

# Benchmarking CMA-ES with Basic Integer Handling on a Mixed-Integer Test Problem Suite

Tristan Marty

Thales Research and Technology,  
Inria and Institut Polytechnique

Yann Semet

Thales Research and Technology

Anne Auger

Inria and Institut Polytechnique

Sébastien Héron

Thales Research and Technology

Nikolaus Hansen

Inria and Institut Polytechnique

## ABSTRACT

We compare the performances of one implementation of CMA-ES (pycma version 3.3.0) for optimizing functions with both continuous and integer variables. The implementation incorporates a lower bound on the variance along the integer coordinates to keep the optimization from stalling. This benchmark will serve as a baseline for further works on pycma. Results show substantial improvement since the last benchmarked version of pycma. Also this implementation is competitive with other mixed integer algorithms.

## CCS CONCEPTS

•Computing methodologies → Continuous space search;

## KEYWORDS

Benchmarking, Black-box optimization, Mixed-integer optimization

### ACM Reference format:

Tristan Marty, Yann Semet, Anne Auger, Sébastien Héron, and Nikolaus Hansen. 2023. Benchmarking CMA-ES with Basic Integer Handling on a Mixed-Integer Test Problem Suite . In *Proceedings of The Genetic and Evolutionary Computation Conference 2023, Lisbon, Portugal, July 15–19, 2023 (GECCO '23)*, 8 pages.

DOI: 10.1145/3583133.3596411

## 1 INTRODUCTION

Mixed optimization problems are of great importance in real-world applications. Typically, variables include real-valued degrees of freedom that correspond, directly or not, to physical measurements (e.g. angles, frequencies, forces, etc.) as well as discrete (integer or binary) settings that correspond to categorical choices (e.g. materials) or service activations (e.g. radar modes). Therefore, developing algorithmic variants that can handle several types of optimization variable at the same time without impairing the search process will be very useful in practice.

In the context of black box optimization for problems with continuous variables, evolutionary algorithms which update a covariance matrix, namely CMA-ES, have been studied for more than 20 years

Permission to make digital or hard copies of part or all 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. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

GECCO '23, Lisbon, Portugal

© 2023 Copyright held by the owner/author(s). 979-8-4007-0120-7/23/07...\$15.00  
DOI: 10.1145/3583133.3596411

[10]. While the CMA-ES was originally not designed to handle integer variables, it can be employed with integer variables by using the integer part of a continuous variable in the fitness evaluation. This is however expected to lead to integer variables getting stuck in the resulting fitness plateaus, and several versions have been proposed to improve the algorithm in this situation [4, 5].

In order to study mixed integer problems within the COCO framework [9], the bboB-mixint testbed was proposed in [13]. This testbed is based on the 24 functions of the bboB testbed with 80% of the variables to be integer with an arity between 2 and 16.

In order to prevent variables getting stuck on the same integer level, for example in CMA-ES, a lower bound on the variance can be set such that there remains a reasonable probability for an integer variable to “jump” to another integer level, as for example in CMA-ES with margin (CMA-ESwM) [4]. The same principle is implemented in the pycma code with Integer Handling (cma-IH in the benchmark) with a more simplistic approach.

The purpose of this paper is to measure the performance of the current implementation of the widely distributed cma-IH code (version 3.3.0) [6] with different options and compare its performance with previous results and enhancements.

## 2 ALGORITHM PRESENTATION

The general idea behind cma-IH is to set a lower bound for the variance of the integer variables. It prevents integer variables from being trapped in one value. We formulate the bounding using the covariance matrix decomposition from the CMA-ES with diagonal decoding [1]. At each new iteration  $t + 1$ , new candidate solutions,  $x^{(t+1)}$ , are sampled according to a multivariate normal distribution.

$$\begin{aligned} x^{(t+1)} &\sim \mathcal{N}(m, \sigma^{(t)^2} \mathbf{C}^{(t)}) \\ &\sim m + \sigma^{(t)} \sqrt{\mathbf{C}^{(t)}} \mathcal{N}(0, \mathbf{I}) , \end{aligned}$$

where  $\mathbf{C}$  is a covariance matrix and  $\sigma$  is the step size. In order to bound the standard deviation of each integer variable, we introduce the diagonal scaling matrix  $\mathbf{D}$ ,

$$x^{(t+1)} \sim m + \sigma^{(t)} \mathbf{D}^{(t)} \sqrt{\mathbf{C}^{(t)}} \mathcal{N}(0, \mathbf{I}) ,$$

where  $\mathbf{D}^{(0)}$  is the identity. The update of  $\sigma$  and  $\mathbf{C}$  are done as in the original CMA-ES algorithm. Then, the diagonal matrix  $\mathbf{D}$  is updated based on the current standard deviation in each coordinate. The standard deviation of the  $i^{\text{th}}$ -th coordinate of  $x$ ,  $\sigma_{\text{std}(i)}^{(t)}$ , is computed with the diagonal elements of  $\mathbf{C}$  and  $\mathbf{D}$ :

---

```

import cocoex
import cma
suite = cocoex.Suite ('bbob-mixint', '', '')
for problem in suite :
    up_bound = list(problem.upper_bounds[
        :problem.number_of_integer_variables]) +
        [None]
    lo_bound = list(problem.lower_bounds[
        :problem.number_of_integer_variables]) +
        [None]
    cma.fmin2(problem , problem.initial_solution, 1,
        {'bounds': [lo_bound, up_bound]
        'CMA_std': (problem.upper_bounds
                    - problem.lower_bounds)/5,
        'integer_variables':list(range(
            problem.number_of_integer_variables)),
        'restarts' :9})

```

---

**Figure 1:** Calling code to benchmark `cma.fmin2`.

$$\sigma_{\text{std}(i)}^{(t)} = \sigma^{(t)} D_{i,i}^{(t)} \sqrt{C_{i,i}^{(t)}}$$

This standard deviation has to be compared with a lower bound  $\sigma_{LB}$ . If the standard deviation in the  $i^{\text{th}}$  coordinate is smaller than the threshold value,  $\sigma_{\text{std}(i)}^{(t)} < \sigma_{LB}$ , then the **D** matrix is scaled up in this coordinate.

$$D_{i,i}^{(t+1)} \leftarrow \begin{cases} \frac{\sigma_{LB}}{\sigma^{(t)} \sqrt{C_{i,i}^{(t)}}} & \text{if } 1 \leq i \leq N_{\text{int}} \text{ and } \sigma_{\text{std}(i)}^{(t)} \leq \sigma_{LB} \\ D_{i,i}^{(t)} & \text{otherwise} \end{cases} \quad (1)$$

where  $N_{\text{int}}$  is the number of integer variables. In the version 3.3.0 of pycma with integer handling (`cma-IH`), the lower bound is set to:

$$\sigma_{LB} = \frac{1}{2N_{\text{int}} + 1} \quad (2)$$

### 3 EXPERIMENTAL PROCEDURE

Figure 1 shows the calling code and options used to run the benchmark for `cma-IH`. A maximum of nine restarts may be performed and each restart doubles the population size starting from the default (IPOP-CMA-ES) [2]. The initial step size is chosen to sample solutions mainly within the upper and the lower bound of the problem. The bounds are enforced on the integer variables only. The maximum number of function evaluations allowed per problem is set to  $10^4$  times the number of dimension of the problem.

This current implementation `cma-IH` is compared to performance data from the bbo-mixint COCO archive<sup>1</sup>. We show results of the algorithms given in Table 1.

The experiments were performed and the plots were produced with COCO [9], version 2.6.3.

### 4 CPU TIMING

In order to evaluate the CPU timing of the algorithm, we have run the `cma-IH` algorithm with restarts on the bbo-mixint test

<sup>1</sup>See <https://numbbo.github.io/data-archive>

**Table 1: Compared algorithms**


---

cma-IH	pycma version 3.3 (current) with the default integer handling option
CMA-ES-pycma	previous integer handling of pycma benchmarked in [13]
CMA-ESwM	CMA-ES with margin benchmarked in [3]
cma-NoIH	pycma version 3.3 (current) without any integer handling option
cmaIH1e-1	as <code>cma-IH</code> with a larger coordinate-wise lower bound on the sample standard deviation of $\sigma_{LB} = 10^{-1}$
DE-scipy	Differential Evolution implemented in scipy benchmarked in [13]
RANDOMSEARCH	Random Search algorithm for baseline

---

suite [13] with a maximum budget of  $10^4 \times$  dimensions function evaluations according to [11].

The Python code was run on a Linux machine with 32 cores, Intel® Xeon® E7 to T3 v4 processor. The time per function evaluation averaged over all functions in the test bed for dimensions 5, 10, 20, 40, 80 and 160 equals 3.7, 4.9, 5.1, 5.6, 6.4 and  $8.4 \times 10^{-4}$  seconds respectively.

## 5 RESULTS

We show results from experiments according to [11] and [8] on the benchmark functions given in [13]. Figures 2 and 3 show summary results for the different function groups. Figure 4 shows the empirical distribution of function evaluations to reach 51 target values on the single functions [9]. Tables 2 and 3 present expected runtimes and statistical tests.<sup>2</sup>

The **expected runtime (ERT)**, used in the tables, depends on a given target precision,  $I_{\text{target}} = f_{\text{opt}} + \Delta f$ , 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  $f_t$  [11, 12]. **Statistical significance** is tested with the rank-sum test for a given target  $\Delta f$  using, for each trial, either the number of needed function evaluations to reach  $f_t$  (inverted and multiplied by  $-1$ ), or, if the target was not reached, the best  $\Delta f$ -value achieved, measured only up to the smallest number of overall function evaluations for any unsuccessful trial under consideration.

*Comparing two integer handling versions of pycma.* We compare the performance of `cma-IH` (the current default implementation) versus CMA-ES-pycma (an older implementation). In comparison, the `cma-IH` performs increasingly better with increasing dimension and is overall about ten times faster than CMA-ES-pycma in dimension 80 (see Figure 3, lower right). The largest differences we observe on the ill-conditioned and multi-modal functions. The difference on the linear slope function is mainly due to the boundary handling which was applied also to the continuous variables in the CMA-ES-pycma experiment, as we confirmed in additional

<sup>2</sup>The entire data can be found at <https://trmarty.github.io/MixedIntegerData/>

experiments (not shown). The cma-IH performs statistically significantly better on 6, 8, 10, 16 and 15 functions in dimension 5, 10, 20, 40 and 80, respectively. The CMA-ES-pycma performs better on the separable ellipsoid function f2 in dimension 20, 40 and 80.

*Comparing pycma with and without the integer handling.* The cma-IH and cma-NoIH perform very similar overall. The only statistically significant difference we observed is on the separable ellipsoid function f2 in dimension 5 (not shown), where cma-IH is about three times faster than cma-NoIH, probably due to preventing additional restarts before to reach better targets. The default lower bound on the sample standard deviations for integer variables has overall little impact on the performance on this test suite.

*A larger lower bound for the standard deviations.* In order to understand the impact of the choice of the standard deviation lower bound value, we have run experiments with a few different bound values and show the results for the lower bound set to  $10^{-1}$ . While in the smaller dimensions the effects are minor, in larger dimension the cmaIH1e-1 performs overall best and is three times faster than cma-IH on the ill-conditioned functions in larger dimension for budgets above 100 times dimension. On the 80-dimensional *sphere* function, cmaIH1e-1 solves the last target ten times quicker than cma-IH and also two times quicker than CMA-ESwM. The cmaIH1e-1 also solves the *Ellipsoid separable* whereas cma-IH gets stuck at around 20% of targets. The cmaIH1e-1 is slower than CMA-ESwM by a factor of approximately 3 on functions 5, 13 and 14, however it is significantly better than the current version cma-IH on functions 2, 5 or 12 while not being worse on all other functions. For this range of dimensions, bounding the standard deviation to  $10^{-1}$  is a simple yet effective solution.

## 6 CONCLUSION

We have assessed the integer handling implemented in the current pycma module [7] (version 3.3.0). The performance with integer handling is overall superior to a previously benchmarked version, however it behaves quite similar to the version without integer handling. We attribute the latter observation to a suboptimal parameter setting: tests with a lower bound for the sample standard

deviations of 1/10, which is considerably larger than the original setting, reveal a remarkably improved performance on a number of functions. In particular in larger dimension, this variant performs overall best.

## REFERENCES

- [1] Y. Akimoto and N. Hansen. 2020. Diagonal Acceleration for Covariance Matrix Adaptation Evolution Strategies. *Evolutionary Computation* 28, 3 (09 2020), 405–435. DOI:[http://dx.doi.org/10.1162/evco\\_a\\_00260](http://dx.doi.org/10.1162/evco_a_00260)
- [2] A. Auger and N. Hansen. 2005. A restart CMA evolution strategy with increasing population size. In *2005 IEEE Congress on Evolutionary Computation*, Vol. 2. 1769–1776 Vol. 2. DOI:<http://dx.doi.org/10.1109/CEC.2005.1554902>
- [3] Ryoki Hamano, Shota Saito, Masahiro Nomura, and Shinichi Shirakawa. 2022. Benchmarking CMA-ES with Margin on the Bbob-Mixint Testbed. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '22)*. Association for Computing Machinery, New York, NY, USA, 1708–1716. DOI:<http://dx.doi.org/10.1145/3520304.3534043>
- [4] Ryoki Hamano, Shota Saito, Masahiro Nomura, and Shinichi Shirakawa. 2022. CMA-ES with margin: lower-bounding marginal probability for mixed-integer black-box optimization. *Proceedings of the Genetic and Evolutionary Computation Conference* (2022).
- [5] Nikolaus Hansen. 2011. *A CMA-ES for Mixed-Integer Nonlinear Optimization*. Research Report RR-7751. INRIA. <https://hal.inria.fr/inria-00629689>
- [6] Nikolaus Hansen. 2022. CMA-ES/pycma 3.3.0. <https://github.com/CMA-ES/pycma/tree/development>. (2022). commit e762ac0.
- [7] Nikolaus Hansen, Youhei Akimoto, and Petr Baudis. 2019. CMA-ES/pycma on Github. Zenodo, DOI:<https://doi.org/10.5281/zenodo.2559634>. (Feb. 2019). DOI:<http://dx.doi.org/10.5281/zenodo.2559634>
- [8] N. Hansen, A. Auger, D. Brockhoff, D. Tušar, and T. Tušar. 2016. COCO: Performance Assessment. *ArXiv e-prints* arXiv:1605.03560 (2016).
- [9] N. Hansen, A. Auger, R. Ros, O. Mersmann, T. Tušar, and D. Brockhoff. 2021. COCO: A Platform for Comparing Continuous Optimizers in a Black-Box Setting. *Optimization Methods and Software* 36 (2021), 114–144. Issue 1. DOI:<http://dx.doi.org/10.1080/10556788.2020.1808977>
- [10] N. Hansen and A. Ostermeier. 1996. Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation. In *Proceedings of IEEE International Conference on Evolutionary Computation*. 312–317. DOI:<http://dx.doi.org/10.1109/ICEC.1996.542381>
- [11] N. Hansen, T. Tušar, O. Mersmann, A. Auger, and D. Brockhoff. 2016. COCO: The Experimental Procedure. *ArXiv e-prints* arXiv:1603.08776 (2016).
- [12] Kenneth Price. 1997. Differential evolution vs. the functions of the second ICEO. In *Proceedings of the IEEE International Congress on Evolutionary Computation*. IEEE, Piscataway, NJ, USA, 153–157. DOI:<http://dx.doi.org/10.1109/ICEC.1997.592287>
- [13] Tea Tušar, Dimo Brockhoff, and Nikolaus Hansen. 2019. Mixed-Integer Benchmark Problems for Single- and Bi-Objective Optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '19)*. Association for Computing Machinery, New York, NY, USA, 718–726. DOI:<http://dx.doi.org/10.1145/3321707.3321868>

Table 2: Expected runtime (ERT) to reach given targets, measured in number of  $f$ -evaluations, in dimension 40. For each function, the ERT and, in braces as dispersion measure, the half difference between 10 and 90%-tile of (bootstrapped) runtimes is shown for the different target  $\Delta f$ -values as shown in the top row. #succ is the number of trials that reached the last target  $f_{\text{opt}} + 10^{-8}$ . The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries, succeeded by a star, are statistically significantly better (according to the rank-sum test) when compared to all other algorithms of the table, with  $p = 0.05$  or  $p = 10^{-k}$  when the number  $k$  following the star is larger than 1, with Bonferroni correction by the number of functions (24). Best results are printed in bold. Data produced with COCO v2.6.3

Table 3: Expected runtime (ERT) to reach given targets, measured in number of  $f$ -evaluations, in dimension 80. For each function, the ERT and, in braces as dispersion measure, the half difference between 10 and 90%-tile of (bootstrapped) runtimes is shown for the different target  $\Delta f$ -values as shown in the top row. #succ is the number of trials that reached the last target  $f_{\text{opt}} + 10^{-8}$ . The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries, succeeded by a star, are statistically significantly better (according to the rank-sum test) when compared to all other algorithms of the table, with  $p = 0.05$  or  $p = 10^{-k}$  when the number  $k$  following the star is larger than 1, with Bonferroni correction by the number of functions (24). Best results are printed in **bold**. Data produced with COCO v2.6.3

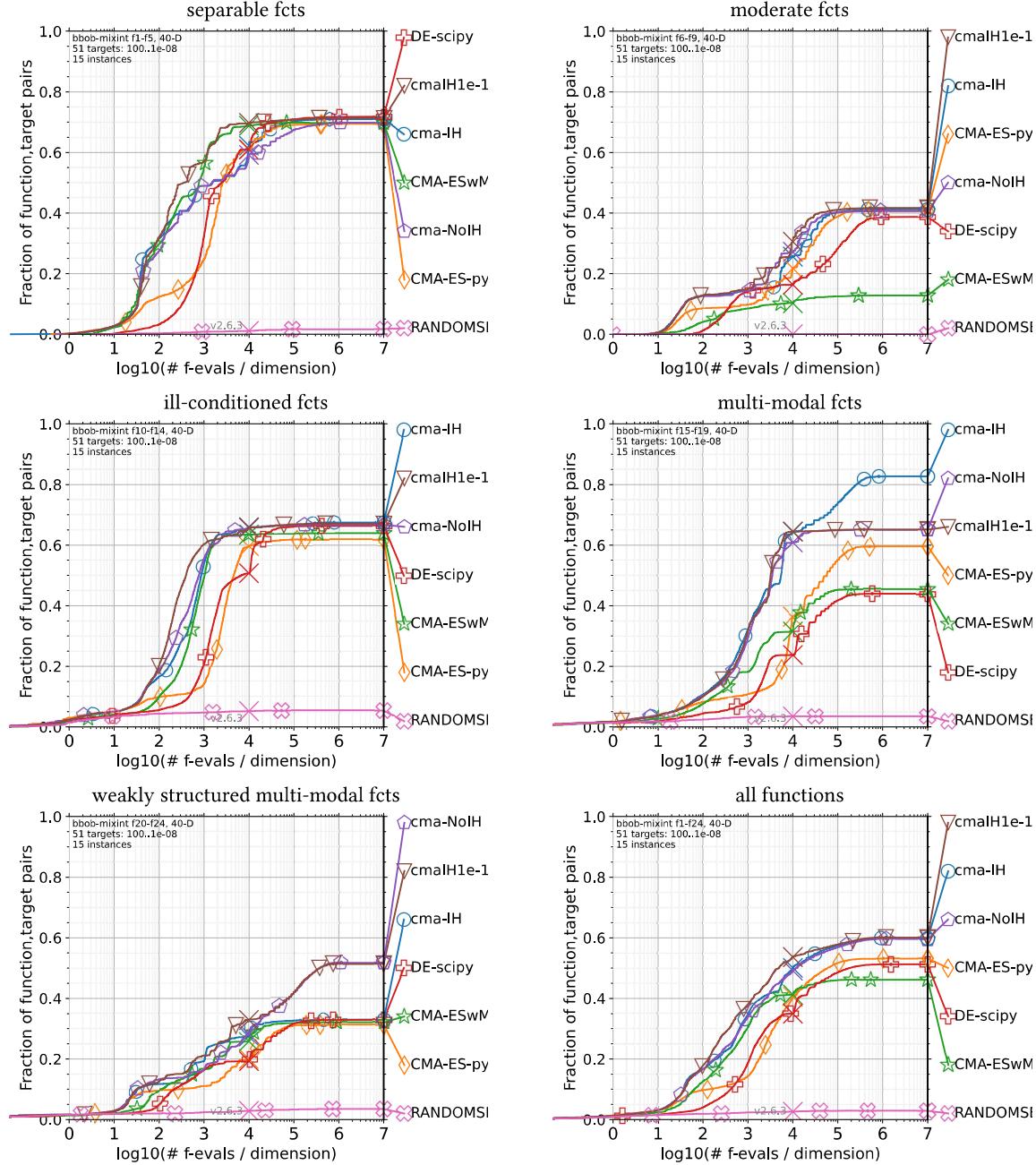
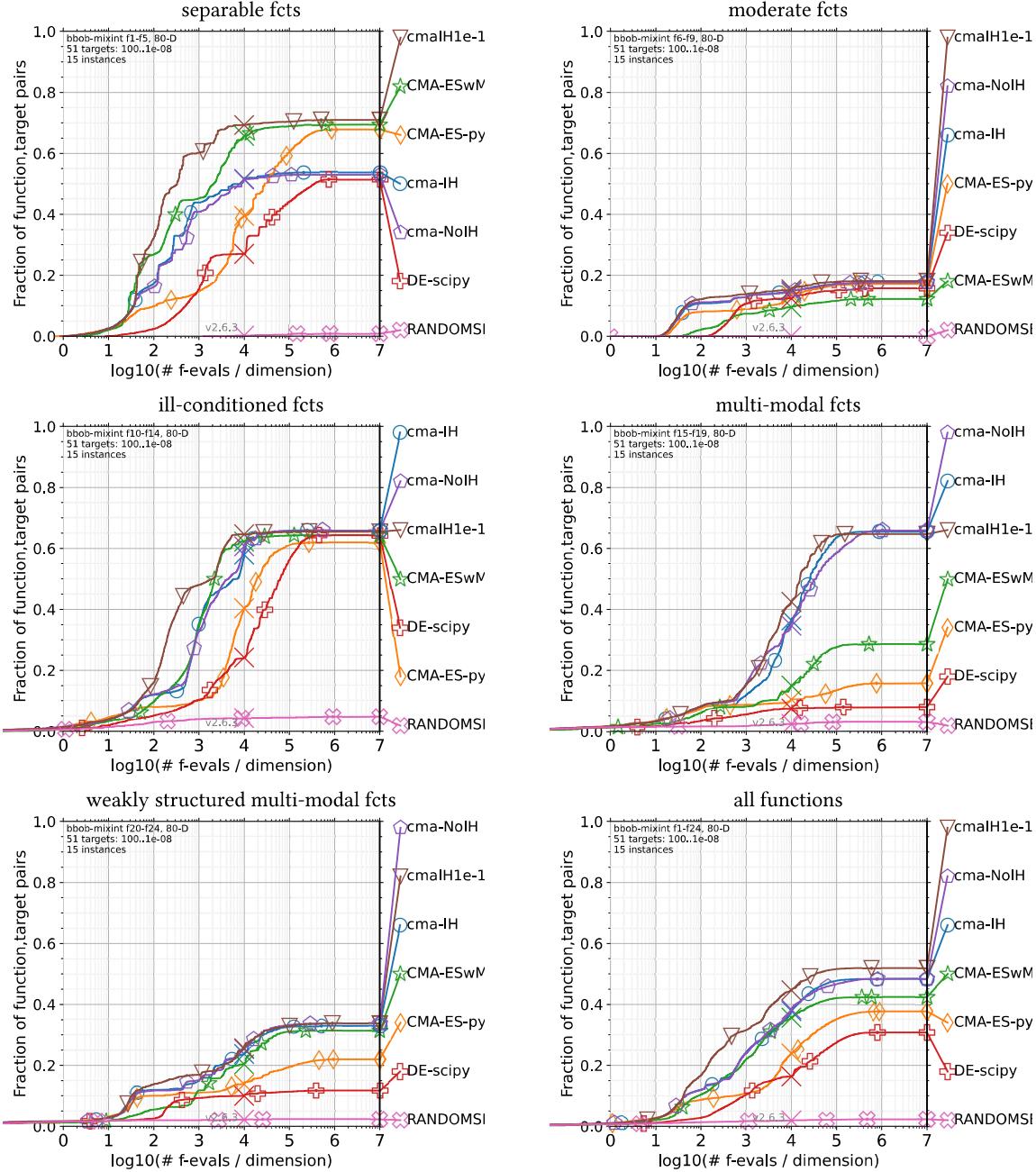
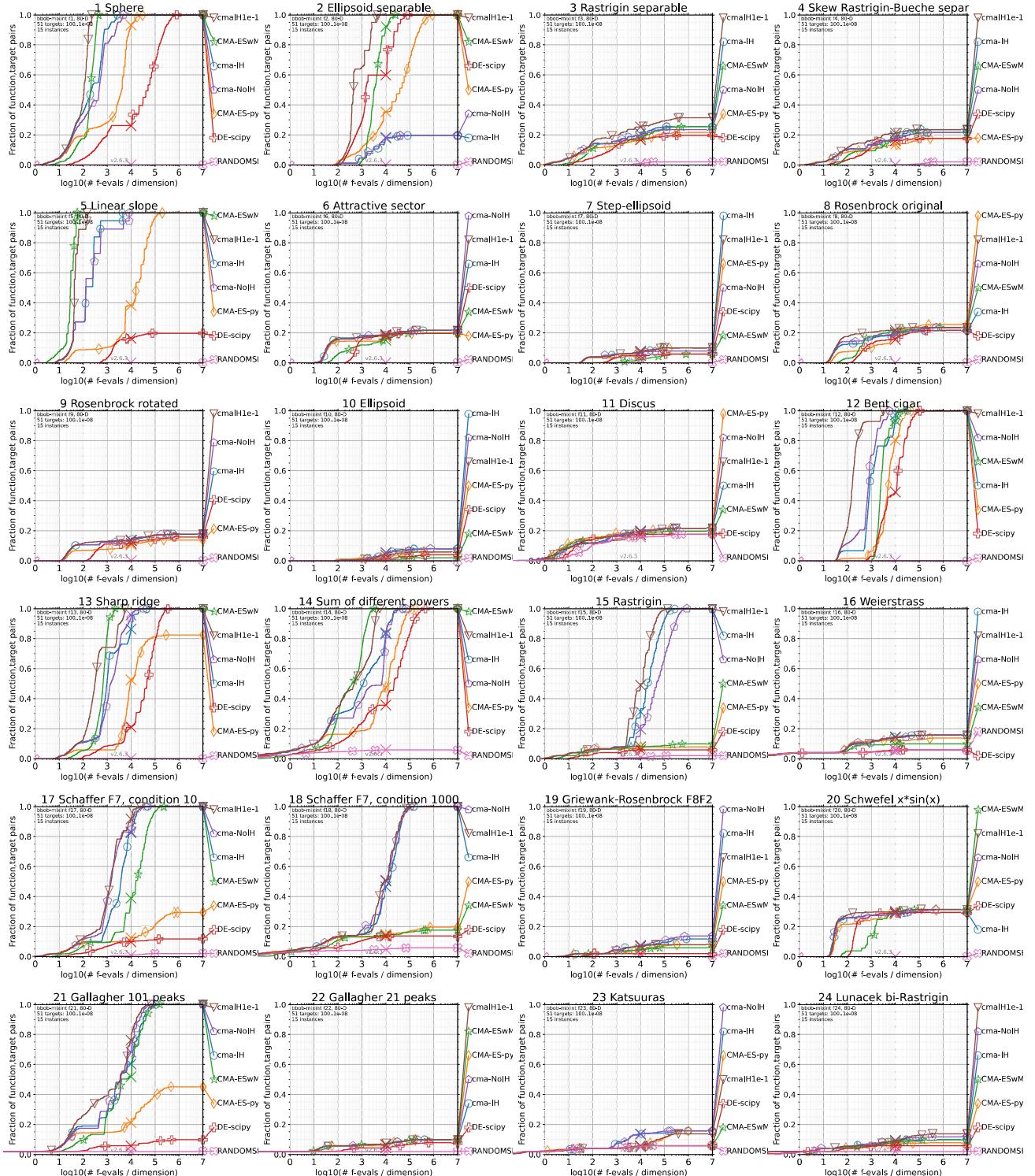


Figure 2: Bootstrapped empirical cumulative distribution of the number of  $f$ -evaluations divided by dimension (FEvals/DIM) for 51 targets with target precision in  $10^{[-8..2]}$  for all functions and subgroups in 40-D.



**Figure 3: Bootstrapped empirical cumulative distribution of the number of  $f$ -evaluations divided by dimension (FEEvals/DIM) for 51 targets with target precision in  $10^{-8..2}$  for all functions and subgroups in 80-D.**



**Figure 4:** Empirical cumulative distribution of simulated (bootstrapped) runtimes, measured in number of  $f$ -evaluations, divided by dimension (FEvals/DIM) for the 51 targets  $10^{[-8..2]}$  in dimension 80.