Benchmarking a Variant of CMAES-APOP on the BBOB Noiseless Testbed

Duc Manh Nguyen[1,2]

[1]Hanoi National University of Education, Vietnam
[2]Sorbonne Université, IRD, JEAI WARM, Unité de Modélisation Mathématiques et Informatique des Systèmes Complexes, UMMISCO, F-93143, Bondy, France

The Genetic and Evolutionary Computation Conference
Kyoto, July 15-19, 2018

Outline

- The CMAES-APOP algorithm
- A Variant of CMAES-APOP algorithm
- Numerical Experiments on the BBOB Noiseless Testbed
- Conclusion and Perspectives

The CMAES-APOP algorithm

- Adapting population size seems to be a right way in the CMA-ES to optimize multi-modal functions.
- Some approaches for adapting population size in the CMAES:
  - IPOP-CMA-ES [1] [AH05, Ros10]: the CMA-ES is restarted with increasing population size by a factor of two whenever one of the stopping criteria is met.
  - BIPOP-CMA-ES [2]: define two restart regimes: one with large populations (IPOP part), and another one with small populations. In each restart, BIPOP-CMA-ES selects the restart regime with less function evaluations used so far.

---

[1] [AH05] A. Auger and N. Hansen, A restart cma evolution strategy with increasing population size, 2005 IEEE Congress on Evolutionary Computation, vol. 2, 2005, pp. 1769-1776.

[2] [Han09] N. Hansen, Benchmarking a bi-population cma-es on the bbob-2009 function testbed, Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, GECCO 09, 2009, pp. 2389-2396.

The CMAES-APOP algorithm

- Ahrari and Shariat-Panahi [3]: An adaptation strategy for the CMA-ES which used the oscillation of objective value of x_{mean} to quantify multimodality of the region under exploration.
- Nishida and Akimoto [4]: An adaptation strategy for the CMA-ES that is based on the estimation accuracy of the natural gradient.

---

[3] [ASP15] A. Ahrari and M. Shariat-Panahi, An improved evolution strategy with adaptive population size, Optimization 64 (2015), no. 12, 2567-2586.

[4] [NA16] K. Nishida and Y. Akimoto, Population size adaptation for the cma-es based on the estimation accuracy of the natural gradient, Proceedings of the Genetic and Evolutionary Computation Conference 2016, GECCO 16, 2016, pp. 237-244

The CMAES-APOP Algorithm [5]

Motivation

- a natural desire when solving any optimization problem
- one prospect when using larger population size to search

  "We want to see the decrease of objective function"

Signal?

- We track the non-decrease of objective function (exactly, $f^{\text{med}} := \text{median}(f(\mathbf{x}_{i:\lambda}), i = 1, ..., \mu)$ - the median of objective function of μ elite solutions in each iteration) in a slot of S successive iterations to adapt the population size in the next S successive iterations
- We do not adapt the population size in each iteration but in each slot of S iterations.

  $\Rightarrow$ The variation of population size takes a staircase form in iterations.

[5][NH17] D. M. Nguyen and N. Hansen, Benchmarking cmaes-apop on the bbob noiseless testbed, Proceedings of the Genetic and Evolutionary Computation Conference Companion (New York, NY, USA), GECCO 17, ACM, 2017, pp. 1756-1763.

A Variant of CMAES-APOP Algorithm

Ideas: $f^{\text{med}} := \text{median}(f(\mathbf{x}_{i:\lambda}), i = 1, ..., \mu)$ is the 25th percentile of objective function values evaluated on λ candidate points.
$\Rightarrow$ What if we change the 25th percentile to the other percentiles?

Some test functions:

$$
\begin{aligned}
f_{\text{Rastrigin}}(\boldsymbol{x}) &= 10n + \sum_{i=1}^{n}(x_i^2 - 10\cos(2\pi x_i)) \\
f_{\text{Schaffer}}(\boldsymbol{x}) &= \sum_{i=1}^{n-1}(x_i^2 + x_{i+1}^2)^{0.25}[\sin^2(50(x_i^2 + x_{i+1}^2)^{0.1}) + 1] \\
f_{\text{Ackley}}(\boldsymbol{x}) &= 20 - 20 \cdot \exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\right) + \mathrm{e} - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) \\
f_{\text{Bohachevsky}}(\boldsymbol{x}) &= \sum_{i=1}^{n-1}(x_i^2 + 2x_{i+1}^2 - 0.3\cos(3\pi x_i) - 0.4\cos(4\pi x_{i+1}) + 0.7)
\end{aligned}
$$

- For each function, 51 runs are conducted.
- $f_{\text{stop}} = 10^{-10}$ ($f_{\text{stop}} = 10^{-8}$ for the Schaffer function).
- the starting point for the functions Rastrigin, Schaffer, Ackley, Bohachevsky is $(5, ..., 5), (55, ..., 55), (15, ..., 15)$, and $(8, ..., 8)$ respectively; the initial step-size σ for these functions is 2, 20, 5, 3 respectively.

A Variant of CMAES-APOP Algorithm

We run the CMAES-APOP algorithm with the small initial population size $\lambda = \lambda_{\text{default}}$ (i.e, set $k_n = 1$) and without the upper bound for the population size in three dimensions $n = 10, 20, 40$.

Function	n	**25-p**	1-p	10-p	50-p	75-p	90-p
	10	**3.317e+04**	4.332e+04	3.527e+04	3.160e+04	3.069e+04	3.250e+04
Rastrigin	20	**9.077e+04**	1.189e+05	9.254e+04	9.212e+04	9.038e+04	9.286e+04
	40	**2.981e+05**	3.992e+05	3.163e+05	3.006e+05	3.034e+05	3.133e+05
	10	**3.098e+04**	5.111e+04	3.334e+04	3.051e+04	3.012e+04	3.147e+04
Schaffer	20	**8.175e+04**	1.663e+05	8.833e+04	8.024e+04	8.233e+04	8.646e+04
	40	**2.255e+05**	4.942e+05	2.266e+05	2.224e+05	2.348e+05	2.325e+05
	10	**1.403e+04**	2.280e+04	1.481e+04	1.369e+04	1.429e+04	1.498e+04
Ackley	20	**3.105e+04**	6.125e+04	3.263e+04	3.024e+04	3.144e+04	3.326e+04
	40	**7.204e+04**	1.275e+05	7.379e+04	6.761e+04	7.164e+04	7.617e+04
	10	**1.002e+04**	1.494e+04	1.052e+04	1.015e+04	1.064e+04	1.085e+04
Bohachevsky	20	**2.397e+04**	4.261e+04	2.533e+04	2.366e+04	2.378e+04	2.494e+04
	40	**5.536e+04**	9.881e+04	5.781e+04	5.627e+04	5.810e+04	6.101e+04

Table: The aRT of some variants of CMAES-APOP: the 25-percentile is replaced by the other percentiles (**aRT** (average Running Time) = number of function evaluations divided by the number of successful trials)

A Variant of CMAES-APOP Algorithm

Some notations:

- P : a set of percentiles.
- $f^p := \text{percentile}(\{f(\mathbf{x}_{i:\lambda}), i = 1, ..., \lambda\}, p)$: the p-percentile of objective function of λ candidates in each iteration, where p can vary from 0 to 100 (in fact p will be chosen from the set of percentiles P); f^p_{prev} and f^p_{cur} denote the p-percentiles in the previous and current iteration respectively.
- n_{up} : the number of times "$f^p_{\text{cur}} - f^p_{\text{prev}} > 0$" occurs during a slot of S iterations.
- t_{up} : the history of n_{up} in each slot recorded.
- no_{up} : the number of most recent slots we do not see the non-decrease.
- $\lambda_{\max} := (20n + 30)\lambda_{\text{default}}$: the maximum number of the population, where $\lambda_{\text{default}} = \lfloor 4 + 3\log(n) \rfloor$.

A Variant of CMAES-APOP Algorithm

Input: $\mathbf{m} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+$

Initialize: $\mathbf{C} = \mathbf{I}, \mathbf{p}_c = 0, \mathbf{p}_\sigma = 0, \lambda = k_n \times \lambda_{\text{default}}$

Set: $\mu = \lfloor \lambda/2 \rfloor, w_i, \mu_w, c_c, c_\sigma, c_1, c_\mu, d_\sigma$, iter $= 0$, $S = 5$, $r_{\max} = 30$, $n_{\text{up}} = 0$, $t_{\text{up}} = [\ ]$.

While not terminate

    iter $=$ iter $+ 1$;

    $\mathbf{x_i} = \mathbf{m} + \sigma \mathbf{y_i}, \mathbf{y_i} \sim \mathbf{N}(\mathbf{0}, \mathbf{C})$, for $i = 1, \ldots, \lambda$

    Take p randomly from the set of percentiles P

    if iter > 1

        if $f^p_{\text{cur}} - f^p_{\text{prev}} > 0$ //Check if f^p increases

            $n_{\text{up}} = n_{\text{up}} + 1$;

        end

    end

    Update $\mathbf{m}, \mathbf{p}_c, \mathbf{p}_\sigma, \mathbf{C}, \sigma$ as in the CMA-ES

    if (mod(iter, S) $= 1$) & (iter > 1) // Adapting the population size

        $t_{\text{up}} = [t_{\text{up}}; n_{\text{up}}]$;

        Adapt the population size according to the information of n_{up} (... details in the next slide)

        $n_{\text{up}} \leftarrow 0$ // Reset n_{up} back to 0

    end

A Variant of CMAES-APOP Algorithm

(16) Adapt the population size according to the information of n_{up}

16.1 **if** $n_{up} > 1$

16.2 $\quad \lambda \leftarrow \left\lfloor \min\left(\exp\left(\frac{n_{up} \cdot (4 + 3\log(n))}{S \cdot \sqrt{\lambda - \lambda_{default} + 1}} \right), r_{max} \right) \times \lambda \right\rfloor;$

16.3 $\quad \lambda \leftarrow \min(\lambda, \lambda_{max});$

16.4 $\quad \sigma \leftarrow \sigma \times \exp\left(\frac{1}{n}\left(\frac{n_{up}}{S} - \frac{1}{5}\right)\right);$ // Enlarge σ a little bit

16.5 **elseif** $n_{up} = 0$

16.6 $\quad no_{up} = length(t_{up}) - \max(find(t_{up} > 0));$

16.7 $\quad$ **if** $\lambda > 2\lambda_{default}$

16.8 $\quad\quad \lambda \leftarrow \max(\lfloor \lambda \times \exp(-no_{up}/10) \rfloor, 2\lambda_{default});$

16.9 $\quad$ **end**

16.10 **end**

16.11 **if** λ is changed // Only when $n_{up} > 1$ or $n_{up} = 0$

16.12 $\quad$ Update $\mu, w_{i=1\ldots\mu}, \mu_w$ w.r.t the new population size λ

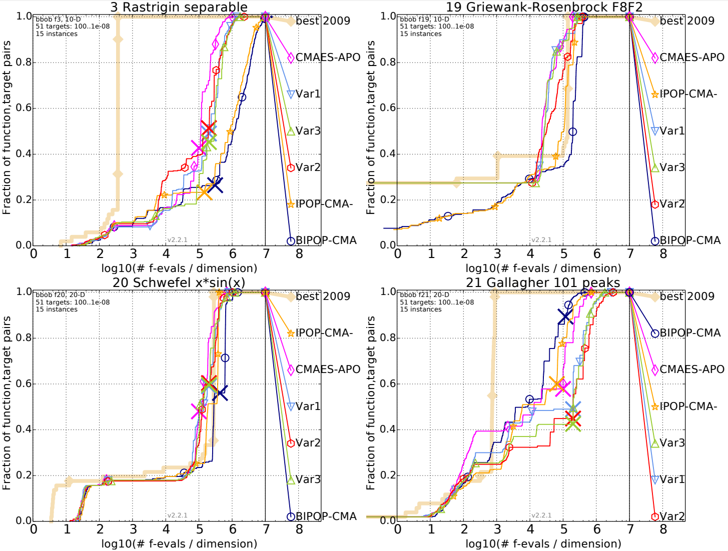
16.13 $\quad$ Update the parameters $c_c, c_\sigma, c_1, c_\mu, d_\sigma$

16.14 **end**

Numerical Experiments on the BBOB Noiseless Testbed

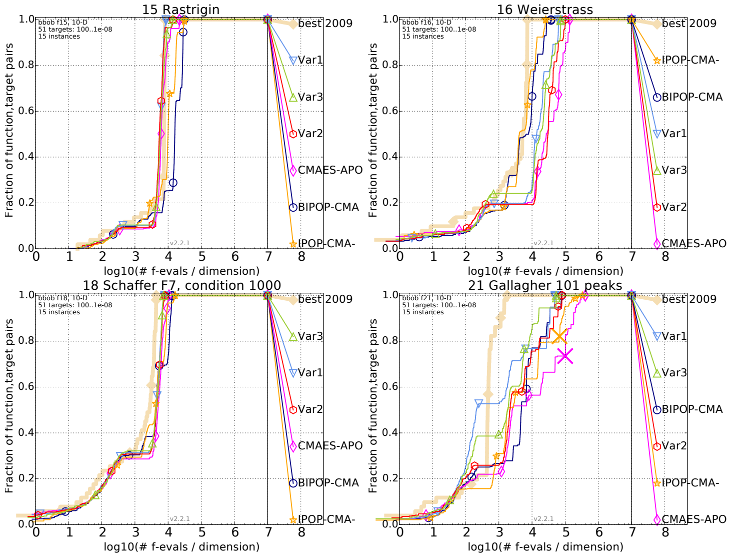
- Test the algorithms with a budget of $2 \times 10^5 \times n$, where n is the problem dimension.
- Denote the variants corresponding to $P_1 = \{1, 25, 50\}$, $P_2 = \{1, 50\}$, and $P_3 = \{1, 50, 75\}$ by Var1, Var2 and Var3 respectively.
- In the first run: the pure CMA-ES with the default population size $\lambda = \lambda_{\text{default}}$. From second run: the pop-size adaptation strategy is applied with the initial population size $\lambda = k_n \times \lambda_{\text{default}}$.
- The parameter k_n is set to 10, 20, 30, 40, 50, 60 for $n = 2, 3, 5, 10, 20, 40$ respectively.
- Take the starting point $\mathbf{m}^0$ uniformly in $[-4, 4]^n$.
- Set the initial step-size $\sigma_0 = 2$ for all run.

The variants < CMAES-APOP

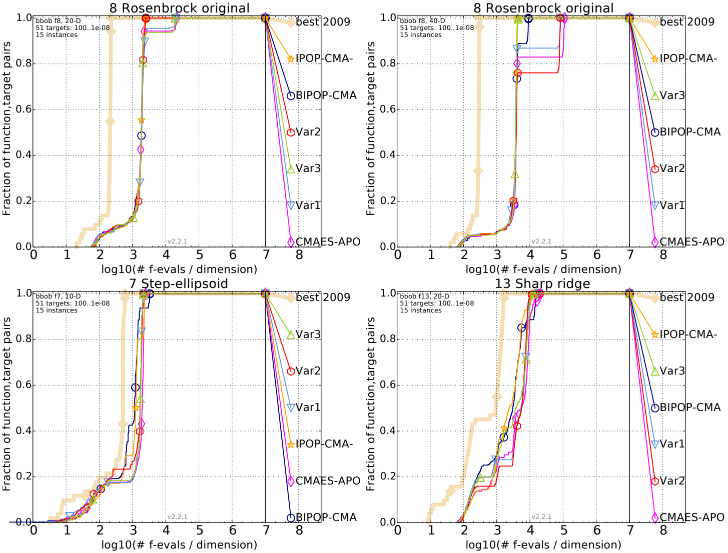


All variants are still better than the IPOP-CMA-ES and BIPOP-CMA-ES on f_3 in 10-D; than the BIPOP-CMA-ES on f_{19} in dimensions 10; and than the BIPOP-CMA-ES on f_{20} in dimensions 20.

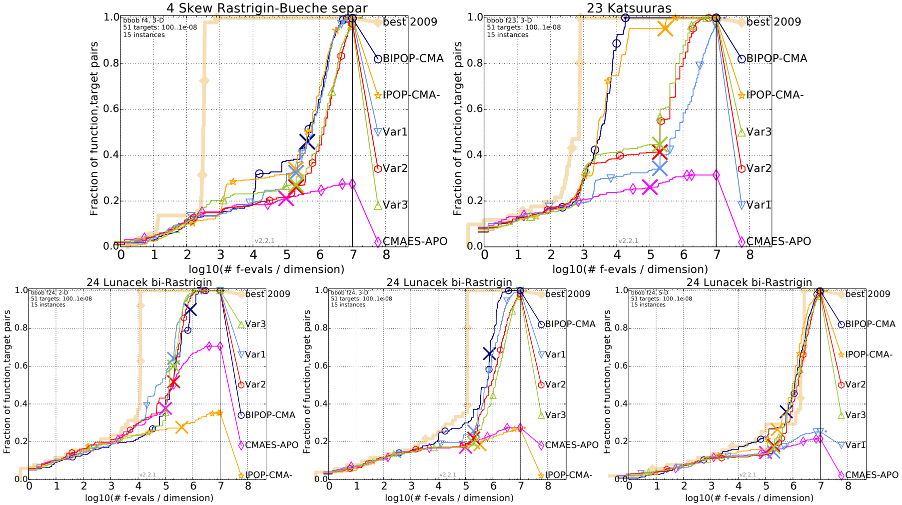
The variants $>_{\text{slightly}}$ CMAES-APOP: f_{15}, f_{16}, f_{18}, f_{21} in 10-D



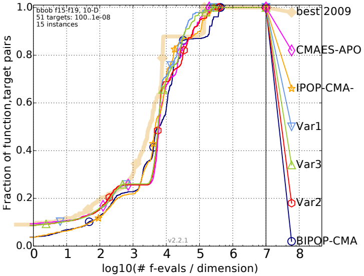
The variants $>_{\text{slightly}}$ CMAES-APOP: f_7, f_8, f_{13}

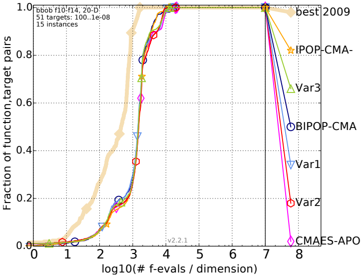


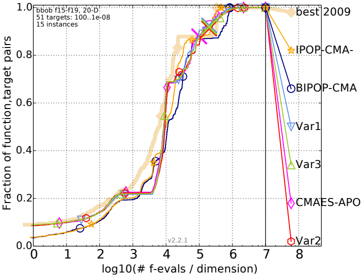
The variants > CMAES-APOP: on f_4, f_{23}, f_{24} in small dimensions



(Var1 ($\{1, 25, 50\}$) **& Var3** ($\{1, 50, 75\}$)) $>_{\text{slightly}}$ **Var2** ($\{1, 50\}$)

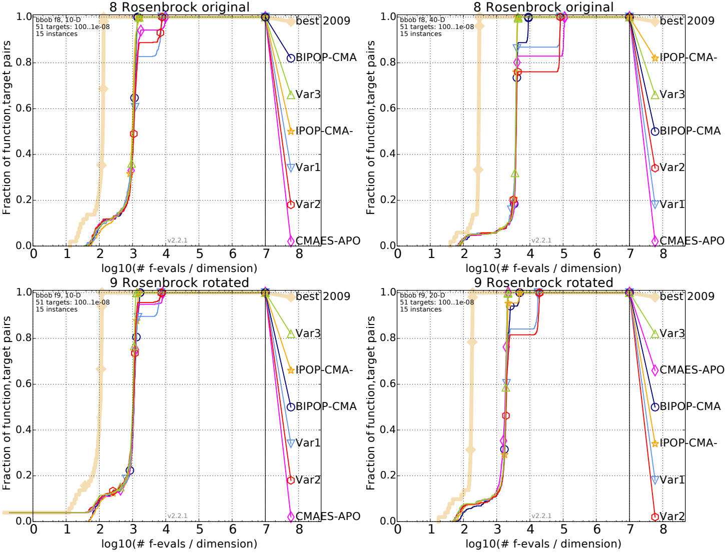




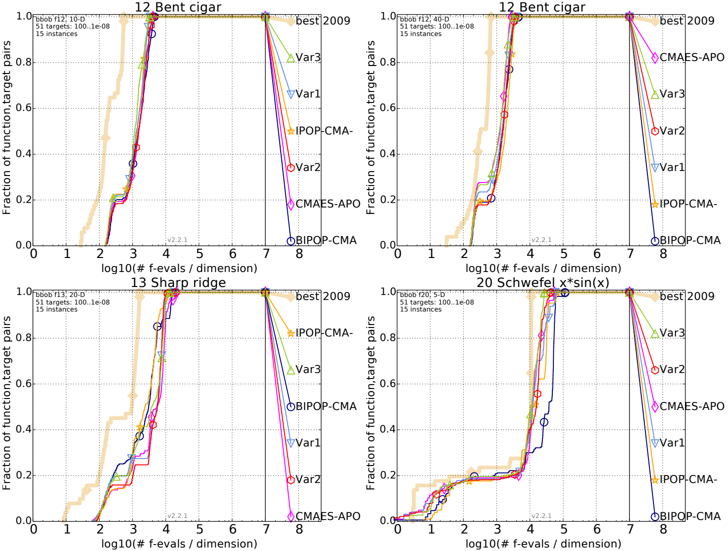


$\Rightarrow$ Tracking more percentiles can help us to make better decisions in adapting population size for the class of conditioned functions, and the class of multi-modal functions with adequate global structure in high dimensions.

Var3 ($\{1, 50, 75\}$) $>_{\text{slightly}}$ ((Var1 ($\{1, 25, 50\}$) & Var2 ($\{1, 50\}$)(1/2)



Var3 $(\{1, 50, 75\}) >_{\text{slightly}}$ ((**Var1** $(\{1, 25, 50\})$ & **Var2** $(\{1, 50\})(2/2)$



⇒ The information of non-elite individuals is also useful to adapt the population size.

Conclusion and Perspectives

Conclusion:

- Present a variant of CMAES-APOP: track the change of some percentiles of objective values rather than one percentile; set the upper bound of the population size depending on the problem dimension.
- This approach improves the performance of CMAES-APOP in some cases when the set of percentiles P is chosen appropriately.

Perspectives:

- How to initialize a good set P and how to evaluate the importance of each percentile p in P during the evolution process?
- The information of percentiles could play a deeper role inside the evolution process of the CMA-ES?

References

- A. Auger and N. Hansen, *A restart cma evolution strategy with increasing population size*, 2005 IEEE Congress on Evolutionary Computation, vol. 2, 2005, pp. 1769–1776 Vol. 2.
- A. Ahrari and M. Shariat-Panahi, *An improved evolution strategy with adaptive population size*, Optimization **64** (2015), no. 12, 2567–2586.
- N. Hansen, *Benchmarking a bi-population cma-es on the bbob-2009 function testbed*, Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, GECCO '09, 2009, pp. 2389–2396.
- K. Nishida and Y. Akimoto, *Population size adaptation for the cma-es based on the estimation accuracy of the natural gradient*, Proceedings of the Genetic and Evolutionary Computation Conference 2016, GECCO '16, 2016, pp. 237–244.
- D. M. Nguyen and N. Hansen, *Benchmarking cmaes-apop on the bbob noiseless testbed*, Proceedings of the Genetic and Evolutionary Computation Conference Companion (New York, NY, USA), GECCO '17, ACM, 2017, pp. 1756–1763.
- R. Ros, *Black-box optimization benchmarking the IPOP-CMA-ES on the noiseless testbed: comparison to the BIPOP-CMA-ES*, GECCO '10: Proceedings of the 12th annual conference comp on Genetic and evolutionary computation (New York, NY, USA), ACM, 2010, pp. 1503–1510.

Thank you for your attention!