CMA-ES with Coordinate Selection for High-Dimensional and Ill-Conditioned Functions

Hiroki Shimizu
The University of Tokyo
shimizu@tkl.iis.u-tokyo.ac.jp

Masashi Toyoda
Institute of Industrial Science, the University of Tokyo
toyoda@tkl.iis.u-tokyo.ac.jp

ABSTRACT

Algorithms for black-box optimization need considering numerous properties of objective functions in advance. The covariance matrix adaptation evolution strategy (CMA-ES) is known as one of the state-of-the-art algorithms for black-box optimization. Despite its achievement, the CMA-ES fails to minimize the objective function which is high-dimensional and ill-conditioned such as 100,000-dimensional Ellipsoid function. This fact is a serious problem to apply the CMA-ES to recent high-dimensional machine learning models. We confirm that the "single" step-size for all coordinates is one of the hindrances to the adaptation of the variance-covariance matrix. To solve this, we propose a CMA-ES with coordinate selection. Coordinate selection enables us to vectorize the step-size and adapt each component of the vector to the scale of selected coordinates. Furthermore, coordinate selection based on estimated curvature reduces the condition number during updating variables in selected coordinate space. Our method is enough simple to easily apply to most of variations of CMA-ES: only execute conventional algorithms in the selected coordinate space. The experimental results show that our method applied to the CMA-ES, the sep-CMA-ES and the VD-CMA outperforms the conventional variations of CMA-ES in terms of function evaluations and an objective value in the optimization of high-dimensional and ill-conditioned functions.

CCS CONCEPTS

• **Theory of computation** → *Design and analysis of algorithms*;

KEYWORDS

CMA-ES, ill-conditioned function, high dimensional function

ACM Reference Format:
Hiroki Shimizu and Masashi Toyoda. 2021. CMA-ES with Coordinate Selection for High-Dimensional and Ill-Conditioned Functions. In *2021 Genetic and Evolutionary Computation Conference Companion (GECCO '21 Companion), July 10–14, 2021, Lille, France.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3449726.3459575

1 INTRODUCTION

Handling high-dimensional problems is challenging even for the covariance matrix adaptation evolution strategy (CMA-ES) [2], which is a state-of-the-art algorithm for black-box optimization, because of its time and space complexities. To deal with this, the sep-CMA-ES [5] limits a covariance matrix to the diagonal components and reduces the complexities. However, Loshchilov [4] pointed out that the sep-CMA-ES did not sufficiently decrease the value of the $100,000$-dimensional ellipsoid function even after 10^6 function evaluations.

Therefore, we propose a CMA-ES with coordinate selection to reduce the condition number and adapt the vectorized step-size to the scale of each coordinate of an objective function. We used our algorithm for several variations of CMA-ES and determined that our method is superior to a conventional CMA-ES in terms of function evaluations that reach optimal value for high-dimensional and ill-conditioned functions.

2 CMA-ES WITH COORDINATE SELECTION

The condition number, which indicates the ill-conditionality of an objective function, is determined by a ratio of max. and min. eigenvalues of the Hessian of the objective function. Our method only updates the parameters in the selected coordinate space at each generation and regards the variables in the coordinates that are not selected as numerical constants. Therefore, the condition number at each generation is recalculated from the selected coordinate space. The original condition number is 10^6 for the $100,000$-dimensional ellipsoid function. The expected value of the condition number in randomly selected 100-coordinates' space is approximately 7.7×10^5 which is about 25% less than the original one. To further reduce the condition number, we propose curvature-based coordinate selection that selects coordinates based on the estimated curvature of each coordinate of the function.

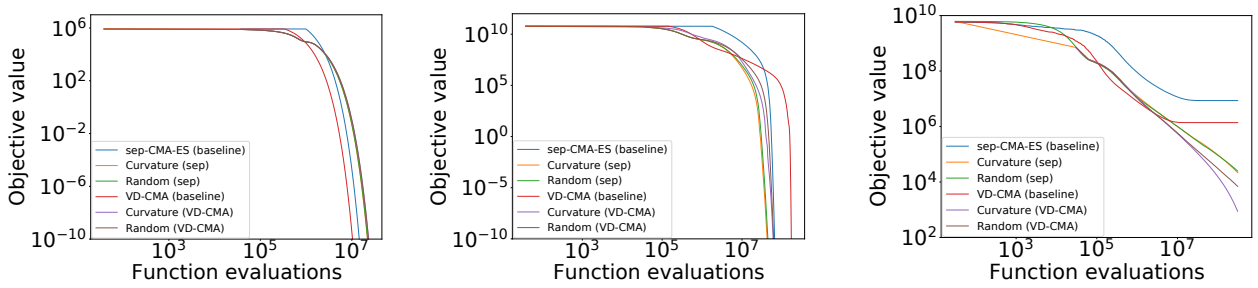
To vectorize the step-size is not useful for conventional CMA-ES because the updater of the step-size is a scalar value for all elements. However, a set of the elements of the vectorized step-size is changed at each generations in our method. Thus, each element is updated to adapt to the scale of each coordinate.

The overall algorithm to minimize an objective function $f(\boldsymbol{x}) \in \mathbb{R}$, $\boldsymbol{x} \in \mathbb{R}^d$ is as follows. To begin, the mean vector $\boldsymbol{m}^{(0)} \in \mathbb{R}^d$, the covariance matrix $\boldsymbol{C}^{(0)} \in \mathbb{R}^{d\times d}$, and the vectorized step-size $\boldsymbol{\sigma}^{(0)} \in \mathbb{R}^n$ are determined based on the search region, and the evolution paths of the covariance matrix and the step-size are initialized as $\boldsymbol{p}_{\boldsymbol{c}}^{(0)} \in \mathbb{R}^d$ and $\boldsymbol{p}_{\boldsymbol{\sigma}}^{(0)} \in \mathbb{R}^d = \boldsymbol{0}$. Next, we introduce an index vector $\boldsymbol{\iota} \in \mathbb{R}^d = [0, 1, ..., d-1]$ as the length of the input dimensions of $f(x)$ and $e \in \mathbb{R}$ (initially $e = 0$) as the end point of the selected vector. The number of dimensions to be selected in each iteration is $s \in \mathbb{R}$. Then, we repeat the following steps until the predetermined termination conditions are met.

[Step 1] From $\boldsymbol{\iota}$, we take the elements from ι_e to ι_{e+s} and let them be the index vector $\boldsymbol{\iota'} \in \mathbb{R}^s$ in the current generation. Then,

GECCO '21 Companion, July 10–14, 2021, Lille, France

ACM ISBN 978-1-4503-8351-6/21/07.
https://doi.org/10.1145/3449726.3459575



(a) **100,000-D Sphere**, $f_{sph}(\boldsymbol{x})=\sum_{i=1}^{d} x_i^2$ (b) **100,000-D Ellipsoid**, $f_{ell}(\boldsymbol{x})=\sum_{i=1}^{d}(1000^{(i-1)/(d-1)}x_i)^2$ (c) **10,000-D Rotated ellipsoid**, $f_{ellrot}(\boldsymbol{x})=f_{ell}(\boldsymbol{Rx})$

Figure 1: Convergence of each function. R **is a** $d \times d$ **orthogonal matrix generated randomly.**

we assign $e+s$ as the new e. For $e > d-1$, we reorder the elements of $\boldsymbol{\iota}$ based on the estimated curvature $\boldsymbol{\kappa}$ (see [Step 7]) and $e \leftarrow 0$.

[Step 2] Based on the generated current generation index vector $\boldsymbol{\iota'}$, we generate a sub-vector and a sub-matrix from each variable: the mean sub-vector $\boldsymbol{m'} \in \mathbb{R}^s$, the covariance sub-matrix $\boldsymbol{C'} \in \mathbb{R}^{s\times s}$, the step-size sub-vector $\boldsymbol{\sigma'} \in \mathbb{R}^s$, and the evolution path sub-vectors $\boldsymbol{p'_c} \in \mathbb{R}^s$ and $\boldsymbol{p'_\sigma} \in \mathbb{R}^s$.

[Step 3] We sample λ individuals $\boldsymbol{x'}^{(t)} \in \mathbb{R}^{\lambda\times d}$ from a multivariate normal distribution as $\boldsymbol{z'}^{(t)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, $\boldsymbol{y'}^{(t)} = \boldsymbol{z'}^{(t)}\sqrt{\boldsymbol{C'}^{(t)}}$,

$$x_i'^{(t)} = \begin{cases} m_i^{(t)} + \sigma_i'^{(t)} y_i'^{(t)} & if\ i \in \boldsymbol{\iota'} \\ m_i^{(t)} & otherwise \end{cases}.$$

$\boldsymbol{x'}^{(t)}$ is a matrix of the same dimension as the normal CMA-ES here and takes same value of $\boldsymbol{m}^{(t)}$ for all the dimensions except for $\boldsymbol{\iota'}$.

[Step 4] As with the normal CMA-ES, each individual is evaluated and ranked by the objective function, then we compute the weighted sum as $\boldsymbol{dy'}^{(t)} = \sum_i^\lambda w_i \boldsymbol{y_i'}^{(t)}$, $\boldsymbol{dz'}^{(t)} = \sum_i^\lambda w_i \boldsymbol{z_i'}^{(t)}$.

[Step 5] We update the evolution paths $h_\sigma^{(t+1)}$, $\boldsymbol{p'_\sigma}^{(t+1)}$ and $\boldsymbol{p'_c}^{(t+1)}$ the same as normal CMA-ES. The elements specified by $\boldsymbol{\iota'}$ are updated here, and the other elements are not updated.

[Step 6] We update $\boldsymbol{m}^{(t)}$, $\boldsymbol{C}^{(t)}$, and $\boldsymbol{\sigma}^{(t)}$ based on the evolution path. As in [Step 5], we only update the elements specified by $\boldsymbol{\iota'}$, and we keep the original values of the other elements.

[Step 7] We estimate the curvature of the objective function as follows. First, we initialize $\boldsymbol{\delta p} = [0.0001, ..., 0.0001]^s$ as a small stride. Then, we adjust to adapt to the scale of each coordinate as $\boldsymbol{\delta q} = \boldsymbol{\sigma'}\boldsymbol{\delta p}\sqrt{\boldsymbol{C'}}$. Finally, we estimate the first-order derivatives $\boldsymbol{d_1}$, second-order derivatives $\boldsymbol{d_2}$, and curvature $\boldsymbol{\kappa}$ as

$$\boldsymbol{d_1} = \frac{f(\boldsymbol{m}+\boldsymbol{\delta q}) - f(\boldsymbol{m}-\boldsymbol{\delta q})}{2 * \boldsymbol{\delta q}},$$

$$\boldsymbol{d_2} = \frac{f(\boldsymbol{m}+\boldsymbol{\delta q}) - 2f(\boldsymbol{m}) + f(\boldsymbol{m}-\boldsymbol{\delta q})}{\boldsymbol{\delta q}^2}, \quad \boldsymbol{\kappa} = \frac{|\boldsymbol{d_2}|}{(1+\boldsymbol{d_1}^2)^{\frac{3}{2}}}.$$

3 EXPERIMENT

We used the values given in the paper [3] for the hyperparameters. We used our curvature-based coordinate selection and random-based coordinate selection (reorder $\boldsymbol{\iota}$ randomly in [step 1]) methods for the sep-CMA-ES and the VD-CMA [1] (Our methods are "Curvature" and "Random" in figure 1). As in the study [4], the initial values were set to $\boldsymbol{m}^{(0)} = U(-5, 5)$, $\boldsymbol{C}^{(0)} = \boldsymbol{I}$, $\sigma^{(0)} = 1.0$ ($\boldsymbol{\sigma^{(0)}} = [1.0, ..., 1.0]^d$ for our CMA-ES), and $\lambda = 4 + 3\lfloor \ln(d) \rfloor$. The target value of the objective function and the max. number of function evaluations were set to 10^{-10} and $\lambda \times 10^7$. The number of coordinates selected at each generation was set to $s = 100$.

Figure 1 (b) and (c) show that our algorithms reached a better solution in fewer function evaluations than the conventional CMA-ES for ill-conditioned functions, ellipsoid and rotated ellipsoid. However, figure 1 (a) shows that slightly more function evaluations reached the target value for the sphere function compared with the conventional CMA-ES. Coordinate selection limits the number of coordinates updated during a single iteration, degrading the performance of the well-conditioned function.

4 CONCLUSION AND FUTURE WORK

We proposed a CMA-ES with coordinate selection to reduce the condition number and adapt the vectorized step-size to the scale of each coordinate of an objective function. We used our method for several variations of CMA-ES and improved the performance of the optimization of high-dimensional and ill-conditioned functions. It is truly important for our method how we select coordinates. If we determined the number of selection dynamically (contrast to our current method $s = 100$), the performance of well-conditioned functions could be improved.

ACKNOWLEDGEMENT

This work was supported by JST CREST JPMJCR19A4.

REFERENCES

[1] Y. Akimoto, A. Auger, and N. Hansen. 2014. Comparison-Based Natural Gradient Optimization in High Dimension. In *Proc. of GECCO '14.* ACM Press, 373–380.
[2] N. Hansen. 2016. The CMA Evolution Strategy: A Tutorial. *arXiv:1604.00772.*
[3] N. Hansen and A. Auger. 2014. Principled Design of Continuous Stochastic Search: From Theory to Practice. *Theory and Principled Methods for the Design of Metaheuristics* (2014), 145–180.
[4] I. Loshchilov. 2014. A Computationally Efficient Limited Memory CMA-ES for Large Scale Optimization. In *Proc. of GECCO '14.* ACM Press, 397–404.
[5] R. Ros and N. Hansen. 2008. A Simple Modification in CMA-ES Achieving Linear Time and Space Complexity. In *PPSN X.* Vol. 5199. Springer, 296–305.