On the Non-convergence of Differential Evolution: Some Generalized Adversarial Conditions and A Remedy

Debolina Paul
Indian Statistical Institute
Kolkata, West Bengal, India
debolinap8@gmail.com

Saptarshi Chakraborty
Indian Statistical Institute
Kolkata, West Bengal, India
saptarshichakraborty27@gmail.com

Swagatam Das
Electronics and Communication Sciences Unit,
Indian Staistical Institute
Kolkata, West Bengal, India
swagatam.das@isical.ac.in

Ivan Zelinka
Electrical Engineering and Computer Science,
Technical University of Ostrava
Ostrava, Czech Republic
ivan.zelinka@vsb.cz

ABSTRACT

In this paper, we analyze the convergence behavior of Differential Evolution (DE) and theoretically prove that under certain adversarial conditions, the generic DE algorithm may not at all converge to the global optimum even on apparently simpler fitness landscapes. We characterize these function classes and initialization conditions theoretically and provide mathematical supports to the non-convergence behavior of DE. To overcome these adversarial conditions, we propose a slightly modified variant of DE called Differential Evolution with Noisy Mutation (DENM), which incorporates a noise term in the mutation step. We analytically show that DENM can converge to the global optima within a finite budget of function evaluations.

CCS CONCEPTS

• **Theory of computation → Evolutionary algorithms; Bioinspired optimization**;

KEYWORDS

Differential Evolution, non-convergence to optima, noisy mutation, generalized adversarial conditions

ACM Reference Format:

Debolina Paul, Saptarshi Chakraborty, Swagatam Das, and Ivan Zelinka. 2019. On the Non-convergence of Differential Evolution: Some Generalized Adversarial Conditions and A Remedy. In *Proceedings of the Genetic and Evolutionary Computation Conference 2019 (GECCO '19).* ACM, New York, NY, USA, 2 pages. https://doi.org/https://doi.org/10.1145/3319619.3322007

1 INTRODUCTION

Storn and Price proposed the Differential Evolution (DE) [1, 5] algorithm in 1995 as an efficient yet simple population-based global optimization algorithm for the continuous parameter space ($\subseteq \mathbb{R}^p$). Even though the DE algorithms are quite popular, little is known about its theoretical convergence behavior [2, 4]. In this paper, we investigate by rigorously exploring the non-convergence behavior of DE for a broader class of functions and under some adversarial conditions imposed on the population initialization. We also propose a remedy to overcome such stagnation or false convergence to a non-optimal point by DE in terms of a modified mutation scheme. We theoretically analyze the convergence behavior of the proposed algorithm, referred to here as DE with Noisy Mutation (DENM).

GECCO '19, July 13–17, 2019, Prague, Czech Republic

ACM ISBN 978-1-4503-6748-6/19/07…$15.00…$15.00
https://doi.org/https://doi.org/10.1145/3319619.3322007

2 ON SOME DEROGATORY PROPERTIES OF THE MUTATION AND CROSSOVER IN DE

The following results hold true for any norm $||\cdot||$, for discussing properties of DE crossover, we in particular take the l_∞ norm. Moreover, $B_M(\mathbf{x})$ and $\bar{B}_M(\mathbf{x})$ respectively denote the open and closed ball of radius M centered at $\mathbf{x}$.

Lemma 1. Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in B_M(\mathbf{x}^*)$ be distinct. Let $\mathbf{x}' = \mathbf{x} + F(\mathbf{y} - \mathbf{z})$ then $\exists M' \geq M$ such that $\mathbf{x}' \in B_{M'}(\mathbf{x}^*)$.

Theorem 1. Let $\mathbf{x}_0, \dots, \mathbf{x}_{2k} \in \mathbf{X}$ and $\mathbf{X}$ is contained in an affine space $\mathcal{A}$ of $\mathbb{R}^p$. Then the vector $\mathbf{x} = \mathbf{x}_{i_0} + \sum_{i=1}^{n} F_i(\mathbf{x}_{2i-1} - \mathbf{x}_{2i}) \in \mathcal{A}$.

Remark 1. The binomial crossover between two vectors $\mathbf{x}$ and $\mathbf{y}$ can be represented as $f_\mathbf{a}(\mathbf{x}, \mathbf{y}) = \mathbf{a} \circ \mathbf{x} + (\mathbf{1} - \mathbf{a}) \circ \mathbf{y}$ for some $\mathbf{a} \in \{0, 1\}^p$, where $\circ$ is the Hadamard product.

Theorem 2. Let $\mathbf{x}$ and $\mathbf{y} \in \mathcal{A}$, where $\mathcal{A}$ is an axis parallel affine space of $\mathbb{R}^p$. Let $f_\mathbf{a} : \mathcal{A} \times \mathcal{A} \to \mathbb{R}^p$ be such that $f_\mathbf{a}(\mathbf{x}, \mathbf{y}) = \mathbf{a} \circ \mathbf{x} + (\mathbf{1} - \mathbf{a}) \circ \mathbf{y}, \mathbf{a} \in \{0, 1\}^p$. Then range of $f_\mathbf{a} \subseteq \mathcal{A}$.

Theorem 3. If $\mathbf{x}, \mathbf{y} \in B_M^\infty(\mathbf{z})$ and $\mathbf{a} \in \{0, 1\}^p$, then $f_\mathbf{a}(\mathbf{x}, \mathbf{y}) \in B_M^\infty(\mathbf{z}) \forall z \in \mathbb{R}^p$.

3 THEORETICAL ANALYSIS OF THE NON-CONVERGENCE OF DE

Adversarial Condition 1. Let $\mathbf{X}^{(0)} = \{\mathbf{x}_1, \dots, \mathbf{x}_{Np}\} \subset \mathcal{A}$, where $\mathcal{A}$ is an axis-parallel affine space. If $X^{(t)}, \forall t$ is updated using only mutation and crossover, then $\forall t \in \mathbb{N}$, $\mathbf{X}^{(t)} \subset \mathcal{A}$, i.e. under initialization on an axis parallel flat $\mathcal{S}$, the population cannot escape it.

Adversarial Condition 2. If the initial population is set too close to the optima and the value of the mutation parameter F is large enough, the standard DE algorithm fails to converge. In other words, if $\exists M' > M$ and $\|\mathbf{x} - \mathbf{x}^*\| > M'$, $f(\mathbf{x}) > \sup_{\mathbf{x} \in B_M(\mathbf{x}^*)} f(\mathbf{x})$, then there exists some $F_0 > 0$ such that DE will not make any changes to the population when $F > F_0$.

Adversarial Condition 3. Let us take those functions which have broad local optimal basins alongside a narrow global optima i.e. $f : \mathcal{S} \to \mathbb{R}$, $\mathcal{S} \subseteq \mathbb{R}^p$ be any function, and $\exists x^*_{local}, M, M'$. If the population is initialized sufficiently inside a local optimal basin then DE fails to reach the global optima in a finite number of iterations, under the following regularity conditions:

(1) $\exists x^*_{local} \in \mathbb{R}^p s.t. f(x^*_{global}) < f(x^*_{local})$ and $x^*_{global} \notin B^{\infty}_{M'}(x^*_{local})$.
(2) $\sup_{x \in B^{\infty}_{M}(x^*_{local})} f(x) \leq \inf_{y \in B^{\infty}_{M'}(x^*_{local}) \setminus B^{\infty}_{M}(x^*_{local})} f(y)$.
(3) $M' > M + 2MF$

In other words, if $x_i^{(0)} \in B^{\infty}_{M}(x^*_{local}) \forall i \in \{1, \ldots, Np\}$, then $x_i^{(t)} \in B_M(x^*_{local}) \forall i \in \{1, \ldots, Np\}, \forall t \in \mathbb{N} \cup \{0\}$.

4 THE PROPOSED ALGORITHM AND ITS CONVERGENCE BEHAVIOUR

We propose a remedy in form of Algorithm 1 to overcome the adversarial conditions in Section 3. Convergence of the algorithm is established through Theorems 4 and 5. The efficacy of the proposed DENM algorithm over DE/rand/1 with respect to adversarial condition 1 is illustrated in Fig. 1.

Algorithm 1: Differential Evolution with Noisy Mutation (DENM)

Input: $\mathcal{F}$, F, Cr, t_{max}. **Output**: $\mathbf{x}^{(tmax)}_{best}$ (The best member at termination).
Generate $\mathbf{x}_1, \ldots, \mathbf{x}_{Np}$ uniformly within the search range.
for *t=1 to* t_{max} **do**
  for *i= 1 to* Np **do**
    Select r_1, r_2, r_3 uniformly at random from $\{1, \ldots, Np\} \setminus \{i\}$;
    Compute $\mathbf{v}_i^{(t)} = \mathbf{x}_{r_1}^{(t)} + F(\mathbf{x}_{r_2}^{(t)} - \mathbf{x}_{r_3}^{(t)})$;
    Generate additive noise $\mathbf{e}_i^{(t)}$ from CDF $\mathcal{F}$ and let $\mathbf{w}_i^{(t)} = \mathbf{v}_i^{(t)} + \mathbf{e}^{(t)}$;
    Select j_{rand} uniformly at random from $\{1, \ldots, p\}$;
    for *j =1 to p* **do**
$$u_{j,i,t} = \begin{cases} v_{j,i,t} & \text{if } rand_{i,j}[0,1] \leq Cr \text{ or } j = j_{rand} \\ x_{j,i,t} & \text{otherwise.} \end{cases}$$
$$z_{j,i,t} = \begin{cases} w_{j,i,t} & \text{if } rand_{i,j}[0,1] \leq Cr \text{ or } j = j_{rand} \\ x_{j,i,t} & \text{otherwise.} \end{cases}$$
    end
    if $f(\mathbf{u}_i^{(t)}) \leq min\{f(\mathbf{x}_i^{(t)}), f(\mathbf{z}_i^{(t)})\}$ **then** replace $\mathbf{x}_i^{(t)}$ by $\mathbf{u}_i^{(t)}$;
    else if $f(\mathbf{z}_i^{(t)}) \leq min\{f(\mathbf{x}_i^{(t)}), f(\mathbf{u}_i^{(t)})\}$ **then** replace $\mathbf{x}_i^{(t)}$ by $\mathbf{z}_i^{(t)}$;
    else keep $\mathbf{x}_i^{(t)}$;
  end
end

Given an indicator function $I(.)$, an objective function $f(.)$, and a set $Q_\epsilon = \{\mathbf{x} | f(\mathbf{x}) - f(\mathbf{x}^*) < \epsilon\}$, the first hitting time $N(\epsilon)$ is defined as: $N(\epsilon) = \inf \left\{ t : \sum_{i=1}^{Np} I(\mathbf{x}_i^{(t)} \in Q_\epsilon) > 0 \right\}$.

We will assume the following on the objective function f.

- $\exists\, \mathbf{x}^*$ such that $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^p$.
- The set Q_ϵ contains a $B_{\delta(\epsilon)}(\mathbf{x}^*)$ for some $\delta(\epsilon) > 0$.
- The set $L_b = \{\mathbf{x} | f(\mathbf{x}) \leq b\}$ is bounded for all $b \in Range(f)$.

We can infer the DENM/rand/1 converges if at least one point of $\mathbf{X}^{(t)}$ lies in Q_ϵ. We impose the following regularity conditions on $\mathcal{F}$:

- $\mathcal{F}$ admits a Radon-Nikodym derivative (say h) with respect to the Lebesgue measure on $\mathbb{R}^p$.
- h has the property that $h(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbb{R}^p$.
- $\exists\, H > 0$ such that $h(\mathbf{x}) \leq H\ \forall \mathbf{x} \in \mathbb{R}^p$.

Theorem 4. The first hitting time $N(\epsilon)$ is finite with probability 1.

Theorem 5. The expectation of the first hitting time is finite, i.e. $\mathbb{E}(N(\epsilon)) < \infty$.

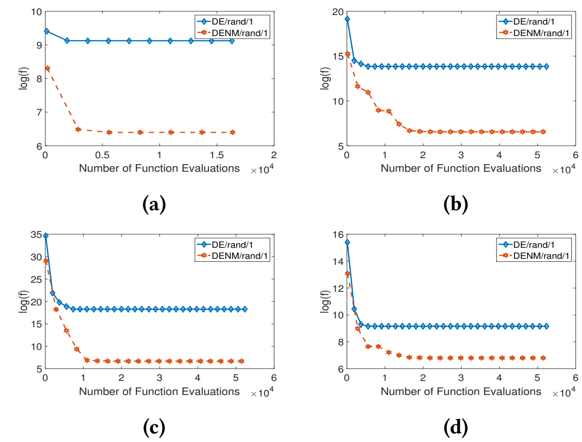

Figure 1: Experimental results on CEC 2013 functions in terms of convergence chracteristics [3] for DE/rand/1. (a) Sphere Function. (b) Rotated High Conditioned Elliptic Function. (c) Rotated Bent Cigar Function. (d) Rotated Discus Function.

5 CONCLUSION

In this paper, we theoretically established the fact that the DE algorithms might not converge for poor initialization and ill-behaved functions. To overcome these adverse situations, we proposed a remedy by adding random noise to the mutant vector and provided a detailed mathematical study on the convergence of the proposed algorithm. One may attempt to improve the bounds on the expectations of the first hitting time and study the convergence behavior of the population distribution to the optima in terms of the Wasserstein distance, as it better captures the metric properties of the underlying distribution in the context of optimal transport.

6 ACKNOWLEDGEMENT

The following grants are acknowledged for the financial support provided for this research: SGS SP2019/137, VSB-Technical University of Ostrava.

REFERENCES

[1] S. Das, S. S. Mullick, and P.N. Suganthan. 2016. Recent advances in differential evolution – An updated survey. *Swarm and Evolutionary Computation* 27 (2016), 1 – 30.
[2] J. Lampinen and I. Zelinka. 2000. On stagnation of the differential evolution algorithm. In *Proceedings of Mendel, 6th International Mendel Conference on Soft Computing*. 76 – 83.
[3] JJ Liang, BY Qu, PN Suganthan, and Alfredo G Hernández-Díaz. 2013. Problem definitions and evaluation criteria for the CEC 2013 special session on real-parameter optimization. *Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou, China and Nanyang Technological University, Singapore, Technical Report* 201212 (2013), 3–18.
[4] M. Locatelli and M. Vasile. 2015. (Non) convergence results for the differential evolution method. *Optimization Letters* 9, 3 (2015), 413–425.
[5] R. Storn and K. Price. 1995. *Differential evolution-a simple and efficient adaptive scheme for global optimization over continuous spaces.* Vol. 3. ICSI Berkeley.