A Note on the CMA-ES for Functions with Periodic Variables

Takahiro Yamaguchi
Shinshu University
Nagano, Japan
17w2097h@shinshu-u.ac.jp

Youhei Akimoto
University of Tsukuba
Tsukuba, Japan
akimoto@cs.tsukuba.ac.jp

ABSTRACT

In this short paper, we reveal the issue of the covariance matrix adaptation evolution strategy when solving a function with periodic variables. We investigate the effect of a simple modification that the coordinate-wise standard deviation of the sampling distribution is restricts to the one-fourth of the period length. This is achieved by pre- and post-multiplying a diagonal matrix to the covariance matrix.

CCS CONCEPTS

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

KEYWORDS

CMA-ES, periodic variables, mirroring

ACM Reference Format:
Takahiro Yamaguchi and Youhei Akimoto. 2018. A Note on the CMA-ES for Functions with Periodic Variables. In *GECCO '18 Companion: Genetic and Evolutionary Computation Conference Companion, July 15–19, 2018, Kyoto, Japan,* Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3205651.3205669

1 INTRODUCTION

The covariance matrix adaptation evolution strategy (CMA-ES) [2–4] is one of the state-of-the-art search algorithm for black-box continuous optimization problems. Sometimes we face a periodic function, where the function topography is repeated in some dimension. A source of the periodic dimension is to have periodic variables such as angle [1]. Another source is to use the box constraint handling based on the mirroring, i.e., the objective function is extended to the outside the feasible domain by $f(\boldsymbol{x}) = f(2 \cdot \ell - \boldsymbol{x})$ and $f(\boldsymbol{x}) = f(2 \cdot u - \boldsymbol{x})$. Then, the variable will be periodic with the period $[\ell - (u - \ell)/2, \ell + (u - \ell)/2]$. When the CMA-ES is applied to a function with periodic topography, we sometimes observe an undesired behavior. The sampling distribution will become so large that the candidate solutions are generated over several periods, and failed to grasp the function landscape. In this short paper, we show the problem of the CMA-ES on a periodic function in a simple test function, and provide a simple device to prevent the problem.

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

ACM ISBN 978-1-4503-5764-7/18/07.
https://doi.org/10.1145/3205651.3205669

2 ISSUE ON A PERIODIC FUNCTION

To visualize the problem of the CMA-ES on a periodic function, we create the following simple test function,

$$f(\boldsymbol{x}) = \boldsymbol{y}^{\mathrm{T}} \mathbf{A} \boldsymbol{y} \ , \quad y_i := x_i - 2 \left\lfloor \frac{x_i + 1}{2} \right\rfloor \ , \tag{1}$$

where x_i and y_i denote the ith coordinate of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, and $\mathbf{A}$ is a symmetric positive definite matrix. Its minimal value is zero, and is located at $x_i^* = 2j$ for any integer j, and $\lfloor a \rfloor$ denotes the maximum integer that is no greater than $a \in \mathbb{R}$. Its period is $[-1, 1]$ for each variable. It has a unique local optimum at the origin within a single period. The Hessian matrix of this function is $2\mathbf{A}$ everywhere in $(-1, 1)^N$. In this paper, we set $N = 10$. We use the standard CMA-ES described in [2]. We start each run with initial mean vector $\boldsymbol{m}^{(0)} = \mathbf{0}$, initial step-size $\sigma^{(0)} = 1$, and initial covariance matrix $\mathbf{C}^{(0)} = \mathbf{I}$.

Figure 1 shows typical successful and unsuccessful trials of the CMA-ES. In the successful trial, although there was a slight stagnation in the early stage, it found the optimum. In the unsuccessful trial, it failed to properly update the probability distribution. The standard deviation of the distribution became too much greater than the period of the test function (see Figure 1b). Then, the function landscape looks like a plateau with noise, hence the CMA-ES fails to find an appropriate updating direction. This problem was observed if we deactivate either the covariance matrix adaptation or the step-size adaptation, implying that both adaptation mechanisms potentially cause the same problem.

3 BOUNDING COORDINATE-WISE STD.

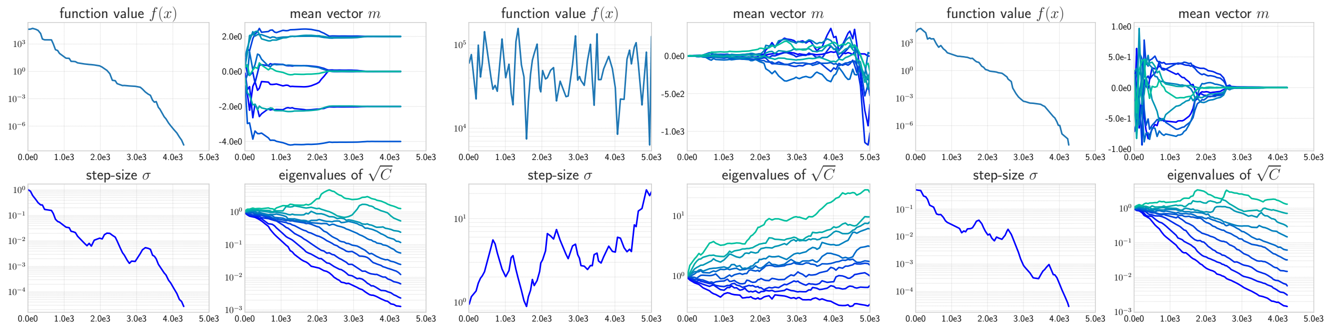
A simple yet effective modification for this problem is to upper bound the coordinate-wise standard deviation of the sampling distribution, i.e., $\sigma \mathbf{C}_{i,i}^{1/2}$. Let r_i be the length of the period for ith coordinate if it is a periodic variable, $+\infty$ otherwise. To guarantee the symmetry and the positive definiteness of $\mathbf{C}$, we correct the covariance matrix as follows

$$\mathbf{C} \leftarrow \mathbf{DCD} \tag{2}$$

where $\mathbf{D}$ is a diagonal matrix whose ith diagonal element is

$$\mathbf{D}_{i,i} = \min \left\{ \frac{r_i}{4\sigma \mathbf{C}_{i,i}^{1/2}}, 1 \right\} \ . \tag{3}$$

This operation forces the coordinate-wise standard deviation to be at greatest $r_i/4$, i.e., each diagonal element of the covariance matrix of the sampling distribution, $\sigma^2 \mathbf{C}$, is upper bounded by $(r_i/4)^2$. The reason of $r_i/4$ is based on the fact that the probability that a normal random variable with the standard deviation $r_i/4$ deviate from the mean value by more than $r_i/2$ is less than 0.05. In other words, approximately 95% of coordinate values of candidate solutions will be produced in one period.



(a) Successful Run **(b) Unsuccessful Run** **(c) Proposed**

Figure 1: Typical successful and unsuccessful runs of the CMA-ES and a typical run of the proposed one on the test function.

Another approach is to shrink σ instead of the coordinate-wise std, as suggested in [1]. Shrinking σ results in shrinking the std. of axes that do not exceed the maximum value. The global search performance of the CMA-ES will be affected and there is a possibility that it will be easily trapped by a local optimum. On the other hand, repairing the coordinate-wise std. results in changing the distribution shape, which may affect the performance on non-separable functions. As shown in Figure 1b, however, the situation requiring the coordinate-wise std. repair is often at the early stage of optimization where the CMA-ES has not learn the shape of the distribution. Therefore, we employ the coordinate-wise std. modification (2).

4 EXPERIMENT

The test function is optimized by the CMA-ES with/without the coordinate-wise std. correction. The same setting has been used as in Section 2, except that the initial step-size is set to 0.5 (quarter period) and 1.0 (half period), the latter exceeds the maximum std. value and repaired immediately in the proposed algorithm, hence is identical to the former. Each run is stooped after 10^4 f-calls as failure, or stopped as success if the function value reaches 10^{-8}. Thirty independent trials have been conducted for each settings.

Table 1: Average f-calls over successful runs and success probability.

Optimizer	Average f-calls	Success Prob.
CMA-ES ($\sigma = 1.0$)	4547.5	8/30
CMA-ES ($\sigma = 0.5$)	4154.4	27/30
CMA-ES with correction	4164.7	30/30

Table 1 shows the summary of the results. The success probability was improved when we set a greater $\sigma^{(0)}$ in the CMA-ES without std. correction. However, even if $\sigma^{(0)} = 0.5$, there were unsuccessful trials, where the std. increased beyond the period length and took time to shrink them. Therefore, this problem may not be fully solved by adjusting the initial condition. On the contrary, a too small initial step-size will lead to a sub-optimal solution when the function is rugged. On the other hand, the CMA-ES with std. correction succeeded in all trials. Figure 1c shows the transition of the parameters of in the CMA-ES with std. correction.

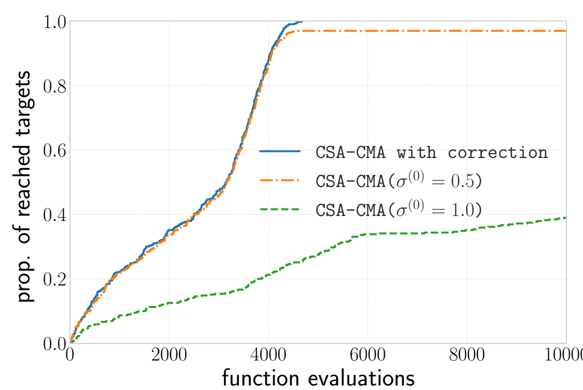


Figure 2: Proportion of reached target levels over 30 runs. Total number of target levels 360: twelve target levels $(10^3, 10^2, \cdots, 10^{-8})$ **for each run times 30 runs.**

Figure 2 shows the proportion of reached targets over 30 independent runs versus the number of f-calls. As Table 1 indicates, three variants converged at more or less the same speed when they succeeded. However, unsuccessful trials took long time (more than double the f-calls in successful trials) until it started converging. The proposed modification is very simple, yet effective in preventing undesired behavior of the CMA-ES.

REFERENCES

[1] Dirk V Arnold. 2014. On the Use of Evolution Strategies for Optimization on Spherical Manifolds. In *International Conference on Parallel Problem Solving from Nature*. Springer, 882–891.
[2] N. Hansen. 2016. The CMA Evolution Strategy: A Tutorial. *ArXiv e-prints* (April 2016). arXiv:cs.LG/1604.00772
[3] Nikolaus Hansen, Sibylle D. Muller, and Petros Koumoutsakos. 2003. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation* 11, 1 (2003), 1–18.
[4] Nikolaus Hansen and Andreas Ostermeier. 2001. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation* 9, 2 (2001), 159–195.