

Improving (1+1) Covariance Matrix Adaptation Evolution Strategy: a Simple Yet Efficient Approach

Fabio Caraffini^{1,a)}, Giovanni Iacca^{2,b)} and Anil Yaman^{3,c)}

¹*School of Computer Science and Informatics, De Montfort University, UK*

²*Department of Information Engineering and Computer Science, University of Trento, Italy*

³*Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands*

^{a)}Corresponding author: fabio.caraffini@dmu.ac.uk

^{b)}giovanni.iacca@unitn.it

^{c)}a.yaman@tue.nl

Abstract. In recent years, part of the meta-heuristic optimisation research community has called for a simplification of the algorithmic design: indeed, while most of the state-of-the-art algorithms are characterised by a high level of complexity, complex algorithms are hard to understand and therefore tune for specific real-world applications. Here, we follow this reductionist approach by combining straightforwardly two methods recently proposed in the literature, namely the Re-sampling Inheritance Search (RIS) and the (1+1) Covariance Matrix Adaptation Evolution Strategy (CMA-ES). We propose an RI-(1+1)-CMA-ES algorithm that on the one hand improves upon the original (1+1)-CMA-ES, on the other it keeps the original spirit of simplicity at the basis of RIS. We show with an extensive experimental campaign that the proposed algorithm efficiently solves a large number of benchmark functions and is competitive with several modern optimisation algorithms much more complex in terms of algorithmic design.

INTRODUCTION

One of the most successful continuous optimisation algorithms currently available in the literature is the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [1]. CMA-ES is a metaheuristic which builds a covariance matrix during the evolutionary process, in the attempt of detecting a coordinate transformation in the search space and adapt the evolutionary operators (mutations) to the new set of coordinates. In a way, this mechanism is similar to a Principal Component Analysis (PCA), as it tries to use the transformed coordinates to exploit the “shape” of the fitness landscape thus guiding the evolution. However, the main drawback of CMA-ES is its algorithmic and computational complexity, which hampers its applicability, especially to large-scale problems.

Several simplifications of CMA-ES have therefore been presented in recent years, for instance the sep-CMA-ES [2], that uses a diagonal covariance matrix and samples each coordinate independently, or the (1+1)-CMA-ES [3], originally devised as building block of the Multi-Objective CMA-ES (MO-CMA-ES) [4], which reduces the complexity from $O(n^3)$ to $O(n^2)$ by employing the Cholesky decomposition of the covariance matrix. The price of these simplifications is a reduced optimisation performance, especially on non-separable, “rotated” or ill-conditioned problems, which are typically presented in modern benchmarks to emulate hard-to-solve real-world problems [5].

To overcome this problem, and obtain an algorithm that may preserve the performance of the original CMA-ES while maintaining a low level of algorithmic and computational complexity, we introduce here a version of (1+1)-CMA-ES that includes the so-called Re-sampling Inheritance Search (RIS), which was proven to be a simple yet efficient framework, in particular on large scale and separable problems [6, 7]. This idea follows the Ockham’s Razor advocated in [8], which states that simplicity should always be considered a crucial aspect of algorithmic design.

The proposed algorithm, dubbed RI-(1+1)-CMA-ES, shares to some extent the working principle of one of the most popular CMA-ES variant, namely g-CMA-ES [9], which makes use of a restart mechanism to reset the covariance matrix and set up a higher population size (i.e. number of individuals sampled from the multivariate distribution), when CMA-ES fails at improving upon the fitness function. However, it is important to notice that we do not propose the same restart process, but rather an iterated local search similar to RIS. More specifically, not

only we reset the covariance matrix, but we also sample a *random new solution*, and subsequently we feed it to an *exponential crossover* (see [6] for details), together with the current best solution (“elite”) found by the algorithm. The resulting solution then shares some variables with the current best solution, and is used after the restart as new start point for (1+1)-CMA-ES. This process is every time independent to the previous ones and has its own evolution. However, if the results improve upon the previous best solution, the latter gets replaced.

EXPERIMENTAL SETUP

The proposed RI-(1+1)-CMA-ES algorithm was thoroughly tested over the 30 problems of the CEC2014 benchmark suite for real-valued global optimisation [10] in 10, 50 and 100 dimension values. Thus, 90 different optimisation problems are considered in this study.

To prove its versatility, the proposed algorithm was compared against nine competitors from different optimisation paradigms. A first comparison was performed against its “predecessor” (1+1)-CMA-ES and two additional classic single-solution algorithms, namely SPSA [11] and Rosenbrock [12]. It is worth noting that these three different algorithms still share a common working principle: they attempt to find a promising direction to follow by implicitly exploiting the gradient information. Indeed, SPSA approximates the gradient in a stochastic way, Rosenbrock by rotating the coordinate axes, and (1+1)-CMA-ES by evolving a covariance matrix that is reflective of the inverse of the Hessian under certain conditions [13, 14]. Furthermore, three state-of-the-art single solution metaheuristics were selected and added to the comparison: ISPO, i.e. a single particle Particle Swarm Optimisation variant [15], cDE-light, i.e. an estimation of distribution algorithm based on Differential Evolution (DE) [16], and PMS, i.e. a Memetic Computing (MC) approach [17]. The latter, can be seen as an evolution of the RIS algorithm. Finally, three population-based algorithms were chosen to compare RI-(1+1)-CMA-ES against more robust and complex algorithmic structures: MDE-pBX [18], JADE [19], and the μ DEA [20]. These algorithms are based on advanced DE schemes and the latter also embeds a particularly efficient MC operator which is activated during the optimisation process.

To produce our numerical results, each optimiser was run 30 times with the parameter settings suggested in the original papers, and a computational budget $B = 5000D$ ($D = 10, 50, 100$ dimensions). As for RI-(1+1)-CMA-ES, the same parameters setting proposed in [6] was used for the “restart” procedure, and the one suggested in [3] for the (1+1)-CMA-ES component. The statistical analysis suggested in [21] was adopted to further validate our conclusions. This means that the aforementioned algorithms were first compared, on each of the 90 problems under consideration, by means of the pairwise Wilcoxon Rank-Sum test [22]. To save space, detailed tables reporting pairwise comparisons are made available online [23] and only commented in the next section. Conversely, the outcome of the Holm-Bonferroni method for multiple hypothesis testing [24] is included and reported in Table 1. The table ranks each j -th algorithm ($j = 1, 2, \dots, 9$) -excluding the reference RI-(1+1)-CMA-ES- based on its average rank on all tested functions (in this case, from rank 1 to rank 10, the higher the better); z_j , p_j and δ/j are calculated as in [25] (being δ the confidence interval set to 0.05); if $p_j < \delta/j$, the null-hypothesis (that RI-(1+1)-CMA-ES has the same performance as the j -th algorithm) is rejected, otherwise is accepted as well as all the subsequent tests.

NUMERICAL RESULTS

Table 1 clearly shows that RI-(1+1)-CMA-ES is a powerful single-solution metaheuristic capable of outperforming several state-of-the-art algorithms, including population-based ones. According to our results, the proposed approach appears to be preferable to (1+1)-CMA-ES, i.e. the corresponding null hypothesis is rejected with a confidence level of 95%. The same conclusion is made for SPSA, JADE, ISPO and Rosenbrock. Moreover, RI-(1+1)-CMA-ES seems to be extremely competitive with modern DE variants as cDE-light and MDE-pBX, and comparable to complex MC frameworks, i.e. PMS and μ DEA. It must be noted that our algorithm displays the higher rank also for the cases with “accepted” null hypothesis. Thus, one can conclude that RI-(1+1)-CMA-ES is on average the best optimiser involved in this study. However, by looking at the pairwise results in [23], it can be observed a moderate deterioration in the performance, with respect to (1+1)-CMA-ES, on unimodal functions (in particular at high dimension values, see Fig. 1). This is an expected consequence of adding a restarting mechanism to better handle multimodal problems, over which (1+1)-CMA-ES performs poorly. In exchange, the proposed approach significantly improves upon its predecessor on multimodal problems, thus justifying our algorithmic choice. This is evident in Fig. 2 as the fitness trends, represented in terms of average error $f_n(i) - f_n^{min}$ ($i = 1, 2, \dots, B$, $n = 7$ in this specific case), show that all algorithms but RI-(1+1)-CMA-ES get stuck in local minima. In such figure, the fitness trend of our algorithm is difficult to be distinguished from the x-axis, as it approaches a negligible error value of magnitude 10^{-13} .

TABLE 1. Holm-Bonferroni procedure (reference: **RI-(1 + 1)-CMA-ES**, Rank = **7.16e + 00**)

j	Optimizer	Rank	z_j	p_j	δ/j	Hypothesis
1	μ DEA	7.00e+00	-3.81e-01	3.52e-01	5.00e-02	Accepted
2	MDE-pBX	6.83e+00	-7.89e-01	2.15e-01	2.50e-02	Accepted
3	PMS	6.68e+00	-1.17e+00	1.21e-01	1.67e-02	Accepted
4	cDE-light	6.63e+00	-1.28e+00	1.00e-01	1.25e-02	Accepted
5	(1 + 1)-CMA-ES	6.12e+00	-2.53e+00	5.68e-03	1.00e-02	Rejected
6	Rosenbrock	5.19e+00	-4.82e+00	7.27e-07	8.33e-03	Rejected
7	ISPO	4.17e+00	-7.32e+00	1.23e-13	7.14e-03	Rejected
8	JADE	3.93e+00	-7.89e+00	1.48e-15	6.25e-03	Rejected
9	SPSA	1.27e+00	-1.44e+01	1.81e-47	5.56e-03	Rejected

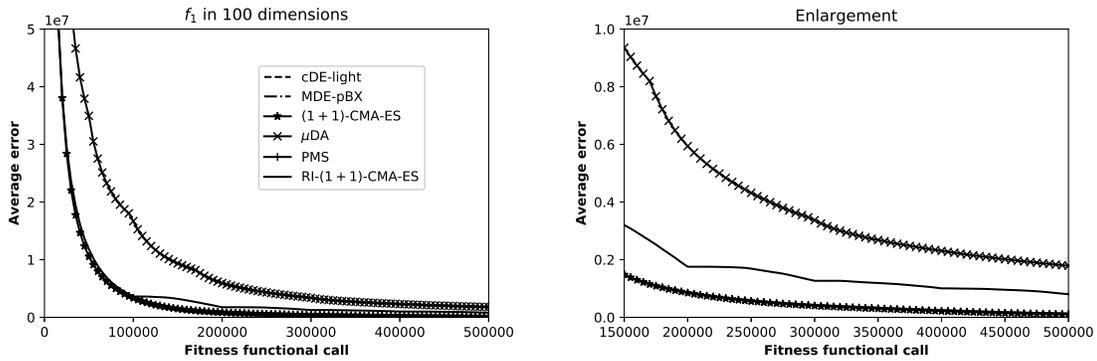


FIGURE 1. Average fitness trends on f_1 from CEC2014 in 100D: complete trends (left), zoom at #func. calls > 150k (right)

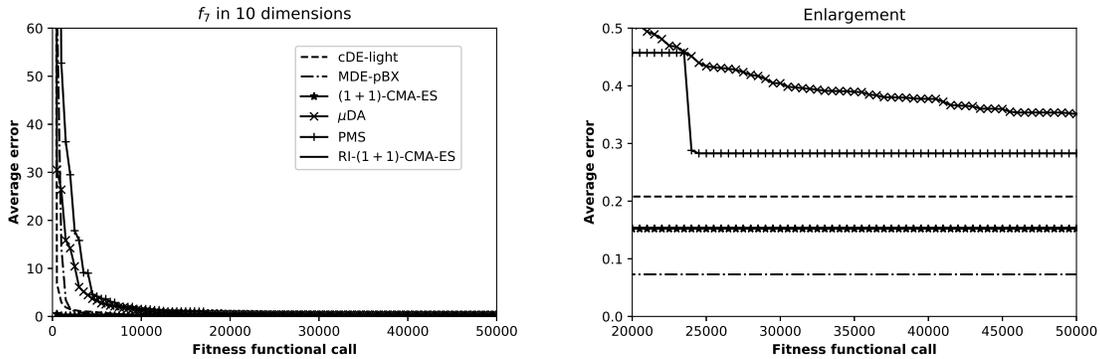


FIGURE 2. Average fitness trends on f_7 from CEC2014 in 10D: complete trends (left), zoom at #func. calls > 20k (right)

CONCLUSIONS

We proposed a novel algorithm which iteratively executes (1+1)-CMA-ES from multiple starting points in order to boost its performances over multimodal and ill-conditioned problems. Initial positions are generated by re-sampling a solution in the search domain, and subsequently “mating” it via exponential crossover with the best ever-found solution in order to inherit promising components. This approach, borrowed from the RIS algorithm, turned out to be surprisingly successful when used with (1+1)-CMA-ES. Indeed, the proposed RI-(1+1)-CMA-ES can efficiently tackle multimodal problems despite the lack of a population of candidate solutions. In this light, the proposed approach makes a significant contribution as it proves that also a simple strategy, if well designed, can outperform more complex population-based algorithmic structures.

ACKNOWLEDGEMENTS

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No: 665347.

REFERENCES

- [1] N. Hansen, “The CMA Evolution Strategy: A Comparing Review,” in *Towards a New Evolutionary Computation: Advances in the Estimation of Distribution Algorithms* (Springer, 2006), pp. 75–102.
- [2] R. Ros and N. Hansen, “A Simple Modification in CMA-ES Achieving Linear Time and Space Complexity,” in *Parallel Problem Solving from Nature – PPSN X*, edited by G. Rudolph, T. Jansen, N. Beume, S. Lucas, and C. Poloni (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008), pp. 296–305.
- [3] C. Igel, T. Suttorp, and N. Hansen, “A Computational Efficient Covariance Matrix Update and a (1+1)-CMA for Evolution Strategies,” in *Proceedings of the Genetic and Evolutionary Computation Conference* (ACM Press, 2006), pp. 453–460.
- [4] C. Igel, N. Hansen, and S. Roth, *Evolutionary Computation* 15, 1–28 (2007).
- [5] F. Caraffini and F. Neri, “Rotation invariance and rotated problems: An experimental study on differential evolution,” in *International Conference on the Applications of Evolutionary Computation* (Springer, Cham, 2018), pp. 597–614.
- [6] F. Caraffini, F. Neri, B. N. Passow, and G. Iacca, *Soft Computing* 17, 2235–2256 (2013).
- [7] F. Caraffini, G. Iacca, F. Neri, L. Picinali, and E. Mininno, “A CMA-ES super-fit scheme for the re-sampled inheritance search,” in *2013 IEEE Congress on Evolutionary Computation* (2013), pp. 1123–1130.
- [8] G. Iacca, F. Neri, E. Mininno, Y.-S. Ong, and M.-H. Lim, *Information Sciences* 188, 17 – 43 (2012).
- [9] A. Auger and N. Hansen, “A restart CMA evolution strategy with increasing population size,” in *2005 IEEE Congress on Evolutionary Computation, Vol. 2* (2005), pp. 1769–1776 Vol. 2.
- [10] J. J. Liang, B. Y. Qu, and Suganthan, “Problem Definitions and Evaluation Criteria for the CEC 2014 Special Session on Real-Parameter Optimization,” Tech. Rep. 201311 (Zhengzhou University and Nanyang Technological University, Zhengzhou China and Singapore, 2013).
- [11] J. C. Spall, *IEEE Transactions on Aerospace and Electronic Systems* 34, 817–823 (1998).
- [12] H. H. Rosenbrock, *The Computer Journal* 3, 175–184 (1960).
- [13] O. M. Shir, J. Roslund, D. Whitley, and H. Rabitz, *Physical Review E* 89, p. 063306 (2014).
- [14] O. M. Shir and A. Yehudayoff, “On the statistical learning ability of evolution strategies,” in *Proceedings of the 14th ACM/SIGEVO Conference on Foundations of Genetic Algorithms* (ACM, 2017), pp. 127–138.
- [15] Z. Ji, J. Zhou, H.-L. Liao, and Q.-H. Wu, *Chinese Journal of Computers* 33, 556–561 (2010).
- [16] G. Iacca, F. Caraffini, and F. Neri, *Journal of Computer Science and Technology* 27, 1056–1076 (2012).
- [17] F. Caraffini, F. Neri, G. Iacca, and A. Mol, *Information Sciences* 227, 60 – 82 (2013).
- [18] M. I. Sk, S. Das, S. Ghosh, S. Roy, and P. Suganthan, *IEEE Transactions on Systems, Man, and Cybernetics. Part B* 42, 482–500 (2011).
- [19] J. Zhang and A. Sanderson, *IEEE Transactions on Evolutionary Computation* 13, 945–958 (2009).
- [20] F. Caraffini, F. Neri, and I. Poikolainen, “Micro-differential evolution with extra moves along the axes,” in *2013 IEEE Symposium on Differential Evolution (SDE)* (2013), pp. 46–53.
- [21] S. Garcia, A. Fernandez, J. Luengo, and F. Herrera, *Soft Computing* 13, 959–977 (2008).
- [22] F. Wilcoxon, *Biometrics Bulletin* 1, 80–83 (1945).
- [23] F. Caraffini, G. Iacca, and A. Yaman, “Improving (1+1) covariance matrix adaptation evolution strategy: a simple yet efficient approach - extended results,” www.tech.dmu.ac.uk/~fcaraf00/NumericalResults/GOW18_RI11CMAES_RESULTS.pdf (2018).
- [24] S. Holm, *Scandinavian Journal of Statistics* 6, 65–70 (1979).
- [25] G. Iacca, F. Caraffini, and F. Neri, *International journal of neural systems* 24, p. 1450008 (2014).