

CMA-ES with Restarts for Solving CEC 2013 Benchmark Problems

Ilya Loshchilov

Laboratory of Intelligent Systems
École Polytechnique Fédérale de Lausanne, Switzerland
Email: ilya.loshchilov@epfl.ch

Abstract—This paper investigates the performance of 6 versions of Covariance Matrix Adaptation Evolution Strategy (CMA-ES) with restarts on a set of 28 noiseless optimization problems (including 23 multi-modal ones) designed for the special session on real-parameter optimization of CEC 2013. The experimental validation of the restart strategies shows that: i). the versions of CMA-ES with weighted active covariance matrix update outperform the original versions of CMA-ES, especially on ill-conditioned problems; ii). the original restart strategies with increasing population size (IPOP) are usually outperformed by the bi-population restart strategies where the initial mutation step-size is also varied; iii). the recently proposed alternative restart strategies for CMA-ES demonstrate a competitive performance and are ranked first w.r.t. the proportion of function-target pairs solved after the full run on all 10-, 30- and 50-dimensional problems.

I. INTRODUCTION

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) proposed by [8], [7] has become a standard for continuous black-box evolutionary optimization. The main advantage of CMA-ES over classical Evolution Strategies comes from the use of correlated mutations instead of axis-parallel ones. The adaptation of the covariance matrix \mathbf{C} allows to steadily learn appropriate mutation distribution and increase the probability of repeating the successful search steps.

However, there are several properties of black-box optimization problems which may lead to a premature convergence of CMA-ES, among the most common are multi-modality and uncertainty. To increase the probability of finding the global optima, IPOP-CMA-ES [2] and BIPOP-CMA-ES [4] restart strategies for CMA-ES have been proposed. The IPOP-CMA-ES was ranked first on the continuous optimization benchmark at CEC 2005 [3]; and BIPOP-CMA-ES showed the best results together with IPOP-CMA-ES on the black-box optimization benchmark (BBOB) in 2009 and 2010 [1]. Later, alternative restart strategies for CMA-ES proposed in [12] demonstrated an even more competitive performance on some of multi-modal functions during the BBOB 2012. The recently proposed *weighted active* covariance matrix update of CMA-ES [11], [9] is also a competitive alternative to the original update procedure, it allows to substantially improve the performance both on unimodal and multi-modal functions [9]. This paper focuses on analyzing the performance of the restart strategies of CMA-ES with the original and weighted active covariance matrix updates on the CEC 2013 benchmark test [10].

The remainder of this paper is organized as follows. Section

II presents the main principles of the CMA-ES algorithm. Section III describes the restart strategies of CMA-ES. Section IV explains the experimental procedure and comments the experimental results. Section V concludes the paper with a discussion and some perspectives for further research.

II. THE $(\mu/\mu_w, \lambda)$ -CMA-ES

The CMA-ES algorithm [8], [7] optimizes an objective function $f : x \in \mathbb{R}^n \rightarrow f(x) \in \mathbb{R}$ by sampling λ candidate solutions from a multivariate normal distribution. It exploits the best μ solutions out of the λ ones to adaptively estimate the local covariance matrix of the objective function, in order to increase the probability of successful samples in the next iteration. More formally, at iteration t , $(\mu/\mu_w, \lambda)$ -CMA-ES samples λ individuals ($k = 1 \dots \lambda$) according to

$$\mathbf{x}_k^{(t+1)} = \mathcal{N}\left(\mathbf{m}^{(t)}, \sigma^{(t)2} \mathbf{C}^{(t)}\right) = \mathbf{m}^{(t)} + \sigma^{(t)} \cdot \mathcal{N}\left(\mathbf{0}, \mathbf{C}^{(t)}\right), \quad (1)$$

where $\mathbf{m}^{(t)}$ denotes the mean of a normally distributed random vector, $\mathbf{C}^{(t)}$ is the covariance matrix and $\sigma^{(t)}$ is the mutation step-size.

These λ individuals are evaluated and ranked. The mean of the distribution is updated and set to the weighted sum of the best μ individuals as $\mathbf{m}^{(t+1)} = \sum_{i=1}^{\mu} w_i \mathbf{x}_{i:\lambda}^{(t)}$, with $w_i > 0$ for $i = 1 \dots \mu$ and $\sum_{i=1}^{\mu} w_i = 1$, where index $i : \lambda$ denotes the i -th best individual after the objective function. In the original CMA-ES the information about the remaining (worst $\lambda - \mu$) solutions is used only implicitly during the selection process.

However, it has been shown in [11] that the information from the worst solutions also can be used to reduce the variance of the mutation distribution in unpromising directions. The corresponding *active* $(\mu/\mu_I, \lambda)$ -CMA-ES algorithm demonstrates a performance gain up to a factor of 2 without loss of performance on any of tested functions in [11]. Later, the *active* update of $(\mu/\mu_I, \lambda)$ -CMA-ES was extended to the *weighted* case of $(\mu/\mu_W, \lambda)$ -CMA-ES, where $w_i > w_{i+1}$ for $i = 1 \dots \lambda - 1$. This *weighted active* $(\mu/\mu_W, \lambda)$ -CMA-ES (also referred to as aCMA-ES) was implemented in the IPOP regime of restarts as IPOP-aCMA-ES and demonstrated improvements up to a factor of 2 on a set of noiseless and noisy functions from the BBOB [9].

More formally, the active CMA-ES only differs from the original CMA-ES in the adaptation of the covariance matrix $\mathbf{C}^{(t)}$. Like for CMA-ES, the covariance matrix is computed from the best μ solutions, $\mathbf{C}_\mu^+ = \sum_{i=1}^{\mu} w_i \frac{\mathbf{x}_{i:\lambda} - \mathbf{m}^t}{\sigma^t} \times$

$\frac{(\mathbf{x}_{i:\lambda-m^t})^T}{\sigma^t}$. The main novelty is to exploit the worst solutions to compute $\mathbf{C}_\mu^- = \sum_{i=0}^{\mu-1} w_{i+1} \mathbf{y}_{\lambda-i:\lambda} \mathbf{y}_{\lambda-i:\lambda}^T$, where $\mathbf{y}_{\lambda-i:\lambda} = \frac{\|\mathbf{C}^{t-1/2}(\mathbf{x}_{\lambda-\mu+1+i:\lambda-m^t})\|}{\|\mathbf{C}^{t-1/2}(\mathbf{x}_{\lambda-i:\lambda-m^t})\|} \times \frac{\mathbf{x}_{\lambda-i:\lambda-m^t}}{\sigma^t}$. The covariance matrix estimation of these worst solutions is used to decrease the variance of the mutation distribution along these directions:

$$\mathbf{C}^{t+1} = (1 - c_1 - c_\mu + c^- \alpha_{old}^-) \mathbf{C}^t + c_1 \mathbf{p}_c^{t+1} \mathbf{p}_c^{t+1 T} + (c_\mu + c^- (1 - \alpha_{old}^-)) \mathbf{C}_\mu^+ - c^- \mathbf{C}_\mu^-, \quad (2)$$

where \mathbf{p}_c^{t+1} is adapted along the evolution path and coefficients c_1 , c_μ , c^- and α_{old}^- are defined such that $c_1 + c_\mu - c^- \alpha_{old}^- \leq 1$. The interested reader is referred to [7], [9] for a more detailed description of these algorithms.

A potential issue of the active update is that the positive definiteness of the covariance matrix cannot be guaranteed anymore, that may result in algorithmic instability. According to [12], this issue is not observed on the BBOB benchmark suite [5]. In our experiments with the CEC 2013 benchmark suite this issue is also never observed.

III. RESTART STRATEGIES FOR CMA-ES

A. Preliminary Analysis

The CMA-ES algorithm is a local search optimizer and its default population size $\lambda_{default}$ has been tuned for unimodal functions. On multi-modal functions, however, it can get stuck in local optima and the convergence to global optima is not guaranteed. Various approaches to increase the probability of finding global optima have been proposed, many of them belong to i). niching approaches and ii). restart strategies.

A representative approach of the first category is the CMA-ES with the fitness sharing [15], where the niche radius is adapted during the search that allows to keep several running individual CMA-ES instances on a certain distance from each other, and, thus, maintain some diversity. Another example is the NBC-CMA-ES algorithm [14] with the niching via Nearest-Better Clustering (NBC) which is employing a radius-free basin identification method. In this approach, the niches are dynamically identified and the corresponding points are used to form populations for individual CMA-ES instances. According to [14], for very highly multi-modal functions, the effort invested into the coordination of local searches often does not pay off as it becomes almost impossible to identify enough basins of attraction to obtain an advantage over uncoordinated restarts.

The second category of restart strategies is not that different from the first one since restarts also can be viewed as a parallelized search, but rather in the time than in space [14]. A milestone paper [6] investigated the probability of reaching the global optimum (and the overall number of function evaluations needed to do so) w.r.t. the population size of CMA-ES. The analysis of empirical results demonstrated that, indeed, this probability is very sensitive to the population size and that the default population size of CMA-ES is rather too small. The restart strategies described in the following sections are inspired by an idea of exploring CMA-ES hyper-parameters such as the population size and the initial step-size.

B. The IPOP-CMA-ES and IPOP-aCMA-ES

As mentioned, [6] demonstrated that increasing the population size improves the performance of CMA-ES on multi-modal functions. The authors of [6] suggested a restart strategy for CMA-ES with successively increasing population size. Such an algorithm was later introduced in [2] as IPOP-CMA-ES. IPOP-CMA-ES only aims at increasing the population size λ . Each time at least one of the stopping criteria is met by the CMA-ES, it launches a new CMA-ES with population size $\lambda = \rho_{inc}^{i_{restart}} \lambda_{default}$, where $i_{restart}$ is the index of the restart and $\lambda_{default}$ is the default population size. Factor ρ_{inc} must be not too large to avoid "overjumping" some possibly optimal population size λ^* ; in [2] it is set to $\rho_{inc} = 2$ that in certain cases allows to keep a potential loss in terms of function evaluations (compared to the "oracle" restart strategy which would directly set the population size to the optimal value λ^*) by about a factor of 2.

The active version of IPOP-CMA-ES (IPOP-aCMA-ES) has been proposed in [9].

C. The BIPOP-CMA-ES and BIPOP-aCMA-ES

In BIPOP-CMA-ES [4] after the first single run with default population size, the algorithm is restarted in one of two possible regimes and account the budget of function evaluations spent in the corresponding regime. Each time the algorithm is restarted, the regime with smallest budget used so far is used.

Under the first regime the population size is doubled as $\lambda_{large} = 2^{i_{restart}} \lambda_{default}$ in each restart $i_{restart}$ and use some fixed initial step-size $\sigma_{large}^0 = \sigma_{default}^0$. This regime corresponds to the IPOP-CMA-ES.

Under the second regime the CMA-ES is restarted with some small population size λ_{small} and step-size σ_{small}^0 , where λ_{small} is set to

$$\lambda_{small} = \left\lfloor \lambda_{default} \left(\frac{1}{2} \frac{\lambda_{large}}{\lambda_{default}} \right)^{U[0,1]^2} \right\rfloor, \quad (3)$$

Here $U[0,1]$ denotes independent uniformly distributed numbers in $[0,1]$ and $\lambda_{small} \in [\lambda_{default}, \lambda/2]$. The initial step-size is set to $\sigma_{small}^0 = \sigma_{default}^0 \times 10^{-2U[0,1]}$.

In each restart, BIPOP-CMA-ES selects the restart regime with less function evaluations used so far. Since the second regime uses a smaller population size, it is therefore launched more often.

The active version of BIPOP-CMA-ES (BIPOP-aCMA-ES) has been proposed in [12].

D. The NIPOP-aCMA-ES

The NIPOP-aCMA-ES [12] is an alternative restart strategy to the IPOP-aCMA-ES, where in addition to increasing of population size in each restart, the initial step-size is also decreased by some factor $k_{\sigma dec}$. In [12], this factor is set to $k_{\sigma dec} = 1.6$ such that σ value after 9 restarts (the default maximum number of restarts in BIPOP-aCMA-ES) roughly corresponds to the minimum possible initial $\sigma = 10^{-2} \sigma_{default}$

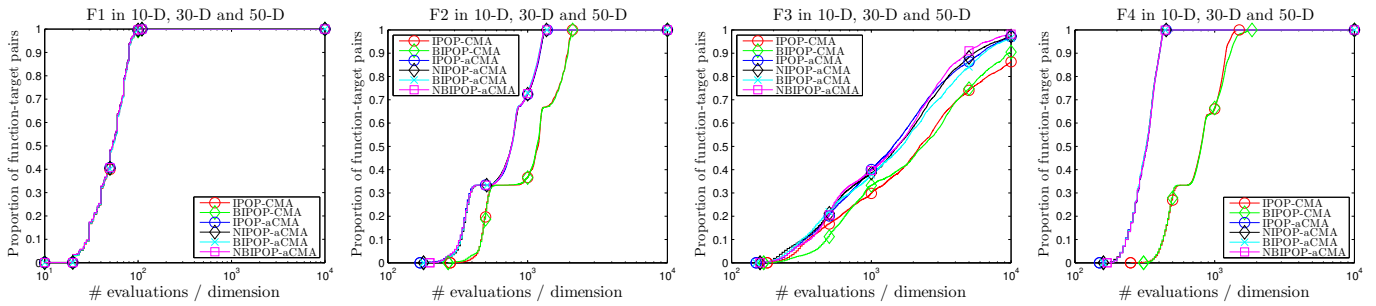


Fig. 2. Empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/D) for 300 function-target pairs in $10^{[-1..4]}$ (100 pairs for each of dimensions 10, 30 and 50) for F1, F2, F3.

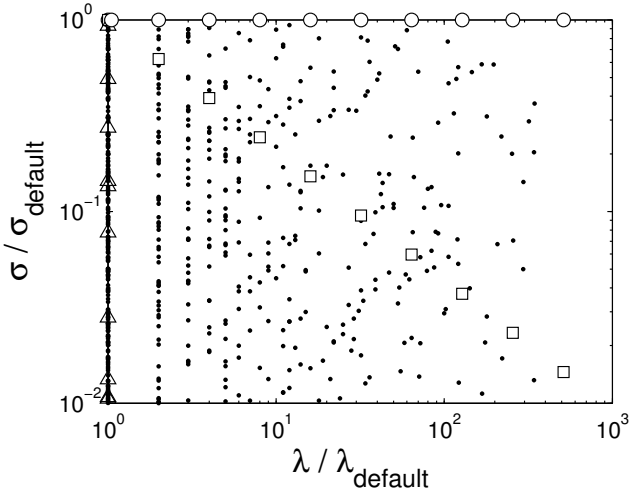


Fig. 1. An illustration of λ and σ hyper-parameters distribution for 9 restarts of IPOP-aCMA-ES (\circ), BIPOP-aCMA-ES (\circ and \cdot for 10 runs), NIPOP-aCMA-ES (\square) and NBIPOP-aCMA-ES (\square and many \triangle for $\lambda/\lambda_{default} = 1$, $\sigma/\sigma_{default} \in [10^{-2}, 10^0]$). The first run of all algorithms corresponds to the point with $\lambda/\lambda_{default} = 1$, $\sigma/\sigma_{default} = 1$.

used for BIPOP-aCMA-ES. This strategy represents an alternative to the BIPOP-aCMA-ES in the case if the restart strategy is restricted to increasing of population size. It also outperforms IPOP-aCMA-ES and is competitive with BIPOP-aCMA-ES on the BBOB noiseless problems [13].

E. The NBIPOP-aCMA-ES

In NBIPOP-aCMA-ES [12] as well as in BIPOP-aCMA-ES there are two restart regimes:

- i). Double the population size and decrease the initial step-size by $k_{\sigma_{dec}} = 1.6$ (NIPOP-aCMA-ES).
- ii). Launch CMA-ES with default population size $\lambda_{default}$ and $\sigma^0 = \sigma_{default}^0 \times 10^{-2U[0,1]}$.

In contrast with BIPOP-aCMA-ES, where both regimes have the same budget, the budget is adapted here according to the performance of the regime: the best solutions x_A^* and x_B^* found by regimes A and B are used as an estimate of the quality of the regimes. Thus, $k_{budget} = 2$ times larger computation budget is allocated for regime A if it performs better than B (i.e., if x_A^* is better than x_B^*), and vice versa.

The NBIPOP-aCMA-ES typically outperforms IPOP-aCMA-ES, BIPOP-aCMA-ES and NIPOP-aCMA-ES on the BBOB noiseless problems [13], especially in larger dimensions.

All the above described algorithms can be viewed as some search algorithms in the space of hyper-parameters λ and σ . The typical patterns of these search algorithms are shown in Fig. 1.

IV. EXPERIMENTAL VALIDATION

The experimental validation investigates the performance of 6 CMA-ES restart strategies: IPOP-CMA-ES, BIPOP-CMA-ES, IPOP-aCMA-ES, BIPOP-aCMA-ES, NBIPOP-aCMA-ES, NBIPOP-aCMA-ES. We use the source code ¹ provided by the authors of [12], which is based on the original MATLAB code ² of CMA-ES provided by N. Hansen. Both for IPOP and BIPOP versions the default parameter settings are used as given in [9], [4], [12]. The initial step-size σ is chosen according to the given search range $[-100; 100]$ as $0.6 \cdot 200 = 120$.

For all functions and dimensions the maximum number of function evaluations was set to $10000n$.

A. Results

The results individually for each function and problem dimension are given according to [10] in Tables II-XIX after the maximum number of function evaluations.

To assess the performance of the algorithms we use a procedure similar to one used in BBOB framework: for each objective function we define a set of function-target pairs Δf_t in the range $[10^{-1}, 10^4]$. The lower bound of 10^{-1} is chosen because for most of *multi-modal* functions the objective values below 10^{-1} are usually difficult to achieve. Fig. 2 and 3 depict the empirical cumulative distribution of running time of the annotated algorithm individually on all objective functions. Importantly, the results for all 3 problem dimensions and 51 runs are aggregated such that if the proportion of function-target pairs equals to 1 after a given number of function evaluations, then all $3 \cdot 100 = 300$ function-target pairs have been solved 51 times (i.e., 15300 problems solved) by the corresponding algorithm. For some functions, e.g., F20, the y-axis is scaled to better illustrate the difference in performances.

¹<https://sites.google.com/site/ppsnbipop/>

²https://www.lri.fr/~hansen/cmaes_inmatlab.html

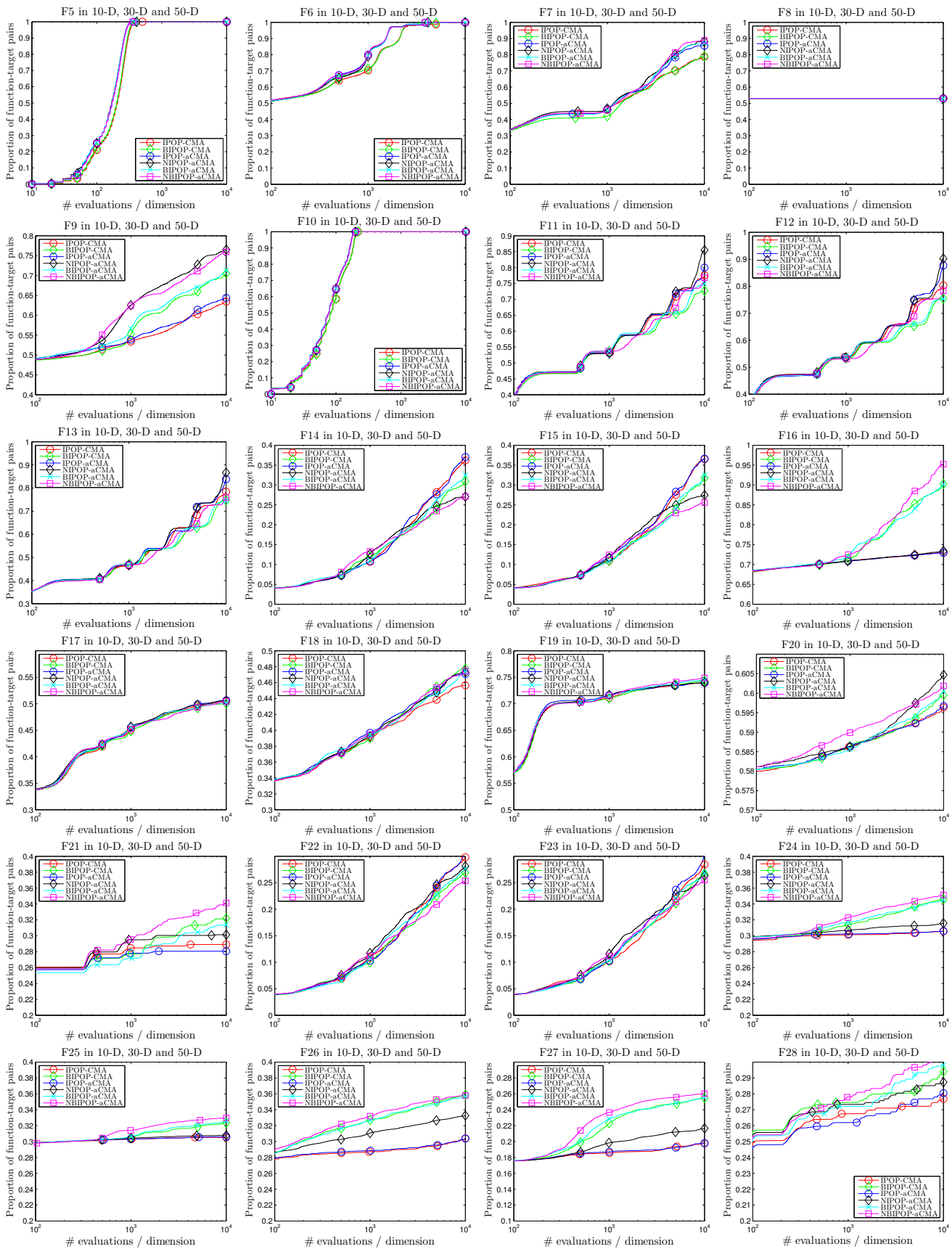


Fig. 3. Continuation of Fig. 2.

TABLE I. COMPUTATIONAL COMPLEXITY OF ALL 6 ALGORITHMS GIVEN FOR 10-, 30- AND 50-DIMENSIONAL SCHWEFEL'S FUNCTION (F14).

	T0	T1	(T2 - T1) / T0 for IPOP-CMA	IPOP-aCMA	BIPOP-CMA	BIPOP-aCMA	NIPOP-aCMA	NBIPOP-aCMA
D=10	0.277	1.778	22.64	25.45	45.97	45.08	59.26	62.39
D=30	0.277	2.929	38.20	45.66	56.20	64.96	63.64	68.11
D=50	0.277	4.159	57.29	69.41	84.06	85.43	102.38	103.79

Active covariance matrix update. The active versions of CMA-ES clearly outperform the original ones on unimodal ill-conditioned functions F2, F3, F4. A substantial improvement is also observed on F5, F6, F7. The only function, where the original versions seem to perform better is F21 composition function of functions F1, F3, F4, F5 and F6, i.e., on which the active versions actually perform better. This is an unexpected result and requires further analysis.

BIPOP vs IPOP. BIPOP-based algorithms outperform IPOP-based algorithms on F9, F14, F16, F20, F21, F24, F25, F26, F27, F28, and are outperformed by the latter on F11, F12, F13, F14 and F15. While in some cases the difference is minor, in overall, BIPOP-based algorithms perform better on composition functions.

NBIPOP and NIPOP vs BIPOP and IPOP. The alternative restart strategies outperform the original ones on F9, F12, F16, F20, F24, F25, F26, F27, F28, and demonstrate a comparable performance on other functions.

Computational Complexity. The results of experimental runs on F14 Schwefel's function are given in Table I according to [10]. The restart strategies where smaller population sizes are used (e.g., NBIPOP-aCMA-ES) spend more time on internal computations per function evaluation, and are typically up to 2 times slower in terms of time than IPOP-CMA-ES.

V. CONCLUSION AND PERSPECTIVES

In this paper, we have compared the original and recently proposed restart strategies for CMA-ES on the CEC 2013 test suite. The aggregated results depicted in Fig. 4 demonstrate a slightly better performance of the NBIPOP-aCMA-ES and NIPOP-aCMA-ES. A possible reason is that a smaller initial step-size is especially useful on composition functions where the basins of attractions are relatively small. The results also confirm some superiority of the active covariance matrix update.

The main limitation of all tested approaches is that the search in the hyper-parameter space of the population size and initial step-size seems to be inefficient and some potentially useful information from the restarts (e.g., the location of the best found solution) is not used. Another important limitation inherited from the CMA-ES is a lack of functionality which would allow to detect and exploit the separability of the objective function. Thus, the algorithms which specifically focus on separable and partially-separable functions will very likely outperform the CMA-ES and its restarts strategies. The above-described issues need to be addressed in future work.

REFERENCES

[1] A. Auger, S. Finck, N. Hansen, and R. Ros. BBOB 2010: Comparison Tables of All Algorithms on All Noiseless Functions. Technical Report RR-7215, INRIA, 2010.

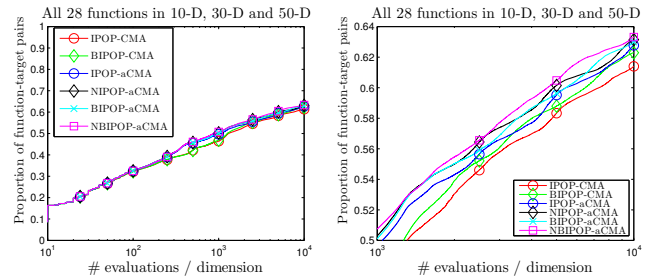


Fig. 4. Empirical cumulative distribution of all function-target pairs solved on all functions, dimensions and runs (in overall, 428400 pairs).

[2] A. Auger and N. Hansen. A Restart CMA Evolution Strategy With Increasing Population Size. In *IEEE Congress on Evolutionary Computation*, pages 1769–1776. IEEE Press, 2005.

[3] N. Hansen. Compilation of results on the 2005 CEC benchmark function set. *Online*, May, 2006.

[4] N. Hansen. Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed. In *GECCO Companion*, pages 2389–2396, 2009.

[5] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-Parameter Black-Box Optimization Benchmarking 2010: Experimental Setup. Technical Report RR-7215, INRIA, 2010.

[6] N. Hansen and S. Kern. Evaluating the CMA Evolution Strategy on Multimodal Test Functions. In *PPSN'04*, pages 282–291, 2004.

[7] N. Hansen, S. Müller, and P. Koumoutsakos. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18, 2003.

[8] N. Hansen and A. Ostermeier. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evol. Comput.*, 9(2):159–195, June 2001.

[9] N. Hansen and R. Ros. Benchmarking a weighted negative covariance matrix update on the BBOB-2010 noiseless testbed. In *GECCO '10: Proceedings of the 12th annual conference comp on Genetic and evolutionary computation*, pages 1673–1680, New York, NY, USA, 2010. ACM.

[10] P. N. S. J. Liang, B.-Y. Qu and A. G. Hernandez-Diaz. Problem Definitions and Evaluation Criteria for the CEC 2013 Special Session and Competition on Real-Parameter Optimization. Technical report, Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou China and Technical Report, Nanyang Technological University, 2013.

[11] G. A. Jastrebski and D. V. Arnold. Improving Evolution Strategies through Active Covariance Matrix Adaptation. In *IEEE Congress on Evolutionary Computation*, pages 2814–2821, 2006.

[12] I. Loshchilov, M. Schoenauer, and M. Sebag. Alternative Restart Strategies for CMA-ES. In V. C. et al., editor, *Parallel Problem Solving from Nature (PPSN XII)*, LNCS, pages 296–305. Springer, 2012.

[13] I. Loshchilov, M. Schoenauer, and M. Sebag. Black-box Optimization Benchmarking of NIPOP-aCMA-ES and NBIPOP-aCMA-ES on the BBOB-2012 Noiseless Testbed. In T. Soule and J. H. Moore, editors, *Genetic and Evolutionary Computation Conference (GECCO Companion)*, pages 269–276. ACM Press, July 2012.

[14] M. Preuss. Niching the cma-es via nearest-better clustering. In *Proceedings of the 12th annual conference companion on Genetic and evolutionary computation*, pages 1711–1718. ACM, 2010.

[15] O. M. Shir, M. Emmerich, and T. Bäck. Adaptive niche radii and niche shapes approaches for niching with the cma-es. *Evolutionary Computation*, 18(1):97–126, 2010.

