$, 5:$\star$, 10:$\circ$, 20:$\square$, 40:$\diamond$.

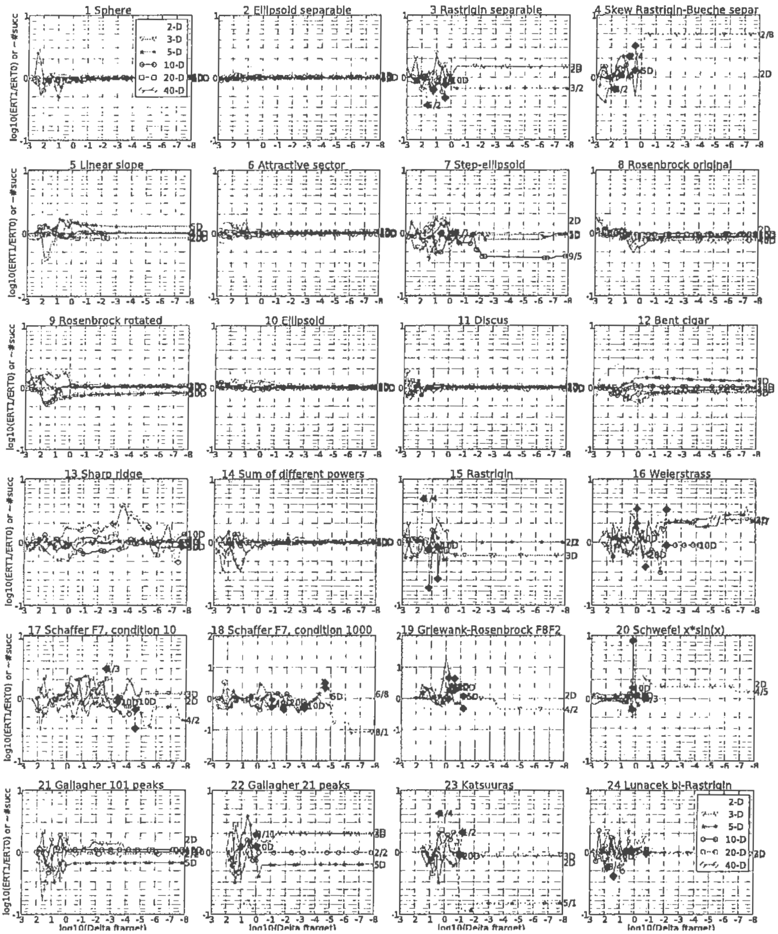


Figure 2: Ratio of ERT for CMA-ES MID over ERT for CMA-ES versus $\log_{10}(\Delta f)$ in 2:+, 3:$\triangledown$, 5:$\star$, 10:$\circ$, 20:$\square$, 40-D:$\Diamond$. Ratios $< 10^0$ indicate an advantage of CMA-ES MID, smaller values are always better. The line becomes dashed when for any algorithm the ERT exceeds thrice the median of the trial-wise overall number of f-evaluations for the same algorithm on this function. Filled symbols indicate the best achieved Δf-value of one algorithm (ERT is 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 CMA-ES MID. 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 CMA-ES MID (1^{st} number) and non-zero for CMA-ES (2^{nd} number). Results are significant with $p = 0.05$ for one star and $p = 10^{-\#\star}$ otherwise, with Bonferroni correction within each figure.

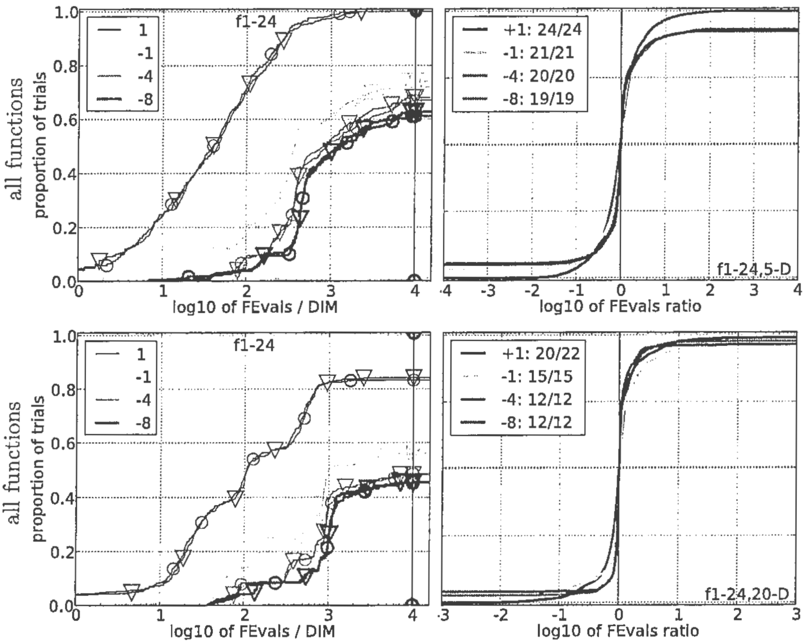


Figure 3: Noiseless functions 5-D (top) and 20-D (bottom). Left: Empirical Cumulative Distribution Function (ECDF) of the running time (number of function evaluations) for CMA-ES MID (∘) and CMA-ES (▽), divided by search space dimension D, to fall below $f_{\mathrm{opt}} + \Delta f$ with $\Delta f = 10^k$ where k is the value in the legend. The vertical black lines indicate the maximum number of function evaluations. Light beige lines in the background show ECDFs for target value 10^{-8} of all algorithms benchmarked during BBOB 2009. Right subplots: ECDF of ERT of CMA-ES MID over ERT of CMA-ES for different Δf.

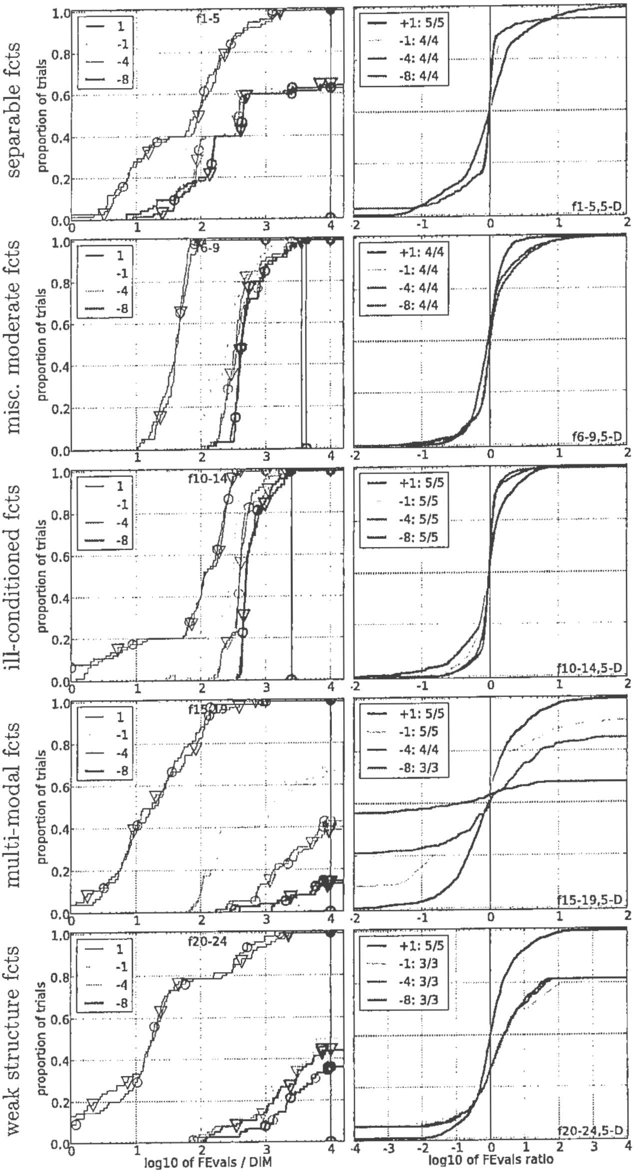


Figure 4: Subgroups of functions 5-D. See caption of Figure 3.

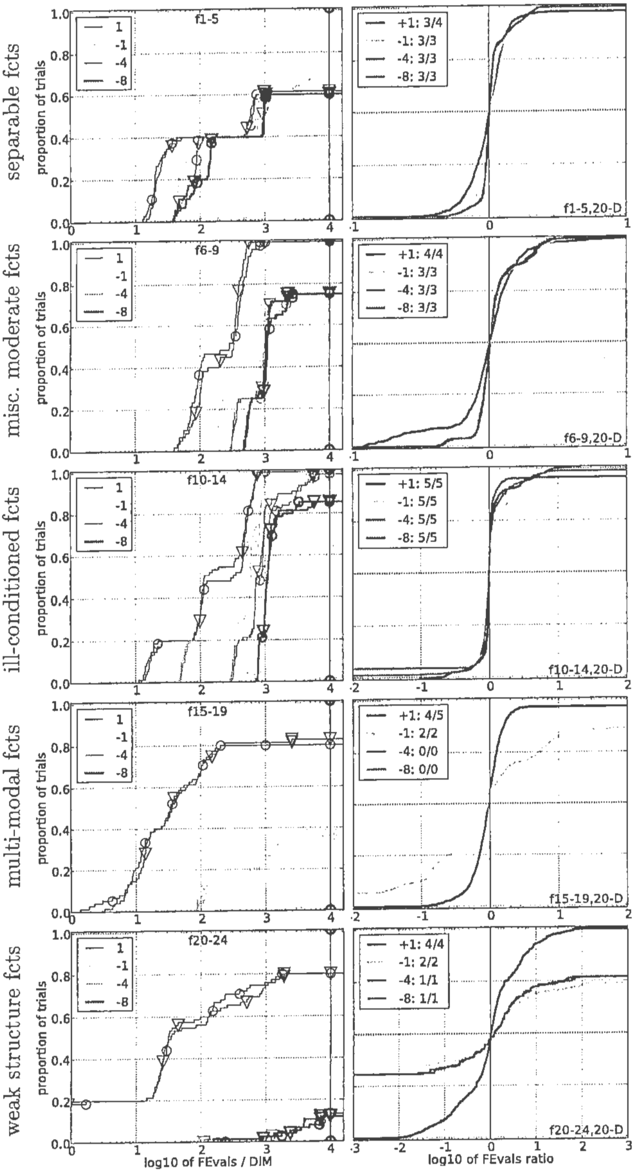


Figure 5: Subgroups of functions 20-D. See caption of Figure 3.

Table 1: ERT in number of function evaluations divided by the best ERT measured during BBOB-2009 given in the respective first row with the central 80% range divided by two in brackets for different Δf values. #succ is the number of trials that reached the final target $f_{\rm opt} + 10^{-8}$. 1:CMA-ES is CMA-ES and 2:CMA-ES MID is CMA-ES MID. Bold entries are statistically significantly better compared to the other algorithm, with $p = 0.05$ or $p = 10^{-k}$ where $k \in \{2, 3, 4, \dots\}$ is the number following the $\star$ symbol, with Bonferroni correction of 48. A $\downarrow$ indicates the same tested against the best BBOB-2009. Results for 5D.

Δf	1e+1	1e-1	1e-3	1e-5	1e-7	#succ
f_1	11	12	12	12	12	15/15
1: CMA-ES	3.7(3)	17(3)	29(4)	41(5)	55(6)	15/15
2: CMA-ES MID	3.0(3)	16(3)	29(4)	42(5)	55(6)	15/15
f_2	83	88	90	92	94	15/15
1: CMA-ES	14(5)	19(2)	20(2)	21(2)	22(2)	15/15
2: CMA-ES MID	14(4)	18(3)	20(2)	22(2)	23(2)	15/15
f_3	716	1637	1646	1650	1654	15/15
1: CMA-ES	1.4(2)	206(244)	205(228)	204(227)	204(242)	2/15
2: CMA-ES MID	1.0(1)	138(138)	137(152)	137(140)	137(151)	3/15
f_4	809	1688	1817	1886	1903	15/15
1: CMA-ES	2.4(4)	∞	∞	∞	∞ *5.0e4*	0/15
2: CMA-ES MID	3.6(4)	∞	∞	∞	∞ *5.0e4*	0/15
f_5	10	10	10	10	10	15/15
1: CMA-ES	4.8(4)	18(12)	19(18)	19(18)	19(18)	15/15
2: CMA-ES MID	6.0(6)	24(17)	25(17)	25(17)	25(17)	15/15
f_6	114	281	580	1038	1332	15/15
1: CMA-ES	1.8(0.9)	2.2(0.4)	1.6(0.2)	1.2(0.2)	1.2(0.2)	15/15
2: CMA-ES MID	2.1(1)	2.1(0.7)	1.7(0.4)	1.2(0.2)	1.2(0.2)	15/15
f_7	24	1171	1572	1572	1597	15/15
1: CMA-ES	5.4(3)	2.7(2)	3.8(4)	3.8(4)	4.0(4)	15/15
2: CMA-ES MID	5.6(3)	2.4(2)	3.0(4)	3.0(4)	3.7(4)	15/15
f_8	73	336	391	410	422	15/15
1: CMA-ES	3.6(0.9)	6.9(4)	6.8(4)	6.9(4)	7.1(4)	15/15
2: CMA-ES MID	4.0(1)	5.8(4)	5.9(3)	6.1(3)	6.3(3)	15/15
f_9	35	214	300	335	369	15/15
1: CMA-ES	6.3(2)	9.0(7)	7.6(5)	7.4(5)	7.1(4)	15/15
2: CMA-ES MID	6.5(1)	6.7(3)	6.0(2)	6.0(2)	5.8(2)	15/15
f_{10}	349	574	626	829	880	15/15
1: CMA-ES	3.2(1.0)	2.7(0.7)	2.8(0.5)	2.3(0.4)	2.4(0.3)	15/15
2: CMA-ES MID	3.2(0.9)	2.8(0.5)	2.9(0.2)	2.4(0.1)	2.4(0.1)	15/15
f_{11}	143	763	1177	1467	1673	15/15
1: CMA-ES	7.9(4)	2.2(0.3)	1.6(0.1)	1.4(0.1)	1.3(0.1)	15/15
2: CMA-ES MID	8.5(2)	2.3(0.3)	1.7(0.2)	1.4(0.2)	1.4(0.1)	15/15
f_{12}	108	371	461	1303	1494	15/15
1: CMA-ES	5.9(3)	6.4(7)	7.2(6)	3.4(3)	3.5(2)	15/15
2: CMA-ES MID	6.5(4)	9.3(6)	10(6)	4.5(3)	4.5(3)	15/15
f_{13}	132	250	1310	1752	2255	15/15
1: CMA-ES	4.0(2)	5.1(2)	1.6(0.5)	1.7(0.8)	1.8(1.0)	15/15
2: CMA-ES MID	3.7(3)	5.9(3)	1.6(0.4)	1.6(0.2)	1.5(0.2)	15/15
f_{14}	10	58	139	251	476	15/15
1: CMA-ES	2.0(2)	4.0(1)	4.8(1)	5.6(1)	4.5(0.4)	15/15
2: CMA-ES MID	2.3(3)	4.1(0.7)	4.9(0.8)	5.6(0.9)	4.6(0.5)	15/15
f_{15}	511	19369	20073	20769	21359	14/15
1: CMA-ES	1.5(0.4)	18(22)	17(20)	17(19)	16(17)	2/15
2: CMA-ES MID	1.8(2)	18(22)	17(19)	17(18)	16(18)	2/15
f_{16}	120	2662	10449	11644	12095	15/15
1: CMA-ES	1.7(1)	4.1(7)	2.7(4)	3.7(4)	3.6(4)	7/15
2: CMA-ES MID	2.1(2)	5.7(10)	6.2(7)	6.7(8)	10(11)	4/15
f_{17}	5.2	899	3669	6351	7934	15/15
1: CMA-ES	5.0(3)	0.74(0.2)	1.4(2)	12(14)	26(32)	2/15
2: CMA-ES MID	5.3(6)	1.6(2)	1.0(0.7)	6.8(7)	19(22)	4/15
f_{18}	103	3968	9280	10905	12469	15/15
1: CMA-ES	1.0(0.6)	0.64(1.0)	10(11)	66(76)	∞ *5.0e4*	0/15
2: CMA-ES MID	1.1(0.7)	1.1(1)	5.1(6)	∞	∞ *5.0e4*	0/15
f_{19}	1	242	1.2e5	1.2e5	1.2e5	15/15
1: CMA-ES	28(28)	868(1017)	∞	∞	∞ *5.0e4*	0/15
2: CMA-ES MID	38(26)	592(716)	∞	∞	∞ *5.0e4*	0/15
f_{20}	16	38111	54470	54861	55313	14/15
1: CMA-ES	4.0(2)	∞	∞	∞	∞ *5.0e4*	0/15
2: CMA-ES MID	4.7(3)	∞	∞	∞	∞ *5.0e4*	0/15
f_{21}	41	1674	1705	1729	1757	14/15
1: CMA-ES	7.8(19)	5.8(5)	5.7(5)	5.7(5)	5.6(5)	15/15
2: CMA-ES MID	6.6(1)	3.9(4)	3.8(4)	3.8(4)	3.8(4)	15/15
f_{22}	71	938	1008	1040	1068	14/15
1: CMA-ES	3.2(1)	37(42)	35(39)	34(38)	33(37)	11/15
2: CMA-ES MID	5.8(11)	23(27)	22(27)	21(24)	21(25)	13/15
f_{23}	3.0	14249	31654	33030	34256	15/15
1: CMA-ES	3.0(3)	25(26)	24(25)	23(24)	22(23)	1/15
2: CMA-ES MID	2.1(2)	5.0(6)	3.6(4)	3.4(4)	3.3(4)	5/15
f_{24}	1622	6.4e6	9.6e6	1.3e7	1.3e7	3/15
1: CMA-ES	1.6(1)	∞	∞	∞	∞ *5.0e4*	0/15
2: CMA-ES MID	2.3(3)	∞	∞	∞	∞ *5.0e4*	0/15

Table 2: ERT in number of function evaluations divided by the best ERT measured during BBOB-2009 given in the respective first row with the central 80% range divided by two in brackets for different Δf values. #succ is the number of trials that reached the final target $f_{\text{opt}} + 10^{-8}$. 1:CMA-ES is CMA-ES and 2:CMA-ES MID is CMA-ES MID. Bold entries are statistically significantly better compared to the other algorithm, with $p = 0.05$ or $p = 10^{-k}$ where $k \in \{2, 3, 4, \dots\}$ is the number following the $\star$ symbol, with Bonferroni correction of 48. A $\downarrow$ indicates the same tested against the best BBOB-2009. Results for 20D.

Δf	1e+1	1e-1	1e-3	1e-5	1e-7	#succ
f_1	43	43	43	43	43	15/15
1: CMA-ES	8.6(2)	22(2)	34(3)	46(3)	59(3)	15/15
2: CMA-ES MID	9.1(2)	21(2)	34(2)	47(3)	59(2)	15/15
f_2	385	387	390	391	393	15/15
1: CMA-ES	33(5)	43(7)	46(2)	48(2)	49(2)	15/15
2: CMA-ES MID	31(4)	43(4)	46(2)	48(2)	49(2)	15/15
f_3	5066	7635	7643	7646	7651	15/15
1: CMA-ES	∞	∞	∞	∞	∞ *2.0e5*	0/15
2: CMA-ES MID	556(632)	∞	∞	∞	∞ *2.0e5*	0/15
f_4	4722	7666	7700	7758	1.4e5	9/15
1: CMA-ES	∞	∞	∞	∞	∞ *2.0e5*	0/15
2: CMA-ES MID	∞	∞	∞	∞	∞ *2.0e5*	0/15
f_5	41	41	41	41	41	15/15
1: CMA-ES	13(5)	31(11)	32(12)	32(12)	32(12)	15/15
2: CMA-ES MID	14(6)	26(9)	27(9)	27(9)	27(9)	15/15
f_6	1296	3413	5220	6728	8409	15/15
1: CMA-ES	1.4(0.3)	1.1(0.1)	1.1(0.1)	1.1(0.1)	1.1(0.1)	15/15
2: CMA-ES MID	1.5(0.2)	1.1(0.1)	1.1(0.1)	1.2(0.1)	1.2(0.1)	15/15
f_7	1351	9503	16524	16524	16969	15/15
1: CMA-ES	1.3(1)	∞	∞	∞	∞ *2.0e5*	0/15
2: CMA-ES MID	2.3(2)	∞	∞	∞	∞ *2.0e5*	0/15
f_8	2039	4040	4219	4371	4484	15/15
1: CMA-ES	4.3(1.0)	5.7(3)	5.8(3)	5.8(3)	5.8(3)	15/15
2: CMA-ES MID	3.7(0.9)	5.1(3)	5.2(3)	5.2(3)	5.2(3)	15/15
f_9	1716	3277	3455	3594	3727	15/15
1: CMA-ES	5.6(3)	7.0(4)	7.1(4)	7.1(3)	7.0(3)	15/15
2: CMA-ES MID	4.7(1)	5.5(0.6)	5.6(0.6)	5.6(0.5)	5.6(0.5)	15/15
f_{10}	7413	10735	14920	17073	17476	15/15
1: CMA-ES	1.7(0.3)	1.6(0.1)	1.2(0.0)	1.1(0.0)	1.1(0.0)	15/15
2: CMA-ES MID	1.7(0.3)	1.6(0.2)	1.2(0.1)	1.1(0.1)	1.1(0.1)	15/15
f_{11}	1002	6278	9762	12285	14831	15/15
1: CMA-ES	9.4(0.8)	1.9(0.1)	1.4(0.0)	1.2(0.0)	1.0(0.0)	15/15
2: CMA-ES MID	9.5(0.8)	1.9(0.1)	1.4(0.1)	1.2(0.0)	1.0(0.0)	15/15
f_{12}	1042	2740	4140	12407	13827	15/15
1: CMA-ES	2.6(2)	3.8(3)	3.9(1)	1.7(0.5)	1.8(0.4)	15/15
2: CMA-ES MID	2.0(0.1)	3.0(2)	3.3(1)	1.5(0.6)	1.7(0.5)	15/15
f_{13}	652	2751	18749	24455	30201	15/15
1: CMA-ES	9.3(9)	7.5(5)	1.7(2)	4.2(3)	16(19)	4/15
2: CMA-ES MID	5.3(4)	6.5(10)	2.2(3)	4.2(5)	12(14)	4/15
f_{14}	75	304	932	1648	15661	15/15
1: CMA-ES	4.3(0.8)	3.6(0.5)	4.0(0.5)	6.1(0.5)	1.2(0.1)	15/15
2: CMA-ES MID	4.8(1)	3.7(0.5)	4.2(0.6)	6.2(0.5)	1.2(0.1)	15/15
f_{15}	30378	3.1e5	3.2e5	4.5e5	4.6e5	15/15
1: CMA-ES	∞	∞	∞	∞	∞ *2.0e5*	0/15
2: CMA-ES MID	44(51)	∞	∞	∞	∞ *2.0e5*	0/15
f_{16}	1384	77015	1.9e5	2.0e5	2.2e5	15/15
1: CMA-ES	1.6(0.7)	∞	∞	∞	∞ *2.0e5*	0/15
2: CMA-ES MID	1.7(0.7)	∞	∞	∞	∞ *2.0e5*	0/15
f_{17}	63	4005	30677	56288	80472	15/15
1: CMA-ES	2.1(1)	1.5(2)	16(17)	∞	∞ *2.0e5*	0/15
2: CMA-ES MID	2.6(2)	1.5(2)	29(30)	∞	∞ *2.0e5*	0/15
f_{18}	621	19561	67569	1.3e5	1.5e5	15/15
1: CMA-ES	1.1(0.3)	6.8(10)	∞	∞	∞ *2.0e5*	0/15
2: CMA-ES MID	1.2(0.4)	5.0(7)	∞	∞	∞ *2.0e5*	0/15
f_{19}	1	3.4e5	6.2e6	6.7e6	6.7e6	15/15
1: CMA-ES	259(78)	∞	∞	∞	∞ *2.0e5*	0/15
2: CMA-ES MID	279(48)	∞	∞	∞	∞ *2.0e5*	0/15
f_{20}	82	3.1e6	5.5e6	5.6e6	5.6e6	14/15
1: CMA-ES	6.1(1)	∞	∞	∞	∞ *2.0e5*	0/15
2: CMA-ES MID	5.7(1)	∞	∞	∞	∞ *2.0e5*	0/15
f_{21}	561	14103	14643	15567	17589	15/15
1: CMA-ES	9.1(10)	13(14)	13(14)	12(13)	11(12)	10/15
2: CMA-ES MID	13(21)	13(16)	13(16)	12(14)	11(13)	9/15
f_{22}	467	23491	24948	26847	1.3e5	12/15
1: CMA-ES	16(30)	∞	∞	∞	∞ *2.0e5*	0/15
2: CMA-ES MID	19(23)	∞	∞	∞	∞ *2.0e5*	0/15
f_{23}	3.2	67457	4.9e5	8.1e5	8.4e5	15/15
1: CMA-ES	2.4(2)	43(47)	∞	∞	∞ *2.0e5*	0/15
2: CMA-ES MID	1.8(2)	42(46)	∞	∞	∞ *2.0e5*	0/15
f_{24}	1.3e6	5.2e7	5.2e7	5.2e7	5.2e7	3/15
1: CMA-ES	∞	∞	∞	∞	∞ *2.0e5*	0/15
2: CMA-ES MID	∞	∞	∞	∞	∞ *2.0e5*	0/15

Technical Report 2009/20, Research Center PPE, 2009. Updated February 2010.

[2] N. Hansen. The cma evolution strategy: a comparing review. In *Towards a new evolutionary computation*, pages 75–102. Springer, 2006.

[3] N. Hansen. CMA-ES source code, 2013.

[4] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2010: Experimental setup. Technical Report RR-7215, INRIA, 2010.

[5] N. Hansen, A. Auger, R. Ros, S. Finck, and P. Pošík. Comparing results of 31 algorithms from the black-box optimization benchmarking bbob-2009. In *Proceedings of the 12th annual conference companion on Genetic and evolutionary computation*, pages 1689–1696. ACM, 2010.

[6] 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.

[7] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2):159–195, 2001.

[8] 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.