

# Benchmarking the Novel CMA-ES Restart Strategy Using the Search History on the BBOB Noiseless Testbed

Takahiro Yamaguchi

Graduate School of Science and Technology  
Shinshu University  
4-17-1 Wakasato  
Nagano 380-8553, Japan  
17w2097h@shinshu-u.ac.jp

Youhei Akimoto

Faculty of Engineering  
Shinshu University  
4-17-1 Wakasato  
Nagano 380-8553, Japan  
y\_akimoto@shinshu-u.ac.jp

## ABSTRACT

In this paper we propose a termination mechanism and the initial step-size control mechanism for restart strategies in the CMA-ES. The proposed mechanism utilizes a history of the distribution parameters from past restarts to early terminate an overlapping exploitation of the search domain. The initial step-size is controlled so that the next restart will not overlap with past restarts. The proposed mechanism is combined with a simple restart, IPOP restart and BIPOP restart strategies. The effectiveness and the drawback of the proposed mechanism is demonstrated on the BBOB noiseless testbed.

## CCS CONCEPTS

• **Mathematics of computing** → **Bio-inspired optimization**;

## KEYWORDS

Benchmarking, black-box optimization, termination criterion, restart strategy, CMA-ES

### ACM Reference format:

Takahiro Yamaguchi and Youhei Akimoto. 2017. Benchmarking the Novel CMA-ES Restart Strategy Using the Search History on the BBOB Noiseless Testbed. In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15-19, 2017*, 8 pages.  
<https://doi.org/http://dx.doi.org/10.1145/3067695.3084203>

## 1 INTRODUCTION

Restart strategies are almost necessities for the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) to find the global optimum on multimodal black-box functions. The IPOP restart strategy [4] that doubles the population size every restart improves the performance on well-structured (big-valley structured) multimodal functions, whereas a simple random restart [3] sometimes performs better on weakly-structured multimodal functions and deceptive functions. The BIPOP restart strategy [6] that is composed of the IPOP regime and the local search regime achieves a

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*GECCO '17 Companion, July 15-19, 2017, Berlin, Germany*

© 2017 Association for Computing Machinery.

ACM ISBN 978-1-4503-4939-0/17/07...\$15.00

<https://doi.org/http://dx.doi.org/10.1145/3067695.3084203>

nearly best performance for medium to large budgets of function evaluations in BBOB 2009. The IPOP scheme is effective on well-structure multimodal function, while the local search scheme with a relatively small population size and a small step-size helps on weakly-structured functions.

In this paper we introduce a search history to efficiently terminate each restart and to control the initial step-size of the search distribution. The distribution parameters are recorded, and are compared to the parameters in the history to check whether the current search distribution is so close that the search area is overlapped with past restarts. The initial step-size is controlled on the basis of the history. The algorithm is combined with the simple random restart, the IPOP restart, and the BIPOP restart where the initial step-size is controlled by the proposed mechanism. The benchmarking reveals a goodness of the proposed mechanism on e.g.  $f_{22}$ , where the number of local minima is a relatively small constant over dimensions, and a weakness on e.g.  $f_{13}$ ,  $f_{17}$  and  $f_{18}$ , where undesired termination of restart is observed.

## 2 ALGORITHM DESCRIPTION

The base line algorithm for this paper is the so-called CSA-CMA-ES, the covariance matrix adaptation evolution strategy with the cumulative step-size adaptation. We apply restart strategies such as a random restart, a restart with incremental population size (IPOP), and a restart with two regimes that combines the IPOP scheme and a restart with a small initial step-size.

### 2.1 Basic Idea

The proposed algorithm consists of two components, a novel termination criterion and a mechanism to control the initial standard deviation. We record the distribution parameters such as the mean vector  $m$  and the covariance matrix  $\sigma^2C$  of the Gaussian sampling distribution  $\mathcal{N}(m, \sigma^2C)$ . If the current search distribution is close enough to a distribution stored in the history of past restarts, the current restart is terminated as the current search is overlapping with a past search. If successive restarts have stopped due to the newly introduced termination criterion with a similar  $\sigma^2C$ , we regard that the initial step-size is too large to escape from a big valley and decrease next initial standard deviation. We expect it will be useful when optimizing a weakly-structured or deceptive functions. Note that the above idea is for weakly-structured multimodal functions such as the double-sphere function described in Section 2.6. If the objective function is globally well structured but highly rugged such as the Rastrigin function, a larger step-size

helps to find a better local minimum. On such functions restarts are barely terminated with our novel stopping mechanism since they often search different local minima. Then, the initial step-size will not be decreased.

To solve multimodal functions, an incremental population size scheme such as IPOP or BIPOP is often required. It is known from [1, 2] that the optimal standard deviation of the distribution is proportional to  $\mu_w = 1/\sum_{i=1}^{\lambda} w_i^2$  if  $\mu_w$  is sufficiently smaller than the dimension  $N$ , where  $\lambda$  is the population size and  $w_i$  is the recombination weight, and is roughly proportional to  $\lambda$  in case of the default  $w_i$  or the truncation weight. Provided that a nearly optimal standard deviation of the distribution is adapted, the current distribution will not be sufficiently close to the distributions in the history that are recorded from restarts with different population sizes, even if their search areas are overlapped. To adapt our termination and restart scheme, we introduce the following normalization of the covariance matrix of the sampling distribution

$$\Sigma = \frac{\sigma^2}{\alpha^2} \mathbf{C} \quad \text{where} \quad \alpha = \frac{\mu_w}{N-1+\mu_w}. \quad (1)$$

This normalization reflects the fact that the optimal standard deviation is proportional to  $\mu_w$  when  $\mu_w \ll N$  and the empirical knowledge that the optimal value tends to level out when  $\mu_w \gg N$ . We record the normalized parameters and compare the normalized parameters to check the termination condition.

The pseudo-code of the proposed termination mechanism using the search history combined with the restart strategy controlling the initial step-size is provided in Algorithm 1. The single run of the CMA-ES with the proposed termination mechanism is written in lines 11–27. The proposed termination mechanism is in lines 15–26. The next step-size is determined in line 30–36.

## 2.2 History of Normalized Parameters

We predefine multiple targets for the standard deviation of the normalized sampling distribution  $\mathcal{N}(m, \Sigma)$ . In this paper, we mean by the term *normalized standard deviation* (normalized std.) the geometric average  $\frac{1}{2N} \ln \det(\Sigma)$  of the square roots of the eigenvalues of the normalized covariance matrix. Given the initial population size  $\lambda^{(0)}$  and the initial normalized std., the target for the normalized std. is defined as

$$T_{\sigma} = \text{initial normalized std.} \times [1, 10^{-1}, \dots, 10^{1-n_{\sigma}^{\text{target}}}] .$$

When the normalized std. crosses a target  $T_{\sigma}[j]$  from above, we record the current normalized parameters to the history  $T_{\mathcal{N}}[j]$ . To prevent the history from growing too much, we record only the last normalized parameters for each entry for each restart. Therefore, the history size will grow by at most one for each entry for each restart.

## 2.3 Termination Based on KL-divergence

We measure the similarity between two normalized sampling distributions  $\mathcal{N}(m_0, \Sigma_0)$  and  $\mathcal{N}(m_1, \Sigma_1)$  in terms of the KL-divergence

$$D_{\text{KL}}(\mathcal{N}_0 \parallel \mathcal{N}_1) = \frac{1}{2} \{ (m_1 - m_0)^T \Sigma_1^{-1} (m_1 - m_0) + \text{Tr}(\Sigma_0^{-1} \Sigma_1) - N + \ln \det(\Sigma_0^{-1} \Sigma_1) \} . \quad (2)$$

---

### Algorithm 1 Proposed Restart Scheme

---

**Require:**  $N \geq 1, n^{\text{restart}} \geq 1, n_{\sigma}^{\text{target}} \geq 1, n_{\text{KL}}^{\text{check}} \geq 1, n_{\sigma}^{\text{dec}} \geq 1, \sigma^{(0)} > 0, \lambda^{(0)} \geq 1, \delta_{\text{KL}}^{\text{thre}} > 0$

- 1:  $\alpha \leftarrow$  compute normalization factor by (1)
- 2:  $T_{\sigma}[j] \leftarrow (\sigma^{(0)}/\alpha) \cdot 10^{-j}$  for all  $j = 0, \dots, n_{\sigma}^{\text{target}} - 1$
- 3:  $T_{\mathcal{N}}[j] \leftarrow []$  for all  $j = 0, \dots, n_{\sigma}^{\text{target}} - 1$
- 4:  $J_{\text{stop}} \leftarrow []$
- 5: CMAES.INITIALIZE( $N, \lambda^{(0)}, \sigma^{(0)}$ )
- 6:  $\bar{\sigma} \leftarrow T_{\sigma}[0]$
- 7: **for**  $r = 0, \dots, n^{\text{restart}} - 1$  **do**
- 8:    $f_{\text{stop}} \leftarrow$  **False**,  $f_{\text{KL-stop}} \leftarrow$  **False**
- 9:    $J_{\text{reach}} \leftarrow []$
- 10:    $S_{\mathcal{N}}[j] \leftarrow \emptyset$  for all  $j = 0, \dots, n_{\sigma}^{\text{target}} - 1$
- 11:   **while**  $f_{\text{stop}} =$  **False** **and**  $f_{\text{KL-stop}} =$  **False** **do**
- 12:     CMAES.ONESTEP()
- 13:      $f_{\text{stop}} \leftarrow$  CMAES.CHECKSTOPPINGCRITERIA()
- 14:      $(m, \Sigma) \leftarrow$  compute normalized parameter by (1)
- 15:      $j \leftarrow -1$
- 16:     **if**  $\frac{1}{2N} \ln(\det(\Sigma)) \leq \ln(T_{\sigma}[0])$  **then**
- 17:        $j \leftarrow \max\{j \in \llbracket 0, n_{\sigma}^{\text{target}} - 1 \rrbracket \mid \frac{\ln(\det(\Sigma))}{2N} \leq \ln(T_{\sigma}[j])\}$
- 18:     **end if**
- 19:     **if**  $\text{SIZE}(J_{\text{reach}}) = 0$  **or**  $J_{\text{reach}}[-1] < j \leq n_{\sigma}^{\text{target}} - 1$  **then**
- 20:        $S_{\mathcal{N}}[j] \leftarrow \mathcal{N}(m, \Sigma)$
- 21:        $D_{\text{min}} \leftarrow \min_{\mathcal{N} \in T_{\mathcal{N}}[j]} D_{\text{KL}}(\mathcal{N} \parallel \mathcal{N}(m, \Sigma))$
- 22:        $J_{\text{reach}}.\text{APPEND}(j)$  **if**  $D_{\text{min}} \leq \delta_{\text{KL}}^{\text{thre}}$  **else**  $J_{\text{reach}} \leftarrow []$
- 23:     **else if**  $J_{\text{reach}}[-1] > j$  **then**
- 24:        $J_{\text{reach}} \leftarrow [j]$
- 25:     **end if**
- 26:      $f_{\text{KL-stop}} \leftarrow \text{SIZE}(J_{\text{reach}}) \geq n_{\text{KL}}^{\text{check}}$
- 27:   **end while**
- 28:    $T_{\mathcal{N}}[i].\text{append}(S_{\mathcal{N}}[i])$  for all  $i = 0, \dots, n_{\sigma}^{\text{target}} - 1$
- 29:    $\lambda \leftarrow \text{NEXTPOP SIZE}()$
- 30:    $J_{\text{stop}}.\text{APPEND}(J_{\text{reach}}[0])$  **if**  $f_{\text{KL-stop}} =$  **True** **else**  $J_{\text{stop}} \leftarrow []$
- 31:   **if**  $\text{SIZE}(J_{\text{stop}}) \geq n_{\sigma}^{\text{dec}}$  **and**  $J_{\text{stop}}[-i] = J_{\text{stop}}[-1]$  for all  $i = 1, \dots, n_{\sigma}^{\text{dec}}$  **then**
- 32:      $\bar{\sigma} \leftarrow T_{\sigma}[J_{\text{stop}}[-1]]$
- 33:      $J_{\text{stop}} \leftarrow []$
- 34:   **end if**
- 35:    $\alpha \leftarrow$  compute normalization factor by (1)
- 36:    $\sigma \leftarrow \min(\sigma^{(0)}, \bar{\sigma}\alpha)$
- 37:   CMAES.INITIALIZE( $N, \lambda, \sigma$ )
- 38: **end for**

---

Note that if the population size is fixed and the same normalization factor  $\alpha$  is applied for both distributions, only the first term is affected by the normalization, which is  $\alpha^2$  times smaller than the KL-divergence between non-normalized distributions.

The similarity check is performed every time the normalized std. of the current sampling distribution crosses a target  $T_{\sigma}[j]$  from above. The KL-divergence between the current normalized distribution and each normalized distribution in the history  $T_{\mathcal{N}}[j]$  of length at most the number of restarts. We check if there exists an entry in  $T_{\mathcal{N}}[j]$  such that the KL-divergence between the current normalized distribution and the entry is less than or equal to the

predefined threshold  $\delta_{\text{KL}}^{\text{thre}}$ . The above conditions are satisfied  $n_{\text{KL}}^{\text{check}}$  times in a row, we terminate and restart the run.

The threshold is derived on the basis of the quality gain analysis. Suppose that the objective function is the sphere function  $f(x) = \|x\|^2$ , the covariance matrix is proportional to the identity, and  $\mu_w \ll N$ . Then, the optimal standard deviation is given by  $\beta\mu_w\sqrt{f(m)}/N$ , where  $\beta = -\sum_{i=1}^{\lambda} w_i \mathbb{E}[\mathcal{N}_{i;\lambda}]$  and is roughly  $\sqrt{\pi}/2$  for the default weight in the CMA-ES. In the optimal situation, two distributions with the same std. reach the upper bound of the KL-divergence when  $m_0$  and  $m_1$  are symmetric around the optimal solution. Then,

$$\begin{aligned} D_{\text{KL}}(\mathcal{N}_0 \parallel \mathcal{N}_1) &= \frac{1}{2}(m_1 - m_0)^T \Sigma_1^{-1} (m_1 - m_0) \\ &= \frac{N^2 \alpha^2 \|m_1 - m_0\|^2}{2\beta^2 \mu_w^2 f(m_1)} = \frac{2N^2 \alpha^2 f(m_1)}{\beta^2 \mu_w^2 f(m_1)} = \frac{2}{\beta^2} \frac{N^2 \alpha^2}{\mu_w^2} \approx \frac{4}{\pi}. \end{aligned} \quad (3)$$

Based on this derivation, we set  $\delta_{\text{KL}}^{\text{thre}} = 2$ . That is, if the global structure of the objective function of interest is similar to that of the sphere function and a reasonable std. is adapted in the past and current restarts, the current restart will be terminated with high probability.

To prevent premature termination on globally well-structured functions such as Rastrigin function, we terminate the run if the KL-divergence between the current distribution and the history is smaller than the threshold for multiple normalized std. targets in a row (line 26).

## 2.4 Initial Normalized STD Selection

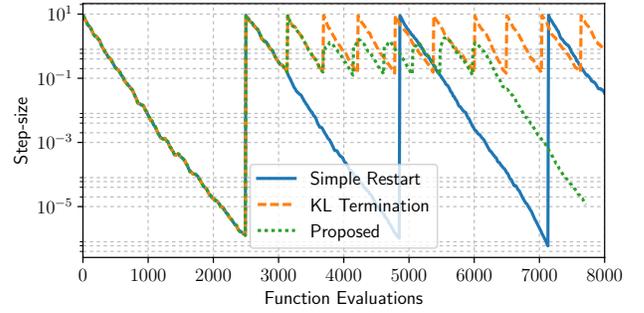
At each restart, the initial normalized std.,  $\bar{\sigma}$ , is taken from the normalized std. target list  $T_{\sigma}$ . The actual step-size  $\sigma$  is then the product of  $\bar{\sigma}$  and the normalization factor  $\alpha$  in (1). We start with  $\bar{\sigma} = \alpha T_{\sigma}[0]$ , which corresponds the given initial step-size  $\sigma^{(0)}$ . If  $n_{\sigma}^{\text{dec}}$  successive restarts have terminated due to the proposed termination criterion and all these restarts have started overlapping the distribution with the history at the same normalized std. target  $T_{\sigma}[j]$ , next restart will be performed with  $\bar{\sigma} = T_{\sigma}[j]$ .

## 2.5 Demonstration on Double-Sphere

The efficiency of the proposed mechanism is demonstrated on the double sphere function defined as follows

$$f(x) = \min(a^2 \|x_o\|^2, \|x_l\|^2 + 1.0), \quad (4)$$

where  $x_o = x - [2.5, \dots, 2.5]$  and  $x_l = x + [2.5, \dots, 2.5]$ . The global optimum is located at  $x_o$  and the radius of the basin of attraction of the global optimum is controlled by a constant  $a = 1.5$ . The initial step-size for the first (re-)start is  $\sigma^{(0)} = 10$ , which is too large to find the basin of the global optimum. Figure 1 shows the evolution of the step-size  $\sigma$  for the simple restart, the restart with the proposed termination mechanism, and the restart with the proposed termination and initial step-size selection mechanism. Compared to the simple restart, the restart strategy with the proposed termination can perform more restarts by early stopping runs searching the same basin of attraction. The proposed restart mechanism can make the initial step-size smaller so that the valley of the global optimum will be more easily found. Only the last restart of the proposed mechanism could find the global optimum. Note that on functions



**Figure 1: Step-size dynamics of the simple restart, restart with the proposed termination criterion (KL Termination), restart with the proposed termination and initial step-size selection (Proposed).**

such as the double-sphere function, a too large step-size results in sampling points away from local optima. In such cases the function values are almost identical to the sphere function  $1 + \|x_l\|^2$  centered at the local (and not the global) optimum. The CMA-ES then tends to converge towards the local optimum.

## 2.6 Restart Scheme

We combine the proposed mechanism with the simple random restart, the IPOP restart, and the BIPOP restart strategies. The parameters for the CMA-ES is set to their default values, and the parameters for the proposed mechanism are as follows:  $n^{\text{restart}} = \infty$ ,  $n_{\sigma}^{\text{target}} = 11$ ,  $n_{\text{KL}}^{\text{check}} = 2$ ,  $n_{\sigma}^{\text{dec}} = 2$ ,  $\delta_{\text{KL}}^{\text{thre}} = 2$ ,  $\lambda^{(0)} = 4 + \lfloor 3 \ln(N) \rfloor$ . The initial mean vector is drawn uniform randomly from  $[-4, 4]^N$  at each restart, and the initial step-size is  $\sigma^{(0)} = 2$  for all cases. The simple random restart strategy with the fixed population size  $\lambda = (\lambda^{(0)})^2$  combined with the proposed termination and initial standard deviation selection mechanism is denoted by KL-Restart. The proposed mechanism with IPOP mechanism (double the population size for each restart) is called KL-IPOP. The proposed mechanism with BIPOP (the initial step-size is determined by the proposed mechanism) is called KL-BIPOP. For KL-IPOP and KL-BIPOP, the maximum population size is set to  $2^8 \lambda^{(0)}$ .

## 3 CPU TIMING

In order to evaluate the CPU timing of the algorithm, we have run the KL-Restart, KL-IPOP, KL-BIPOP on the BBOB noiseless suite [10] with restarts for a maximum budget equal to  $400(D+2)$  function evaluations according to [11]. The Python code was run on a Mac Intel(R) Core(TM) i7-3615QM CPU @ 2.30GHz with 1 processor and 4 cores. The time per function evaluation for dimensions 2, 3, 5, 10, 20 equals 107, 97.2, 86.5, 63.3, 67.1 microseconds respectively for KL-restart, 351, 329, 339, 354, 457 microseconds respectively for KL-IPOP, 357, 378, 394, 416, 471 microseconds respectively for KL-BIPOP.

## 4 RESULTS

Results from experiments according to [11] and [7] on the benchmark functions given in [5, 10] are presented in Figures 2, 3 and 4. The aRT graphs for selected functions of each dimension are

displayed in Figure 5. The experiments were performed with the old BBOB code version 15.03 to compare BIPOP-CMA-ES, the plots were produced with version 2.1 of COCO [9].

The **average runtime (aRT)**, used in the figures and tables, depends on a given target function value,  $f_t = f_{\text{opt}} + \Delta f$ , 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  $f_t$  [8, 12]. **Statistical significance** is tested with the rank-sum test for a given target  $\Delta f_t$  using, for each trial, either the number of needed function evaluations to reach  $\Delta f_t$  (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.

## 5 DISCUSSION

*KL-IPOP vs IPOP.* We observed that the KL-IPOP found the optimum with less number of function evaluations than IPOP on  $f_{22}$ , and  $f_{24}$  of dimension less than 5. Since  $f_{22}$  has a small and constant number of local optima, the proposed termination and initial step-size control mechanism are expected to work effectively. The reason of the superior performance of KL-IPOP on  $f_{24}$  to IPOP is because the IPOP strategy does not decrease the initial step-size, which is required to solve it.

On the other hand, the IPOP showed better performance for  $f_{13}$ ,  $f_{17}$  and  $f_{18}$ . On  $f_{13}$ , the CMA-ES needs a restart, even though the function is unimodal. In such cases the proposed termination mechanism tends to stop a restart despite that it is searching the right big valley. The Schaffer functions  $f_{17}$  and  $f_{18}$  are globally well-structured multimodal functions with many local optima having relatively good function values around the global optimum. The proposed mechanism fails to distinguish the traces of the search distributions that approach the global optimum and a local optimum. We did not observe a similar premature termination on the Rastrigin function  $f_{15}$ , another well-structured multimodal function.

*KL-BIPOP vs BIPOP.* The difference between KL-BIPOP and BIPOP is similar to the difference between KL-IPOP and IPOP. KL-BIPOP showed better performance on  $f_{21}$  and  $f_{22}$ , while BIPOP performed better on  $f_{13}$ ,  $f_{17}$  and  $f_{18}$ . On  $f_{23}$  and  $f_{24}$ , where we did not observe significant difference between IPOP and KL-IPOP, we observed a superior performance of BIPOP to KL-BIPOP. KL-BIPOP shrinks the initial step-size only when overlapping restarts have been observed, while BIPOP starts with a randomly generated step-size. Since both  $f_{23}$  and  $f_{24}$  have many local minima in each big valley, it is not likely that a restart is regarded as overlapping. Figure 6 compares BIPOP, KL-BIPOP, and KL-BIPOP\* which is the BIPOP strategy with the proposed termination mechanism. It shows that  $f_{21}$  and  $f_{22}$  can be optimized by KL-BIPOP\*, implying that the problem on these functions are due to the initial step-size selection mechanism.

*KL-Restart vs KL-IPOP vs KL-BIPOP.* The differences between IPOP and BIPOP are simply inherited to the comparison between KL-IPOP and KL-BIPOP. For globally well-structured functions, the IPOP strategy tends to be better, whereas on weakly-structured functions BIPOP tends to be better.

KL-Restart spent more function evaluations to solve unimodal functions due to the fixed large population size. On  $f_{13}$ ,  $f_{17}$  and  $f_{18}$ , where KL-IPOP and KL-BIPOP suffered from premature termination as discussed above, KL-Restart often finds the target function value at the first (re-)start, hence it works better than KL-IPOP and KL-BIPOP. However, it is definitely not a fundamental solution to the defect of the proposed mechanism.

## 6 SUMMARY AND FUTURE WORK

We proposed the termination criterion and the initial step-size control mechanism for the restart CMA-ES by introducing the history of the distribution parameters and detecting overlapping search. The proposed mechanism is combined with the simple restart, the IPOP restart, and the BIPOP restart scheme and they are compared with the IPOP and the BIPOP restart CMA-ES on the BBOB noiseless testbed. A promising performance has been observed on  $f_{22}$ , that has a relatively small number of local minima with a weak global structure. Meanwhile, a drawback has been delighted on the Schaffer functions and the sharp ridge function. The comparison between the BIPOP strategy and the BIPOP strategy combined with the proposed initial step-size control mechanism revealed the shortcoming of the proposed initial step-size control mechanism, on e.g. the Katsuuras function  $f_{23}$ . Improvement both on the termination criterion and the initial step-size control is necessary in the future work.

**Acknowledgements.** This work is supported by JSPS KAKENHI Grant Number 15K16063.

## REFERENCES

- [1] Youhei Akimoto, Anne Auger, and Nikolaus Hansen. 2017. Quality Gain Analysis of the Weighted Recombination Evolution Strategy on General Convex Quadratic Functions. In *Foundations of Genetic Algorithms, FOGA XIV, Copenhagen, Denmark, January 12-15, 2017*. ACM, 111–126. DOI: <http://dx.doi.org/10.1145/3040718.3040720>
- [2] Dirk V. Arnold. 2005. Optimal weighted recombination. In *Foundations of Genetic Algorithms*. Springer, 215–237.
- [3] Anne Auger and Nikolaus Hansen. 2005. Performance Evaluation of an Advanced Local Search Evolutionary Algorithm. In *2005 IEEE Congress on Evolutionary Computation*. Ieee, 1777–1784.
- [4] Anne Auger and Nikolaus Hansen. 2005. A Restart CMA Evolution Strategy With Increasing Population Size. In *2005 IEEE Congress on Evolutionary Computation*. Ieee, 1769–1776.
- [5] S. Finck, N. Hansen, R. Ros, and A. Auger. 2009. *Real-Parameter Black-Box Optimization Benchmarking 2009: Presentation of the Noiseless Functions*. Technical Report 2009/20. Research Center PPE. <http://coco.lri.fr/downloads/download15.03/bbobdocfunctions.pdf> Updated February 2010.
- [6] Nikolaus Hansen. 2009. Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed. In *Workshop Proceedings of the GECCO Genetic and Evolutionary Computation Conference*. ACM Press, New York, New York, USA, 2389–2395.
- [7] N. Hansen, A. Auger, D. Brockhoff, D. Tušar, and T. Tušar. 2016. COCO: Performance Assessment. *ArXiv e-prints* arXiv:1605.03560 (2016).
- [8] N. Hansen, A. Auger, S. Finck, and R. Ros. 2012. *Real-Parameter Black-Box Optimization Benchmarking 2012: Experimental Setup*. Technical Report. INRIA. <http://coco.gforge.inria.fr/bbob2012-downloads>
- [9] N. Hansen, A. Auger, O. Mersmann, T. Tušar, and D. Brockhoff. 2016. COCO: A Platform for Comparing Continuous Optimizers in a Black-Box Setting. *ArXiv e-prints* arXiv:1603.08785 (2016).
- [10] N. Hansen, S. Finck, R. Ros, and A. Auger. 2009. *Real-Parameter Black-Box Optimization Benchmarking 2009: Noiseless Functions Definitions*. Technical Report RR-6829. INRIA. <http://coco.lri.fr/downloads/download15.03/bbobdocfunctions.pdf> Updated February 2010.
- [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*. 153–157.

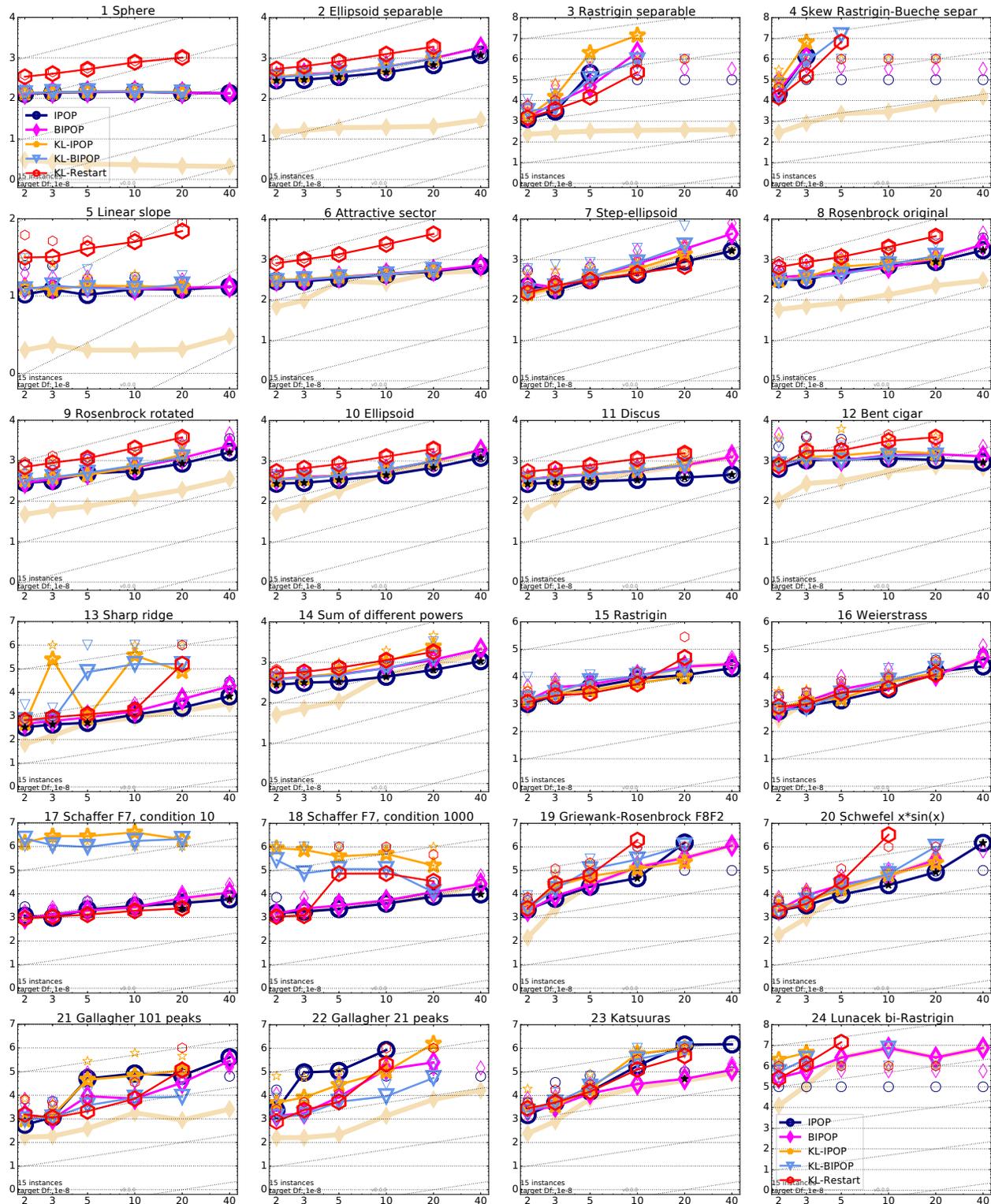


Figure 2: Average running time (ART in number of  $f$ -evaluations as  $\log_{10}$  value), divided by dimension for target function value  $10^{-8}$  versus dimension. Slanted grid lines indicate quadratic scaling with the dimension. Different symbols correspond to different algorithms given in the legend of  $f_1$  and  $f_{24}$ . Light symbols give the maximum number of function evaluations from the longest trial divided by dimension. Black stars indicate a statistically better result compared to all other algorithms with  $p < 0.01$  and Bonferroni correction number of dimensions (six). Legend:  $\circ$ : BIPOP-CMAES,  $\diamond$ : IPOP-CMAES,  $\star$ : KL-CMAES,  $\nabla$ : KL-IPOP-CMAES,  $\ominus$ : KL-Restart

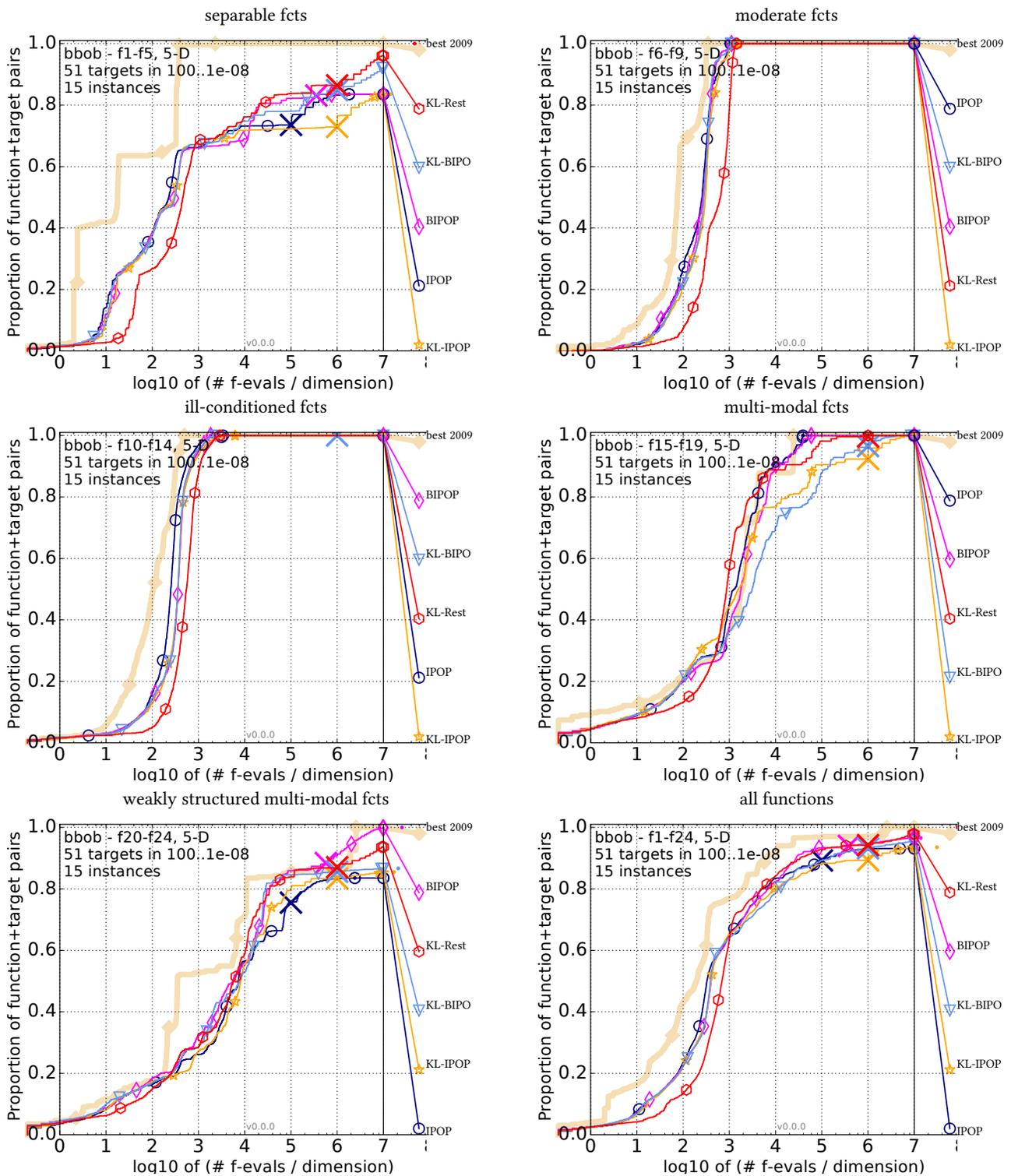


Figure 3: Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/DIM) for 51 targets with target precision in  $10^{[-8..2]}$  for all functions and subgroups in 5-D. The “best 2009” line corresponds to the best aRT observed during BBOB 2009 for each selected target.

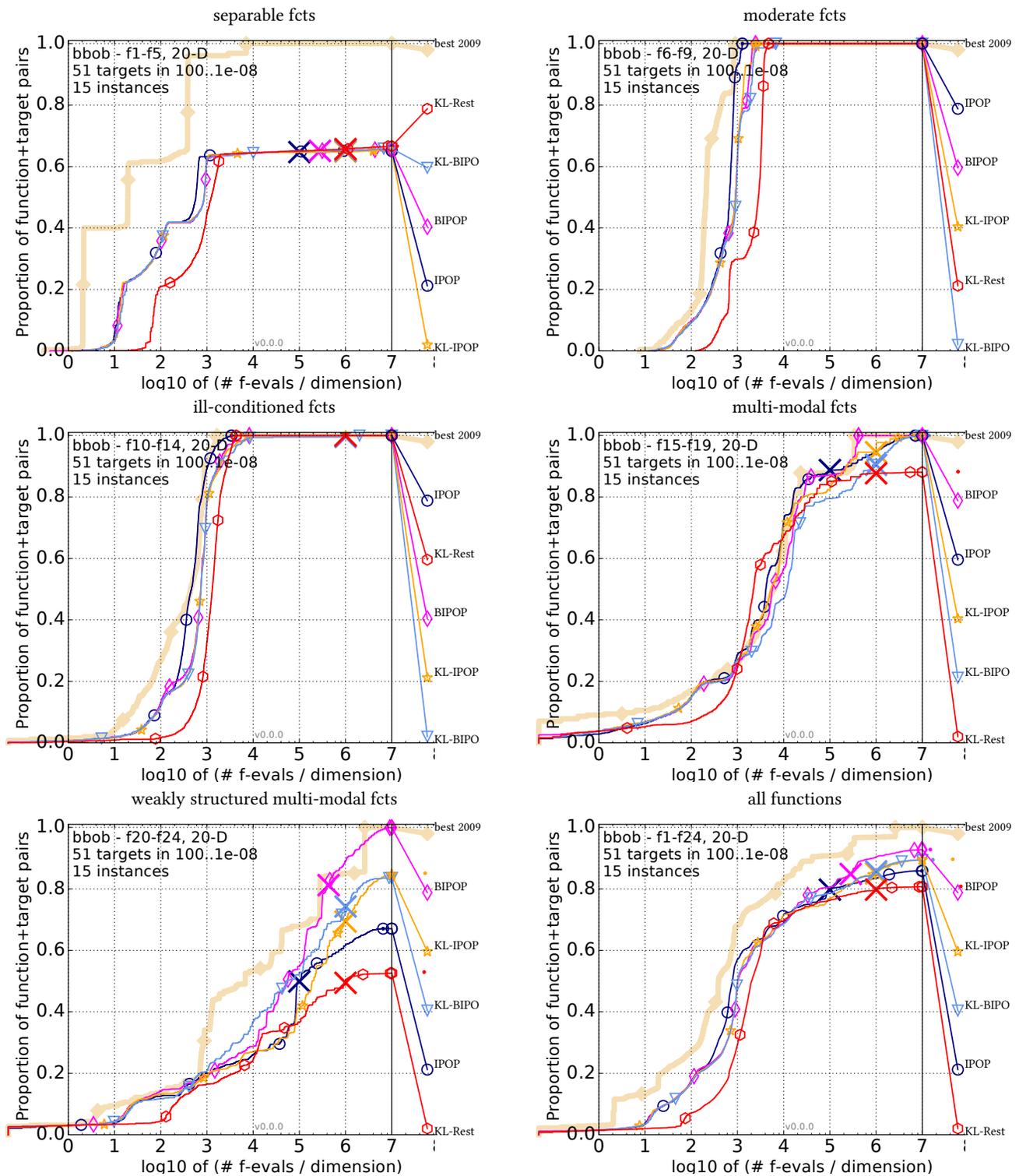


Figure 4: Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/DIM) for 51 targets with target precision in  $10^{[-8..2]}$  for all functions and subgroups in 20-D. The “best 2009” line corresponds to the best aRT observed during BBOB 2009 for each selected target.

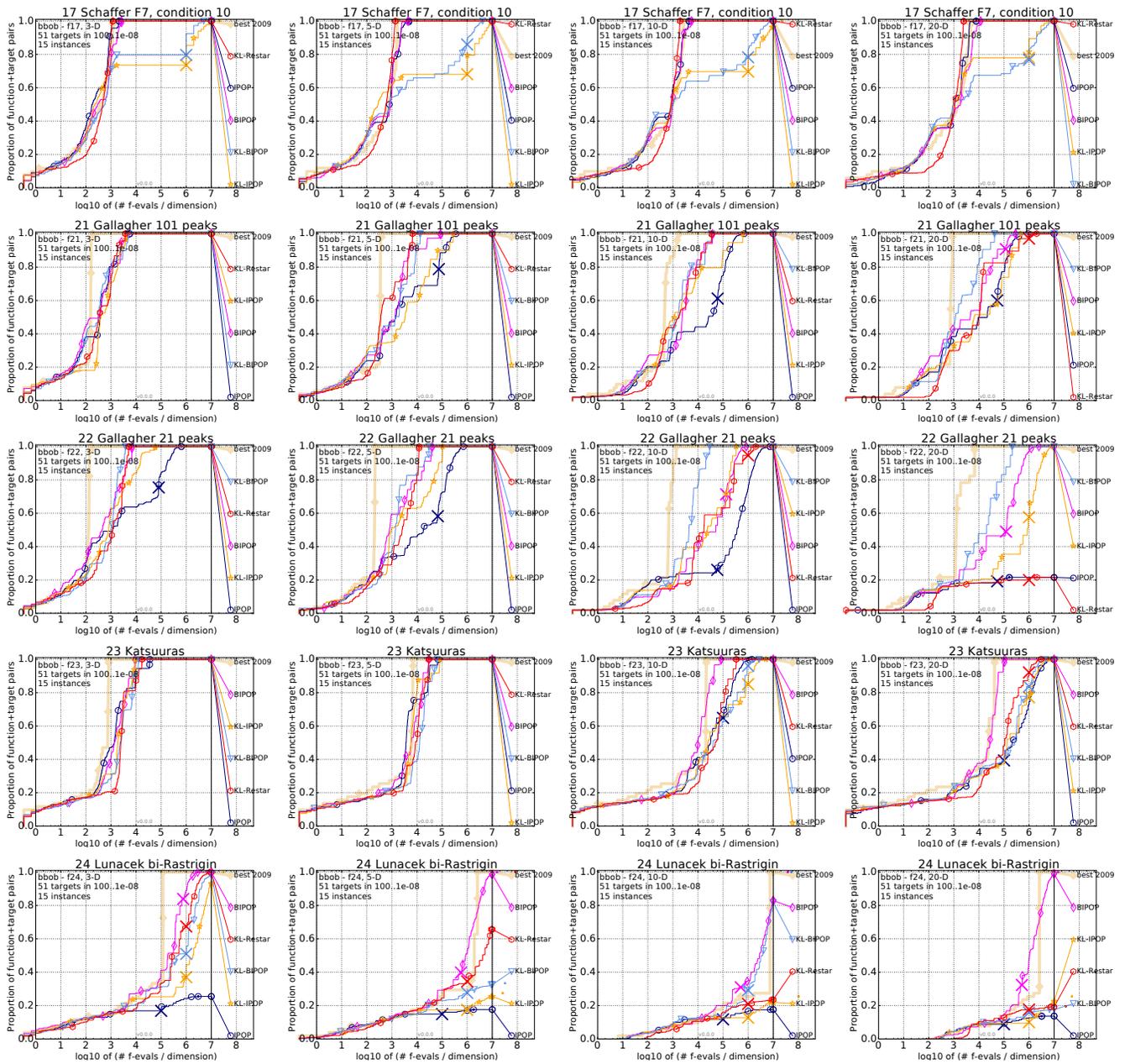


Figure 5: Results on  $f_{17}$ ,  $f_{21}$ ,  $f_{22}$ ,  $f_{23}$  and  $f_{24}$  (from top to bottom) of dimension 3, 5, 10, and 20 (from left to right).

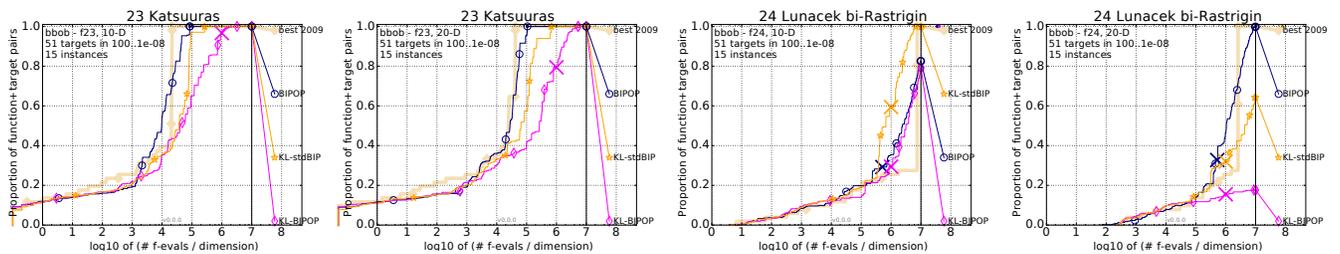


Figure 6: Results on  $f_{23}$  and  $f_{24}$  (left top to right) of dimension 10, and 20 (from left to right).