ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

A Fast and Simple Evolution Strategy with Covariance Matrix Estimation

Oliver Kramer

Computational Intelligence Lab
Department of Computer Science
Carl-von-Ossietzky University of Oldenburg
26111 Oldenburg, Germany
oliver.kramer@uni-oldenburg.de

Abstract. With the rise of A.I. methods the demand for efficient optimization methods that are easy to implement and use increases. This paper introduces a simple optimization method for numerical blackbox optimization. It proposes to apply covariance matrix estimation for the (1+1)-ES with Rechenberg's step size control. Experiments on a small set of benchmark functions demonstrate that the approach outperforms its isotropic variant allowing competitive convergence on problems with scaled and correlated dimensions.

1 Introduction

The covariance matrix adaptation evolution strategy (CMA-ES) has set standards in the field of blackbox numerical optimization. It is based on learning a covariance matrix that allows adaptation to local search space characteristics. Several variants have been proposed throughout the history of evolutionary strategies. A simple but effective one is the (1+1)-CMA-ES [5], which combines the one (1+1)-ES with a variant of the Rechenberg rule. The covariance matrix is adapted like in the original CMA-ES, a Cholesky variant is also proposed.

This paper goes one step further and proposes an even easier to implement and handle variant that is based on the estimation of the covariance matrix. The integration of a simple covariance matrix estimator based on a sliding window of the best solutions for a (1+1)-ES with Rechenberg's 1/5th success rule allows the approximation of optima in solution spaces with correlated dimensions. The approach assumes that the estimation of the covariance matrix in black-box optimization is not the computational bottleneck. More important is the minimization of the number of fitness function calls, i.e., for estimation of the covariance matrix and for the optimization process with its precise estimate.

This paper is structured as follows. Section 2 gives an overview of ES variants related to step size control and covariance matrix adaptation. Section 3 introduces the covariance matrix estimation ES with Rechenberg's step size control (CMR-ES), which is experimentally analyzed in Section 4. Conclusions are drawn in Section 5.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

2 Related Work

Parameter control has an important part to play in ES. The choice of optimal strategy variables or hyperparameters like population sizes and step sizes can have a significant impact on its effectiveness. Besides constant parameters or static parameter control strategies like linear mutation rate decrease, adaptive techniques like Rechenberg's rule, see Section 3, take into account feedback from the search.

Self-adaptation [10] allows multivariate Gaussian mutations with independent step sizes per dimension by adding the step size vector to the chromosome of the solution. While being evolved together with the objective variables with crossover and mutation, successful step sizes have a high probability to dominate during the optimization process. Selection noise is the case of a random decoupling between the magnitude of step sizes and the realizations of their random numbers. Large population sizes or the derandomized self-adaptive ES [4] overcome selection noise. The latter changes step sizes and objective variables with the same randomly drawn numbers applying them to strategy and objective variables.

To consider the movement of solutions during the search, an evolution path ES [8] estimates the length of the advancement of movements in certain directions of the solution space and compares it to the expected size of mutations for step size adaptation.

The CMA-ES [3] approximates the covariance matrix of the best solutions and combines it with the evolution path principle for step sizes and the covariance matrix with its rank-1 and rank-μ updates. The (1+1)-CMA-ES [5] allows the approximation of a covariance matrix with a simple CMA-ES with Rechenberg 1/5th success rule. Based on an extended success rule that has influence on the covariance matrix update, this relatively simple variant shows competitive results. The ES proposed in this paper is an even simpler variant than the (1+1)-CMA-ES, but is still competitive.

Other CMA-ES variants exist, which mainly aim to simplify the original, e.g., the self-adaptive CMSA-ES [1], which applies the principle of self-adaptation to the covariance matrix. Another example is the matrix adaptation ES (MA-ES) [2], which replaces the construction of the covariance matrix by a transformation matrix. A memory-efficient variant has been introduced in [6].

A relatively late development is the natural evolution strategy (NA-ES) [11], which treats the search parameters as distribution and performs gradient ascent in the space of distribution parameters. With Monte Carlo estimation the distribution parameters like covariance matrix and step sizes for scaling the distribution are determined. Convergence of NA-ES on the Sphere function has been analyzed in [9].

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

3 CMR-ES

The goal of covariance matrix adaptation is to learn covariance matrix $\mathbf{C}$ and step size σ for efficient approximation in multimodal blackbox optimization problems. Decoupling learning covariance matrix and step sizes simplifies the learning process.

As a simple covariance matrix ES this paper proposes to employ covariance matrix estimation in a (1+1)-ES using Rechenberg's 1/5th step size control, see Figure 1. For the estimation a simple covariance matrix estimator available in many programming languages can be used.

The estimation of $\mathbf{C}$ is based on a sliding window approach that keeps an archive A of the η best solutions of the last successful generations. At the beginning A is empty. When A reaches size η, covariance matrix $\mathbf{C}$ is estimated to capture the curvature of the fitness landscape, see Figure 2. The (1+1)-CMR-ES generates an offspring solution $\mathbf{x}'$ from a parental solution $\mathbf{x}$ with Gaussian mutation:

$$\mathbf{x}' := \mathbf{x} + \sigma \cdot \mathcal{N}(\mathbf{0}, \mathbf{C}) \tag{1}$$

where $\mathcal{N}(\mathbf{0}, \mathbf{C})$ generates an N-dimensional vector of random numbers with center $\mathbf{0}$ and covariance matrix $\mathbf{C}$.

Algorithm 1: CMR-ES

given $d \approx \sqrt{N+1}, \kappa, \eta$
initialize $\mathbf{x} \in \mathbb{R}^N, \sigma \in \mathbb{R}^+$, $A = []$
repeat
  every κ gen. estimate $\mathbf{C}$ with A
  $\mathbf{x}' := \mathbf{x} + \sigma \cdot \mathcal{N}(\mathbf{0}, \mathbf{C})$
  $\sigma := \sigma \cdot \exp^{1/d}(\mathbb{I}_{f(\mathbf{x}') \leq f(\mathbf{x})} - 1/5)$
  if $f(\mathbf{x}') \leq f(\mathbf{x})$ **then**
    replace $\mathbf{x}$ with $\mathbf{x}'$
    $A = A[1:\eta].\text{append}(\mathbf{x})$
  end if
until termination condition

Fig. 1: $(1+1)$-CMR-ES with Rechenberg's step size control

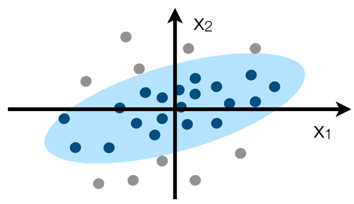


Fig. 2: Estimation of $\mathbf{C}$ with η last best solutions (blue) neglecting older/worse offspring (grey).

The Rechenberg 1/5th rule increases step size σ by multiplication with $\exp(4/5)$, if the success probability is larger than 1/5, and decreases σ by multiplication with $\exp(-1/5)$, if the success probability is smaller than 1/5:

$$\sigma := \sigma \cdot \exp^{1/d}(\mathbb{I}_{f(\mathbf{x}') \leq f(\mathbf{x})} - 1/5) \tag{2}$$

with scaling parameter d. Indicator function $\mathbb{I}_{f(\mathbf{x}') \leq f(\mathbf{x})}$ delivers 1 in case of improvement, i.e., $f(\mathbf{x}') \leq f(\mathbf{x})$ is true for minimization problems, and 0 otherwise. This adaptation reaches a stationary case, if the success probability is 1/5, which has been shown to allow nearly optimal progress on many functions.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Based on their fitness, offspring or parent are selected for the next generation. A new best solution is put into archive A, while the worst solution is removed. The problem dependent frequency of re-estimations of $\mathbf{C}$ is specified by parameter κ.

4 Experiments

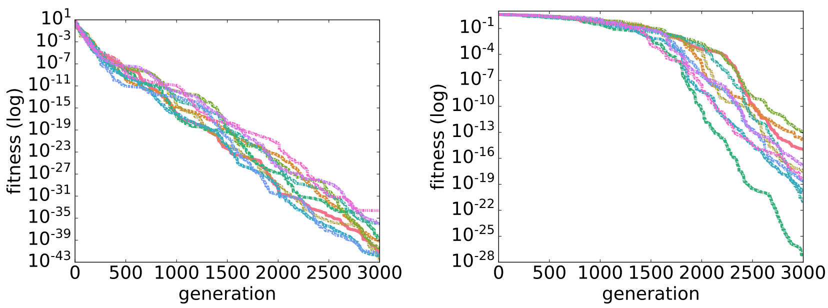
This section presents an experimental analysis on a scaled version of the Sphere function, i.e., Scaled Sphere with $N = 5$ and scaling vector $\mathbf{w} = (10^3, 1, 1, \ldots)$, and on Rosenbrock with $N = 5$, see Section A.

	Scaled Sphere	Rosenbrock
ES	21.87 ± 18.47	$0.28 \pm 3.6 \cdot 10^{-2}$
CMA-ES	$4.5 \cdot 10^{-29} \pm 6.5 \cdot 10^{-29}$	$\mathbf{7.0 \cdot 10^{-23} \pm 1.1 \cdot 10^{-22}}$
CMR-ES	$\mathbf{2.2 \cdot 10^{-35} \pm 6.7 \cdot 10^{-35}}$	$1.2 \cdot 10^{-14} \pm 3.2 \cdot 10^{-14}$

Table 1: Comparison between (1+1)-ES, CMA-ES, and (1+1)-CMR-ES on Scaled Sphere and Rosenbrock with $N = 5$ after $3,000$ fitness function evaluations.

Table 1 shows means and standard deviations of final fitness values achieved at the end of each 100 runs of the (1+1)-ES with Rechenberg's 1/5th success rule, a (1,8)-CMA-ES, and the (1+1)-CMR-ES. Both (1+1)-ES variants use mutation parameter $d = \sqrt{6}$. All ES employ $3,000$ fitness function evaluations. The CMR-ES uses archive size $\eta = 20$ and estimates $\mathbf{C}$ every $\kappa = 200$ generations.

The results show the CMR-ES clearly outperforms the standard (1+1)-ES, showing its approximation capabilities in solution spaces with scaled and correlated dimensions. It outperforms the CMA-ES on Scaled Sphere and is



(a) Scaled Sphere (b) Rosenbrock

Fig. 3: Logarithmic fitness development of each ten exemplary runs of the (1+1)-CMR-ES (a) on Scaled Sphere and (b) on Rosenbrock with $N = 5$.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

outperformed on Rosenbrock, while not getting stuck on the latter.

The desired log-linear approximation of optima can also be observed in Figure 3, which shows the fitness developments of each 10 random runs of the (1+1)-CMR-ES on Scaled Sphere and on Rosenbrock on a logarithmic scale. The covariance matrix re-estimation results in numerous phases of different approximation speeds.

5 Conclusions

Covariance matrix adaptation strategies are state-of-the-art blackbox optimization methods in numerical blackbox optimization. To simplify things, this paper revisits the (1+1)-CMA-ES principle and proposes a combination of covariance estimation and the Rechenberg 1/5th success rule. The covariance matrix estimation is based on a sliding window of the best solutions generated during evolution. The computational complexity of the covariance matrix estimation step of modern implementations can rarely be a bottleneck in practical optimization. More important are savings of potentially expensive fitness function evaluations. Furthermore, the CMR-ES has few problems with numerical instabilities, when handling the covariance matrix.

Future experimental comparisons will concentrate on tuning hyperparameters and experiments on an extended set of numerical test functions considering comparisons to CMA-ES variants like the (1+1)-Cholesky-CMA-ES [5].

A Benchmark Functions

The experimental analysis is based on the following numerical blackbox minimization test functions:

1. Scaled Sphere:
$$f(\mathbf{x}) := (\mathbf{w}\mathbf{x})^T \cdot (\mathbf{w}\mathbf{x}) \tag{3}$$
with $\mathbf{w} = (10^3, 1, 1, \ldots)$, initialization $\mathbf{x}^0 = \mathbf{1}$, optimum $\mathbf{x}^* = \mathbf{0}$ with $f(\mathbf{x}) = 0.0$

2. Rosenbrock:
$$f(\mathbf{x}) := \sum_{i=1}^{N-1} \left[100 \cdot (x_{i+1} - x_i^2)^2 + (1 - x_i)^2\right] \tag{4}$$
initialization $\mathbf{x}^0 = \mathbf{0}$, optimum $\mathbf{x}^* = \mathbf{1}$ with $f(\mathbf{x}) = 0.0$

B Source Code

The CMR-ES in the experimental analysis is implemented in Python. The covariance matrix estimation is based on the function numpy.cov. The source code is available on GitHub. The CMA-ES experiments are based on the CMA-ES of the CMAES Python package (Version 0.8.2) [7].

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

References

[1] H. Beyer and B. Sendhoff. Covariance matrix adaptation revisited - the CMSA evolution strategy. In *Parallel Problem Solving from Nature (PPSN)*, pages 123–132. Springer, 2008.

[2] H.-G. Beyer and B. Sendhoff. Simplify your covariance matrix adaptation evolution strategy. *IEEE Transactions on Evolutionary Computation*, 21(5):746–759, 2017.

[3] N. Hansen, S. D. 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.

[4] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[5] C. Igel, T. Suttorp, and N. Hansen. A computational efficient covariance matrix update and a (1+1)-CMA for evolution strategies. In *Genetic and Evolutionary Computation Conference (GECCO)*, pages 453–460. ACM, 2006.

[6] I. Loshchilov, T. Glasmachers, and H. Beyer. Large scale black-box optimization by limited-memory matrix adaptation. *IEEE Transactions on Evolutionary Computation*, 23(2):353–358, 2019.

[7] M. Nomura, S. Watanabe, Y. Akimoto, Y. Ozaki, and M. Onishi. Warm starting CMA-ES for hyperparameter optimization. In *Conference on Artificial Intelligence (AAAI)*, pages 9188–9196. AAAI Press, 2021.

[8] A. Ostermeier, A. Gawelczyk, and N. Hansen. A derandomized approach to self adaptation of evolution strategies. *Evolutionary Computation*, 2(4):369–380, 1994.

[9] T. Schaul. Natural evolution strategies converge on sphere functions. In *Genetic and Evolutionary Computation Conference (GECCO)*, pages 329–336. ACM, 2012.

[10] H. Schwefel. *Evolution and Optimum Seeking*. Wiley, New York, 1995.

[11] D. Wierstra, T. Schaul, T. Glasmachers, Y. Sun, J. Peters, and J. Schmidhuber. Natural evolution strategies. *Journal of Machine Learning Research*, 15(1), 2014.