An adapting population size approach in the CMA-ES for multimodal functions

Duc Manh Nguyen
Department of Mathematics and Informatics, Hanoi National University of Education
Hanoi, Vietnam
nguyendm@hnue.edu.vn

ABSTRACT

In this work, we give some numerical investigations of the CMAES-APOP on some multi-modal functions and introduce a relaxed version of this algorithm to cope with the hard Schwefel function. The numerical simulations will show the efficiency of our approach.

CCS CONCEPTS

• **Computing methodologies → Continuous space search**;

ACM Reference Format:

Duc Manh Nguyen. 2018. An adapting population size approach in the CMA-ES for multimodal functions. In *GECCO '18 Companion: Genetic and Evolutionary Computation Conference Companion, July 15–19, 2018, Kyoto, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3205651.3205801

1 INTRODUCTION

For adapting population size in the CMAES, in the literature there are well-known and successful methods [1, 2]. In [4], we have introduced the CMAES-APOP algorithm which is inspired from a natural desire when solving an optimization problem as well as one prospect when using larger population size (pop-size for short) to search: *we want to see the decrease of the objective function.* In this method, the non-decrease of objective function in a slot of $S = 5$ successive iterations is tracked to adapt the pop-size for next S successive iterations. Because the pop-size is adapted in each slot of S iterations, its variation takes a staircase form in iterations. In this work, we give some numerical investigations for this method. Besides testing the performance of CMAES-APOP on some multi-modal functions, we also study the effect of percentiles on CMAES-APOP's performance. Moreover, we present briefly a relaxed version of the CMAES-APOP for solving efficiently the Schwefel function with weak global structure.

2 NUMERICAL EXPERIMENT

In the following experiments, we run the CMAES-APOP with small initial pop-size $\lambda = \lambda_{\text{default}}$ (i.e, set $k_n = 1$, see [4]); and except for a case in section 2.2, there is no upper bound for pop-size. The unconstrained multi-modal test problems [3] are summarized in Table 1. The initial parameters for our algorithms and the IPOP-CMAES algorithm are given in the Table 2. These functions have a high number of local optima, are scalable in the problem dimension, and have a minimal function value of 0. The known global minimum is located at $\boldsymbol{x} = 0$, except for the Schwefel function, where the global minimum within $[-500, 500]^n$ equals 420.96874636 in each coordinate. The performance of algorithm is tested for dimensions $n = [5, 10, 20]$ on the Schwefel function, and for dimensions $n = [10, 20, 40]$ on the other functions. Also, the bound constraints for the Ackley function in $[-32.768, 32.768]^n$ and Schwefel function in $[-500, 500]^n$ are considered via quadratic penalty terms. That is, the following functions $f_{\text{Schwefel}}(\boldsymbol{x}) + 10^4 \sum_{i=1}^n \theta(|x_i| - 500).(|x_i| - 500)^2$ and $f_{\text{Ackley}}(\boldsymbol{x}) + \sum_{i=1}^n \theta(|x_i| - 32.768).(|x_i| - 32.768)^2$ will be minimized, where $\theta(x) = 1$ if $x > 0$ and $\theta(x) = 0$ if $x \leq 0$.

Name	Function
Rastrigin	$f_{\text{Rastrigin}}(\boldsymbol{x}) = 10n + \sum_{i=1}^n (x_i^2 - 10\cos(2\pi x_i))$
Scale Rastrigin	$f_{\text{RastScale}}(\boldsymbol{x}) = 10n + \sum_{i=1}^n ((10^{\frac{i-1}{n-1}} x_i)^2 - 10\cos(2\pi 10^{\frac{i-1}{n-1}} x_i))$
Schaffer	$f_{\text{Schaffer}}(\boldsymbol{x}) = \sum_{i=1}^{n-1} (x_i^2 + x_{i+1}^2)^{0.25}[\sin^2(50(x_i^2 + x_{i+1}^2)^{0.1}) + 1]$
Griewank	$f_{\text{Griewank}}(\boldsymbol{x}) = \frac{1}{4000}\sum_{i=1}^n x_i^2 - \Pi_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$
Ackley	$f_{\text{Ackley}}(\boldsymbol{x}) = 20 - 20 \cdot \exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^n x_i^2}\right) + \mathrm{e} - \exp\left(\frac{1}{n}\sum_{i=1}^n \cos(2\pi x_i)\right)$
Bohachevsky	$f_{\text{Bohachevsky}}(\boldsymbol{x}) = \sum_{i=1}^{n-1}(x_i^2 + 2x_{i+1}^2 - 0.3\cos(3\pi x_i) - 0.4\cos(4\pi x_{i+1}) + 0.7)$
Schwefel	$f_{\text{Schwefel}}(\boldsymbol{x}) = 418.9828872724339 \cdot n - \sum_{i=1}^n x_i \cdot \sin(\sqrt{\lvert x_i \rvert})$

Table 1: Test functions.

Function	Initial point	σ
Rastrigin	$\boldsymbol{x}^0 = (5, \ldots, 5)$	2
Scale Rastrigin	$\boldsymbol{x}^0 = (5, \ldots, 5)$	2
Schaffer	$\boldsymbol{x}^0 = (55, \ldots, 55)$	20
Griewank	$\boldsymbol{x}^0 = (305, \ldots, 305)$	100
Ackley	$\boldsymbol{x}^0 = (15, \ldots, 15)$	5
Bohachevsky	$\boldsymbol{x}^0 = (8, \ldots, 8)$	3
Schwefel	$\boldsymbol{x}^0 = (0, \ldots, 0)$	500

Table 2: Initial conditions.

For each function, 51 runs are conducted. Each run is stopped and regarded as successful, when the function value is smaller than $f_{\text{stop}} = 10^{-10}$ ($f_{\text{stop}} = 10^{-8}$ for the Schaffer function). Some additional stopping conditions that are added to the Schaffer function are: $\texttt{TolX} = 10^{-30}, \texttt{TolFun} = 10^{-20}, \texttt{TolHistFun} = 10^{-20}$. We used the matlab implementation of CMA-ES, version 3.40.beta to make CMAES-APOP and its variants.

2.1 CMAES-APOP on the first six functions

From table 3 we can see that CMAES-APOP provides high success rates (more than 80%) for the first six functions. CMAES-APOP is worse than IPOP-CMAES on the Griewank and Bohachevsky functions. It is because they are quite easy multi-modal functions, and we have wasted a lot of populations at the beginning. Nevertheless, CMAES-APOP is better than IPOP-CMAES on Rastrigin, Scale Rastrigin, Schaffer and Ackley functions. It runs faster than IPOP-CMAES does about 2-3 times on the Rastrigin, Scale Rastrigin (well-structured) functions in all dimensions; 1.5 times on the Schaffer function in all dimensions. On the Ackley function, it also

GECCO '18 Companion, July 15–19, 2018, Kyoto, Japan

ACM ISBN 978-1-4503-5764-7/18/07…$15.00
https://doi.org/10.1145/3205651.3205801

runs faster 1.3 times in dimension 10, 4 times in dimension 20, and 16.5 times in dimension 40. This is because whenever the global solution is located, CMAES-APOP gradually reduces the pop-size and profits from the robustness of CMA-ES to find it.

Function	n	SR	Feval	Feval/SR	aRT	IPOP-CMAES aRT
		CMAES-APOP				
Rastrigin	10	0.86	2.797e+04	3.242e+04	3.317e+04	7.471e+04
	20	0.94	8.491e+04	9.022e+04	9.077e+04	2.594e+05
	40	1.00	2.981e+05	2.981e+05	2.981e+05	7.744e+05
Scale Rastrigin	10	0.80	2.758e+04	3.431e+04	3.527e+04	8.498e+04
	20	0.80	8.994e+04	1.118e+05	1.111e+05	2.794e+05
	40	0.90	3.056e+05	3.388e+05	3.427e+05	9.179e+05
Schaffer	10	0.96	2.945e+04	3.065e+04	3.098e+04	4.801e+04
	20	0.94	7.660e+04	8.138e+04	8.175e+04	1.285e+05
	40	0.90	2.016e+05	2.235e+05	2.255e+05	3.206e+05
Griewank	10	0.98	1.183e+04	1.206e+04	1.215e+04	7.192e+03
	20	0.98	2.419e+04	2.468e+04	2.479e+04	7.170e+03
	40	1.00	5.769e+04	5.769e+04	5.769e+04	1.186e+04
Ackley	10	1.00	1.403e+04	1.403e+04	1.403e+04	1.890e+04
	20	0.98	3.055e+04	3.116e+04	3.105e+04	1.249e+05
	40	0.92	6.756e+04	7.331e+04	7.204e+04	1.162e+06
Bohachevsky	10	1.00	1.002e+04	1.002e+04	1.002e+04	5.947e+03
	20	1.00	2.397e+04	2.397e+04	2.397e+04	1.813e+04
	40	0.98	5.424e+04	5.533e+04	5.536e+04	4.537e+04

Table 3: CMAES-APOP vs IPOP-CMAES (n: dimension, SR: success rate, Feval: number of functions evaluations, aRT (average Running Time) = number of function evaluations divided by the number of successful trials).

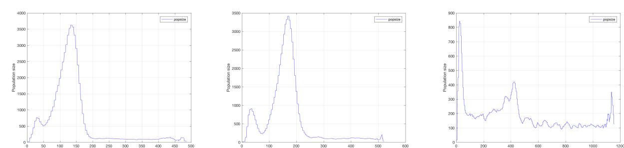

Figure 1: Adapting population size of CMAES-APOP in 40-D (Left: Rastrigin function, Center: Scale Rastrigin function, Right: Schaffer function).

Function	n	25%	10%	50%	75%	90%
Rastrigin	10	**3.317e+04**	3.527e+04	3.160e+04	3.069e+04	3.250e+04
	20	**9.077e+04**	9.254e+04	9.212e+04	9.038e+04	9.286e+04
	40	**2.981e+05**	3.163e+05	3.006e+05	3.034e+05	3.133e+05
Scale Rastrigin	10	**3.527e+04**	4.044e+04	3.638e+04	3.609e+04	3.818e+04
	20	**1.111e+05**	1.040e+05	1.141e+05	1.053e+05	1.083e+05
	40	**3.427e+05**	3.486e+05	3.579e+05	3.367e+05	3.594e+05
Schaffer	10	**3.098e+04**	3.334e+04	3.051e+04	3.012e+04	3.147e+04
	20	**8.175e+04**	8.833e+04	8.024e+04	8.233e+04	8.646e+04
	40	**2.255e+05**	2.266e+05	2.224e+05	2.348e+05	2.325e+05
Griewank	10	**1.215e+04**	1.298e+04	1.198e+04	1.480e+04	1.740e+04
	20	**2.479e+04**	2.601e+04	2.498e+04	2.569e+04	2.891e+04
	40	**5.769e+04**	5.948e+04	5.614e+04	5.843e+04	6.327e+04
Ackley	10	**1.403e+04**	1.481e+04	1.369e+04	1.429e+04	1.498e+04
	20	**3.105e+04**	3.263e+04	3.024e+04	3.144e+04	3.326e+04
	40	**7.204e+04**	7.379e+04	6.761e+04	7.164e+04	7.617e+04
Bohachevsky	10	**1.002e+04**	1.052e+04	1.015e+04	1.064e+04	1.085e+04
	20	**2.397e+04**	2.533e+04	2.366e+04	2.378e+04	2.494e+04
	40	**5.536e+04**	5.781e+04	5.627e+04	5.810e+04	6.101e+04

Table 4: The aRT of some variants of CMAES-APOP.

Figures 1 shows the variation of pop-size in iterations (in average over successful runs) on the Rastrigin, Scale Rastrigin, and Schaffer functions in the 40-D. For the Rastrigin function, the pop-size increases to about 3600 in 130 iterations to locate the optimal solution. Then it decreases to about 80-90 at iteration 190. Since then the pop-size becomes quite stable until the convergence. On the Scale Rastrigin function, the variation of pop-size is more complicated than on the Rastrigin function. Firstly, the pop-size increases to 900 after 40 iterations, then decreases until iteration 80. This is because the effect of scale operator, and the algorithm does not learn so much about this operator. But when re-detecting the ruggedness of function, the algorithm one more time has to increase pop-size to escape the local attraction zone to find out the global one. After that, the pop-size adaptation process is quite similar to on the normal Rastrigin function. On the Schaffer function, the pop-size must be adapted many times to approach the global solution. It is not stable during the searching process, even when the algorithm moves the distribution close to global solution, the adaptation is still going on.

In the CMAES-APOP we compare $f^{\text{med}}_{\text{prev}}$ with $f^{\text{med}}_{\text{cur}}$ to consider a "going up" time. This quantity f^{med} could be seen as 25th percentile of all objective values (denoted by $f^{25\%}$) in one iteration. Table 4 shows performance of CMAES-APOP when we replace the 25th percentile of objective function values with the other percentiles. We can see that for these functions, the aRT of CMAES-APOP does not change so much. It is because these considered functions are not too complicated or are well-structured.

2.2 On the Schwefel function

For the hard Schwefel function, tracking the change of $f^{25\%}$ to adapt the pop-size seems not to be useful. Therefore, we propose a relaxed version of CMAES-APOP in which we try to change the condition "$f^{25\%}_{\text{cur}} - f^{25\%}_{\text{prev}} > 0$" to the condition "$f^{p}_{\text{cur}} - f^{25\%}_{\text{prev}} > 0$", where p is 30% or 50% to relax the condition for a "going up" time. This modification still keeps the relaxed version invariant to scaling and shifting operator on the objective function. Additionally, we set an upper bound on pop-size, say $\lambda_{\max} = (20n + 30)\lambda_{\text{default}}$, to the relaxed version. A version of CMAES-APOP using this upper bound for pop-size is also tested. Note that the step-size σ probably increases although the upper bound for pop-size is set.

n	CMAES-APOP				Relaxed-CMAES-APOP				IPOP-CMAES
	no UB		UB		p = 30%		p = 50%		
	SR	aRT	SR	aRT	SR	aRT	SR	aRT	aRT
5	0.43	5.97e+04	0.41	5.68e+04	0.60	1.01e+05	0.78	1.01e+05	8.49e+04
10	0.13	5.68e+05	0.11	7.52e+05	0.80	4.14e+05	0.90	4.24e+05	5.55e+05
20	0.00	NA	0.00	NA	0.74	2.59e+06	0.90	2.30e+06	3.55e+06

Table 5: On the Schwefel function.

From table 5 we see that CMAES-APOP with/without upper bound for population size has low success rates, can not reach the target in dimension 20, and gives large aRT in dimensions 5 and 10. However, the relaxed versions show quite good performance. Especially, the version with p = 50% provides a high success rate even in dimension 20, and runs faster than IPOP-CMAES does about 1.3 times in dimension 10, and about 1.5 times in dimension 20.

Acknowledgments. This work was completed while the author was visiting Vietnam Institute for Advanced Study in Mathematics (VIASM). He would like to thank the Institute for the support.

REFERENCES

[1] 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.

[2] N. Hansen. 2009. Benchmarking a BI-population CMA-ES on the BBOB-2009 Function Testbed. In *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers (GECCO '09)*. 2389–2396.

[3] N. Hansen and S. Kern. 2004. *Evaluating the CMA Evolution Strategy on Multimodal Test Functions.* 282–291 pages. https://doi.org/10.1007/978-3-540-30217-9_29

[4] D. M. Nguyen and N. Hansen. 2017. Benchmarking CMAES-APOP on the BBOB Noiseless Testbed. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '17)*. ACM, New York, NY, USA, 1756–1763.