A comparison of three Differential Evolution strategies in terms of early convergence with different population sizes

Anil Yaman[1,a)], Giovanni Iacca[2,b)] and Fabio Caraffini[3,c)]

[1] *Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands*
[2] *Department of Information Engineering and Computer Science, University of Trento, Italy*
[3] *School of Computer Science and Informatics, De Montfort University, UK*

[a)] Corresponding author: a.yaman@tue.nl
[b)] giovanni.iacca@unitn.it
[c)] fabio.caraffini@dmu.ac.uk

Abstract. Differential Evolution (DE) is a popular population-based continuous optimization algorithm that generates new candidate solutions by perturbing the existing ones, using scaled differences of randomly selected solutions in the population. While the number of generation increases, the differences between the solutions in the population decrease and the population tends to converge to a small hyper-volume within the search space. When these differences become too small, the evolutionary process becomes inefficient as no further improvements on the fitness value can be made - unless specific mechanisms for diversity preservation or restart are implemented. In this work, we present a set of preliminary results on measuring the population diversity during the DE process, to investigate how different DE strategies and population sizes can lead to early convergence. In particular, we compare two standard DE strategies, namely "DE/rand/1/bin" and "DE/rand/1/exp", and a rotation-invariant strategy, "DE/current-to-random/1", with populations of $10, 30, 50, 100, 200$ solutions. Our results show, quite intuitively, that the lower is the population size, the higher is the chance of observing early convergence. Furthermore, the comparison of the different strategies shows that "DE/rand/1/exp" preserves the population diversity the most, whereas "DE/current-to-random/1" preserves diversity the least.

INTRODUCTION

Differential Evolution (DE) is a powerful yet simple metaheuristic for real-valued optimization which only requires tuning three parameters, i.e. the scale factor F, the crossover ratio CR and the population size N [1, 2, 3]. As its performance heavily depends on these parameters, a great research effort has been put in the last two decades to try to understand the effect and the optimal values of F and CR, with positive results [4, 5, 6, 7, 8, 9].

On the contrary, less clear is still the choice of the population size. In [5], it is recommended to use a number of solutions which is 10 times greater than the dimensionality of the problem. Such recommendation is arguable, as unfeasible in many real-world and large-scale scenarios [10, 11]. Indeed, an excessive number of candidate solutions can introduce a number of undesirable side effects, e.g. it deleteriously strengthens the algorithmic structural bias, which was proven to correlate with the population size in most population-based metaheuristics [12]. Moreover, it has been noted that shrinking the population size can be beneficial to avoid stagnation, an implicit drawback of DE [13], as shown in [14, 15] where the number of solutions is resized on-the-fly to improve the algorithmic performance in terms of robustness and convergence. In this direction, the proximity-based mutation scheme proposed in [16] theoretically guarantees a similar behavior, independently of the population size. However, in practice this method turns out to be unsuitable in presence of large populations, due to its computational cost. Operators resizing and controlling the population are commonly used also to prevent premature convergence [13, 17]. In this regard, it is worth mentioning the DE variant with auto-enhanced population diversity [18], as well as the mechanism introduced in [19], that preserves diversity by replacing more frequently solutions with higher fitness values.

In this light, it is clear that a good choice for the population size can make the difference in DE and all its variants [20]. Moreover, it must be observed that while the literature is abundant with enhanced DE variants, to the best of our knowledge a thorough study on the impact of the population size on premature convergence has never been properly conducted. This piece of research aims at investigating the relation between the population size and the dynamic behavior of DE, in order to learn lessons on how to address premature convergence. To investigate that, we measure the population diversity during the DE process with three commonly used strategies and various population sizes. This knowledge is in fact key to properly tune the algorithm and successfully use DE on real-world applications.

METHODS

We perform an analysis of *early convergence* in DE. Early convergence is observed when the solutions in a population become so similar that the differences between them approach to zero. In this case, DE is not able to produce any further improvement, since it generates new solutions based on scaled differences of randomly selected solutions. Here, we analyze the diversity of the population to investigate early convergence. At each generation of the tested algorithms (see below), we measure diversity by simply calculating the maximum value of the differences between the maximum and minimum values along each dimension of the candidate solutions: $diversity = \max_j \left[\max_i x^i_j - \min_i x^i_j\right]$, where i indicates the ith solution, and j indicates the jth dimension of the candidate solutions.

EXPERIMENTAL SETUP

We focus our analysis on three DE strategies: "DE/rand/1/bin" [1], "DE/rand/1/exp" [21] and "DE/current-to-random/1" [22] (in the following, we refer to these strategies as BIN, EXP and CURR, respectively), with populations of $10, 30, 50, 100, 200$ solutions. We test these configurations on the 28 functions of the CEC 2013 benchmark [23]. In all the experiments, F and CR are fixed, set to 0.4 and 0.3, respectively. The strategies and their parameters were selected based on their usage in previous works [1, 21]. It is important to note that while the crossover effect with a fixed CR is proportional to the problem dimensionality in case of BIN, this is not true for EXP: in this case, unless an adaptive CR is used [24], the crossover effect decreases with the dimensionality. The additional parameter K used in "DE/current-to-random/1" is randomly sampled in [0, 1], for each trial vector generation [22, 25]. The stop condition is set to $5000 \times D$ function evaluations, where D is the dimensionality of the problem. Each algorithm with specified experimental settings is executed for 30 independent runs, and the median of the results are provided.

RESULTS

Figure 1 shows the number of functions on which the BIN, EXP and CURR strategies performs the best on the benchmark functions with 10, 30 and 50 dimensions. On all tested dimensions, when the population size is smaller, EXP tends to perform better. The performance of the CURR tends to increase when the population size increases.

Figure 2 shows the number of functions for which each strategy obtains the lowest diversity at the last generation (ties are considered multiple times). According to the results, in all cases CURR reduces diversity more than the other two strategies, while EXP promotes diversity the most. In Fig 3, we report an example of the diversity trend observed on an evolutionary run of all the tested configurations on the benchmark function f_1 (sphere function). We provide an extended set of results as supplementary material available online [26].

CONCLUSIONS

Our results show that, on most of the tested functions, regardless of the population size, the rank of the strategies from the most diversity promoting strategy to the least one is as follows: EXP, BIN and CURR. Especially in a small population, CURR converges very quickly, thus further improvements in the fitness cannot be made. Larger populations instead decrease the chance of observing this effect, independently from the specific strategy. In future works, we will extend this study to provide adaptive mechanisms for optimal trade-off between strategies and population size.

ACKNOWLEDGMENTS

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

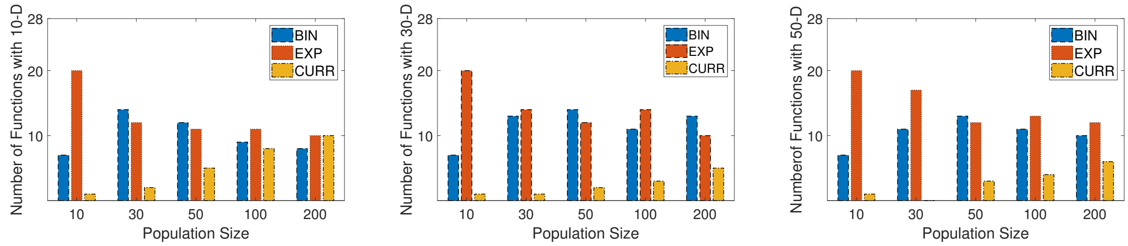

FIGURE 1. No. of functions for which each strategy obtains the smallest error (w.r.t. the optimal fitness) at the last generation.

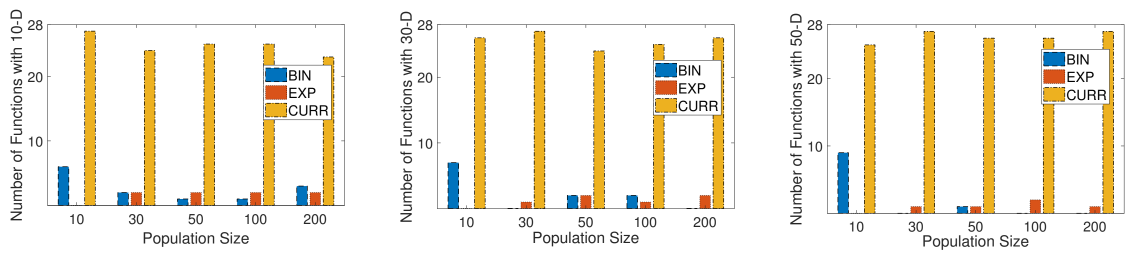

FIGURE 2. No. of functions for which each strategy obtains the lowest diversity at the last generation.

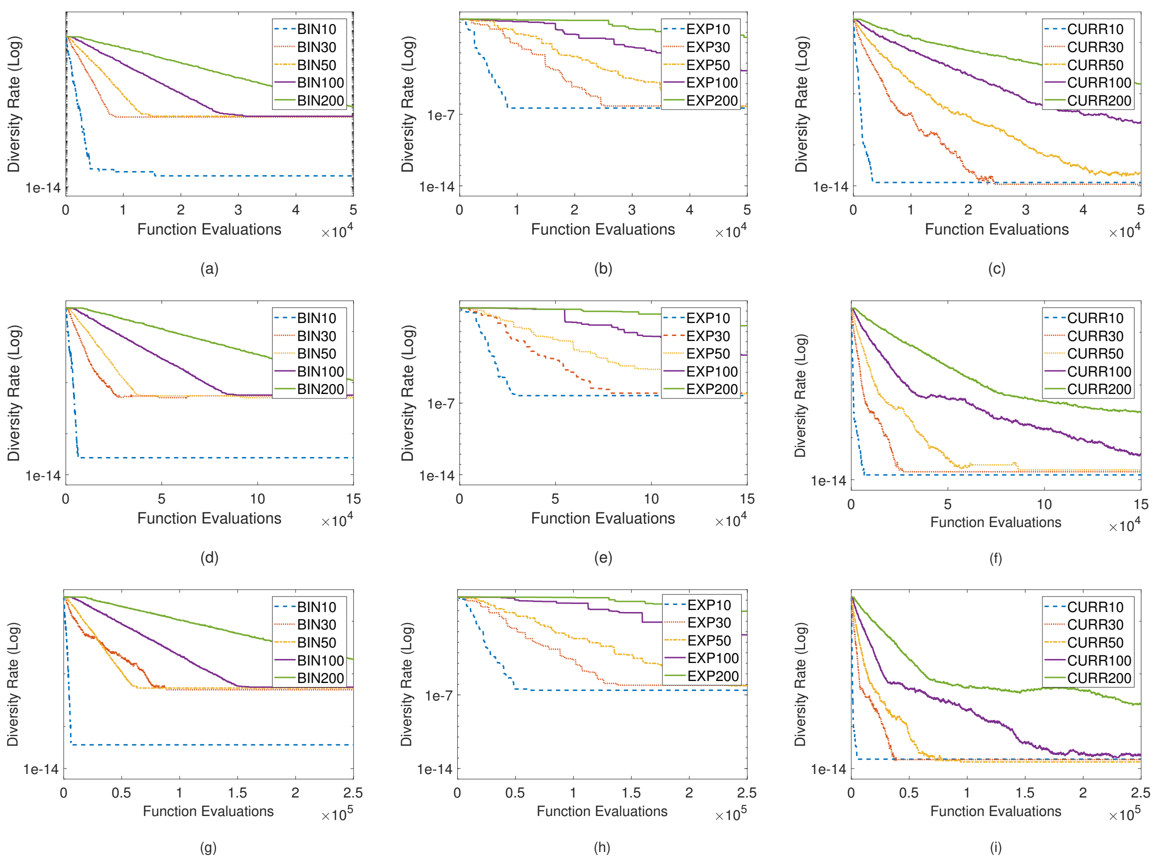

FIGURE 3. Diversity trends of BIN, EXP and CURR, with different population sizes, during an evolutionary run on f_1. (a), (b) and (c): results on 10 dimensions. (d), (e) and (f): results on 30 dimensions. (g), (h) and (i): results on 50 dimensions.

REFERENCES

[1] R. Storn and K. Price, "Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces," Tech. Rep. TR-95-012 (ICSI, 1995).

[2] D. E. Kvasov and M. S. Mukhametzhanov, Applied Mathematics and Computation **318**, 245–259 (2018).

[3] Y. D. Sergeyev, D. Kvasov, and M. Mukhametzhanov, Scientific reports **8**, p. 453 (2018).

[4] D. Zaharie, "Critical values for control parameters of differential evolution algorithm," in *Proceedings MENDEL* (2002), pp. 62–67.

[5] J. Liu and J. Lampinen, Soft Computing **9**, 448–462 (2005).

[6] R. Mallipeddi, G. Iacca, P. N. Suganthan, F. Neri, and E. Mininno, "Ensemble strategies in compact differential evolution," in *IEEE Congress on Evolutionary Computation (CEC)* (2011), pp. 1972–1977.

[7] G. Iacca, F. Neri, F. Caraffini, and P. N. Suganthan, "A differential evolution framework with ensemble of parameters and strategies and pool of local search algorithms," in *Applications of Evolutionary Computation* (Springer, 2014), pp. 615–626.

[8] G. Iacca, F. Caraffini, and F. Neri, "Continuous parameter pools in ensemble differential evolution," in *IEEE Symposium Series on Computational Intelligence* (IEEE, 2015), pp. 1529–1536.

[9] A. Yaman, G. Iacca, M. Coler, G. Fletcher, and M. Pechenizkiy, "Multi-strategy differential evolution," in *Applications of Evolutionary Computation* (Springer, 2018), pp. 617–633.

[10] K. E. Parsopoulos, "Cooperative micro-differential evolution for high-dimensional problems," in *Proceedings of the conference on Genetic and evolutionary computation* (2009), pp. 531–538.

[11] A. P. Piotrowski, Swarm and Evolutionary Computation **32**, 1–24 (2017).

[12] A. V. Kononova, D. W. Corne, P. D. Wilde, V. Shneer, and F. Caraffini, Information Sciences **298**, 468–490 (2015).

[13] J. Lampinen, I. Zelinka, *et al.*, "On stagnation of the differential evolution algorithm," in *Proceedings of MENDEL* (2000), pp. 76–83.

[14] J. Brest and M. S. Maučec, Applied Intelligence **29**, 228–247 (2008).

[15] G. Iacca, R. Mallipeddi, E. Mininno, F. Neri, and P. N. Suganthan, "Super-fit and population size reduction in compact differential evolution," in *IEEE Workshop on Memetic Computing (MC)* (2011), pp. 1–8.

[16] M. G. Epitropakis, D. K. Tasoulis, N. G. Pavlidis, V. P. Plagianakos, and M. N. Vrahatis, IEEE Transactions on Evolutionary Computation **15**, 99–119 (2011).

[17] R. Poláková, "Controlling population size in differential evolution by diversity mechanism," in *International Conference on Artificial Intelligence and Soft Computing* (Springer, 2017), pp. 408–417.

[18] M. Yang, C. Li, Z. Cai, and J. Guan, IEEE transactions on cybernetics **45**, 302–315 (2015).

[19] R. Reddy and G. Jeyakumar, International Journal of Applied Engineering Research **10**, 13579–13594 (2015).

[20] R. Mallipeddi and P. N. Suganthan, "Empirical study on the effect of population size on differential evolution algorithm," in *IEEE Congress on Evolutionary Computation* (IEEE, 2008), pp. 3663–3670.

[21] K. V. Price, R. Storn, and J. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization* (Springer, 2005).

[22] J. Lampinen and I. Zelinka, "Mechanical engineering design optimization by differential evolution," in *New Ideas in Optimization*, edited by D. Corne, M. Dorigo, and F. Glover (McGraw-Hill, 1999), pp. 127–146.

[23] J. Liang, B. Qu, P. Suganthan, and A. G. Hernández-Díaz, "Problem definitions and evaluation criteria for the CEC 2013 special session on real-parameter optimization," Tech. Rep. 201212 (Zhengzhou University, China and Nanyang Technological University, Singapore, 2013).

[24] F. Neri, G. Iacca, and E. Mininno, Information Sciences **181**, 2469 – 2487 (2011).

[25] K. Price, "An introduction to differential evolution," in *New Ideas in Optimization*, edited by D. Corne, M. Dorigo, F. Glover, D. Dasgupta, P. Moscato, R. Poli, and K. V. Price (McGraw-Hill, 1999), pp. 79–108.

[26] A. Yaman, G. Iacca, and F. Caraffini, A comparison of three differential evolution strategies in terms of early convergence with different population sizes, `www.tech.dmu.ac.uk/~fcaraf00/NumericalResults/DE_Early_Convergence_RESULTS.pdf` (2018).