Some Convergence Properties of Differential Evolution

James Montgomery
Information and Communication Technology
University of Tasmania
Hobart, Australia
ORCID 0000-0002-5360-7514

Antonio Bolufé-Röhler
Mathematical and Computational Sciences
University of Prince Edward Island
Charlottetown, Canada
ORCID 0000-0002-3181-1864

Dania Tamayo-Vera
Thinking Big Inc.
Charlottetown, Canada
ORCID 0000-0002-8682-8604

Stephen Chen
Information Technology
York University
Toronto, Canada
ORCID 0000-0001-8657-6200

Tim Hendtlass
Science, Engineering and Technology
Swinburne University of Technology
Melbourne, Australia
ORCID 0000-0002-2582-7309

***Abstract*—Differential Evolution uses fitness-based selection – a *new* solution will replace a reference *target* solution only if it is at least as fit. In a multi-modal search space, we are more interested in the (fitness of the) nearby local optimum that a search solution could lead to than its current fitness. Fitness-based selection can in fact lead to "Failed Exploration" – the rejection of a search solution from a region of the search space that could lead to fitter (local) optimum because it is less fit than the reference solution (which leads to a less fit local optimum). We conduct several experiments to observe the occurrence of Failed Exploration in Differential Evolution, and we provide some insights into its operation and convergence properties.**

***Index Terms*—Exploration, Exploitation, Multi-Modal Search Spaces, Differential Evolution**

I. Introduction

Many metaheuristics generate search solutions through the influence of reference solutions. For example, Genetic Algorithms [1] store a population of parent (reference) solutions, and offspring (search) solutions are created from the parents by applying operators such as crossover and mutation. Another example is Particle Swarm Optimization [2] in which a set of personal best (reference) positions are used to guide the search trajectories of moving particles, and these trajectories affect the sampled (search) solutions. In general, the exploratory search solutions that will be generated by a metaheuristic are strongly influenced by a set of retained reference solutions.

The progression of search in a metaheuristic depends on changes to the reference solutions. This updating of reference solutions usually involves a comparison of the current/actual fitness of a search solution with the current/actual fitness of a reference solution. If the search solution is fitter, it replaces the reference solution, and if the search solution is less fit, it is rejected. However, in multi-modal search spaces, it may be useful to retain some reference solutions in highly promising areas of the search space. In particular, a previous study of exploration and exploitation has shown that reference solutions which approach the fitness of their nearby local optima are more likely to cause the rejection of highly promising (but less fit) exploratory search solutions [3].

A metaheuristic which rejects exploratory search solutions from more promising regions of the search space will likely concentrate its search and (prematurely) converge in the less promising region of the search space represented by its existing reference solutions. The success rate that exploratory search solutions experience in their attempts to replace existing reference solutions thus has a considerable effect on the operation and performance of metaheuristics. A key determinant of this success rate is the occurrence of "Failed Exploration" – the case when an exploratory search solution from a more promising region of the search space is rejected because its current/actual fitness is less than that of a reference solution from a less promising region of the search space [4].

This paper presents some observations of Failed Exploration during the convergence of Differential Evolution (DE) [5]. First, precise definitions for exploration and exploitation are reviewed as part of the background in Section II. Initial experiments in Section III show the (relative) fitness profile of a sinusoidal attraction basin. The relative fitness of exploratory search solutions for DE on the Rastrigin function (which has sinusoidal attraction basins) is examined in Section IV. The paper closes with a discussion in Section V.

II. Background

The ability to measure Failed Exploration depends on precise definitions for exploration and exploitation. We begin by defining a multi-modal search space to consist of attraction basins, each with a single local optimum. An attraction basin around an optimum includes all the points in the search space that would reach that optimum by following a path on which every point has a monotonically decreasing fitness (for minimization problems) or increasing fitness (for maximization problems). We further define that moving a search solution within the same attraction basin as its reference solution

represents *exploitation*, and conversely that moving a search solution into a different attraction basin represents *exploration*.

Exploitation, as defined above, allows for a simple comparison of the current fitness of the search solution and the current fitness of the reference solution. However, this same comparison for exploration can lead to two types of errors. These two errors arise from the four possible cases that can occur based on the fitness of the exploratory search solution, the reference solution, and the local optima of their respective attraction basins. For brevity, we will refer to the fitness of the local optimum of an attraction basin as the fitness of that attraction basin.

Case one is "Successful Exploration": an exploratory search solution from a fitter attraction basin is compared against a less fit reference solution that represents a less fit attraction basin. Based on its superior fitness, most metaheuristics will accept the new search solution and use it to replace the existing reference solution. Case two is "Successful Rejection": a less fit search solution from a less fit attraction basin is rejected after it is compared against a fitter reference solution from a fitter attraction basin. Case three is "Deceptive Exploration": an exploratory search solution from a less fit attraction basin is accepted because it is fitter than a reference solution from a fitter attraction basin. Case four is "Failed Exploration": an exploratory search solution from a fitter attraction basin is rejected because it is less fit than a reference solution from a less fit attraction basin [4].

We define a solution with a fitness similar to that of its attraction basin to have exceptional "relative fitness". Previous research in [3] has demonstrated that the probability of Failed Exploration increases as the relative fitness of the reference solution improves (i.e. approaches the local optimum). The relative fitness of a reference solution is improved by exploitation, and metaheuristics such as Differential Evolution which allow (and even encourage) exploitation in the attraction basins of existing reference solutions often have reference solutions of exceptional relative fitness. As the fitness of the reference solution improves, the proportion of other attraction basins that represent exploratory search solutions with better fitness will decrease, and this leads to increased rates of Failed Exploration [3], [4].

To make precise measurements of exploration, it is necessary to be able to determine the attraction basin that is associated with every solution in the search space. For general search spaces (e.g. real-world problems), this can often be impossible or computationally infeasible. These experiments are thus limited to artificial search spaces in which the attraction basin and its fitness can be easily determined for every solution. One such function is the Rastrigin function shown in Equation 1. As first described by Gonzalez-Fernandez and Chen [3], this function has a regular fitness landscape in which every point with integer values in all dimensions is a local optimum, and every other point belongs to the attraction basin of the local optimum that is determined by rounding each solution term to its nearest integer value. These features make it possible to quickly and easily calculate both the fitness of a solution and the fitness of the local optimum of its attraction basin, i.e. what we call the fitness of an attraction basin.

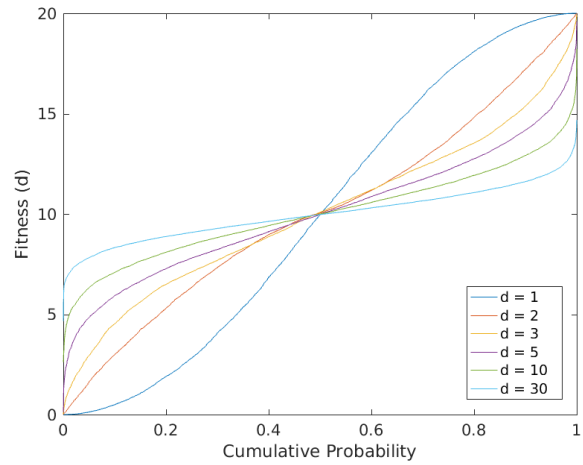


Fig. 1: Cumulative distribution functions for 10,000 uniform random points in sinusoidal attraction basins of $d = 1, 2, 3, 5, 10, 30$ dimensions. As dimensionality increases, the recorded distribution becomes more vertical at the edges – indicating that the proportion of solutions with exceptional relative fitness (close to zero) reduces drastically.

$$f(x) = 10d + \sum_{i=1}^{d} \left(x_i^2 - 10cos(2\pi x_i)\right) \tag{1}$$

The experiments in this paper are based on a DE/rand/1/bin version of DE with typical parameters of population size p = 50, crossover Cr = 0.9, and scale factor F = 0.8 [5], [6]. In each iteration, each population member x_i is considered as a *target* for replacement by a candidate solution that is constructed in two steps: creation of an intermediate solution and crossover with x_i. During the creation of an intermediate solution y_i from three distinct random solutions r_1, r_2, and r_3 in Equation 2, the scale factor F affects the "step size" from r_1 taken in the direction of the "difference vector" created with r_2 and r_3.

$$y_i = r_1 + F(r_2 - r_3) \tag{2}$$

In Equation 3, this intermediate solution is then crossed term-by-term in each dimension d of the search space with the *target* solution x_i to produce the *new* candidate solution x'_i.

$$x'_{i,d} = \begin{cases} y_{i,d} & u_d \leq Cr \\ x_{i,d} & u_d > Cr \end{cases} \tag{3}$$

III. Experiment 1 – Uniform Random Sampling of Sinusoidal Attraction Basins

We start with a search space of d dimensions where $d = 1, 2, 3, 5, 10, 30$. The range in each dimension of the search space is –0.5 to +0.5, and the fitness function for this

search space is as given in Equation 1. We then generate 10,000 uniform random points in this search space and record the fitness of each point. Fig. 1 shows the cumulative distribution functions (CDFs) for uniform random points in each attraction basin that we tested. Note: we switch the axes used in typical CDF plots since (relative) fitness on the y-axis is useful for comparisons in subsequent figures.

In $d = 1$ dimensions, uniform random sampling leads to good coverage of the search range – some points near the local optimum in the middle of the attraction basin which lead to exceptional fitness and some points near the edges of the attraction basin which lead to poor fitness. The actual fitness range is 0.00 to 20.0 from a possible range of 0 to 20. As dimensionality increases, a point near the middle of the attraction basin with exceptional fitness requires all dimensions to independently pick a uniform random value near the middle, and the probability of picking such a point decreases exponentially with increasing dimensionality. Similarly, the probability of a point near the edges with low fitness also decreases exponentially with increasing dimensionality. For $d = 30$, it can be seen that the majority of the 10,000 points have an "average" fitness with the actual fitness range being 140 to 440 from a possible range of 0 to 600.

The CDFs of the fitness distributions in Fig. 1 clearly indicate a trend that increasing dimensionality can lead to attraction basins with a smaller proportion of solutions with exceptional fitness. Since "Successful Exploration" depends on finding an exploratory search solution with a better fitness than its corresponding reference solution, this decreasing proportion of solutions with exceptional fitness can lead to more "Failed Exploration". This conjecture naturally depends on the distributions for exploratory search solutions in metaheuristics to be similar to those of uniform random points, and this has previously been shown for Particle Swarm Optimization [7].

IV. Experiment 2 – Relative fitness of exploratory solutions in DE

In Fig. 2, we show the cumulative distribution functions for the relative fitness of both exploratory and exploitative search solutions for the first, middle, and last 10,000 solutions in DE for $d = 1, 2, 3, 5, 10, 30$ dimensions when $10{,}000d$ total function evaluations are used. In Fig. 3, we plot the relative fitness of every exploratory search solution that represents either successful exploration or failed exploration for the same execution of DE. For $d = 1, 2, 3, 5$, all trials converge to the global optimum. For $d = 10$, four trials converge (but not to the global optimum), 10 trials display convergence similar to $d = 30$, and 16 trials show no convergence. For the first five sub-figures, we have chosen the slowest converging and/or worst performing of 30 independent trials because these trials show more data/insight into the operation of DE – especially into its potential failure mechanisms. For $d = 30$, we have shown the best performing trial (one of the two that reaches convergence) to allow for the greatest contrast between the data for $d = 10$ and $d = 30$.

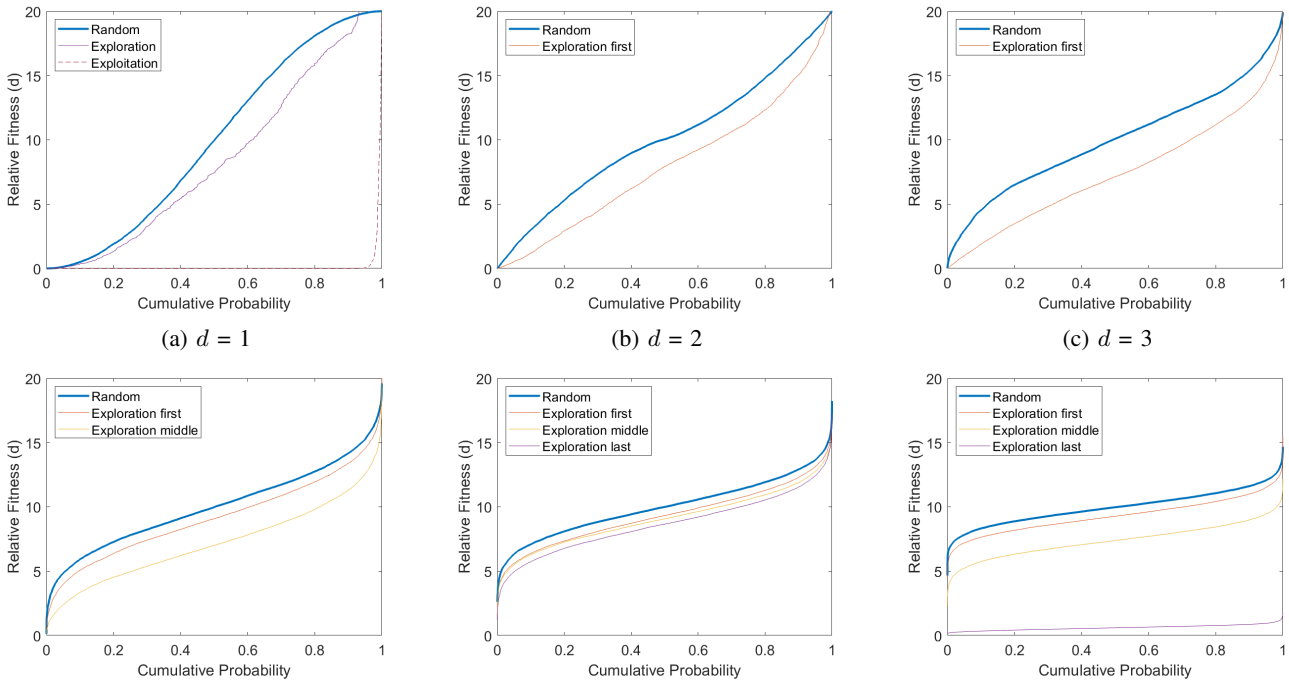
For $d = 1$, the first, middle, and last 10,000 search solutions are identically the same in Fig. 2a. For $d = 2, 3$, all trials converge before the middle 10,000 solutions, so only one plot line is available. The exploitative search solutions in DE also tend to have exceptional relative fitness as indicated by plot lines hugging the x-axis, so CDFs for exploitative solutions are again not shown for $d > 1$ to improve the readability of the remaining sub-figures. In all of the sub-figures, the CDFs for uniform random points in a sinusoidal attraction basin are replicated from Fig. 1 for reference.

The CDFs for relative fitness of exploratory search solutions in DE are similar to those for random points in the base sinusoidal search space. Considering the relative fitness of individual solutions representing Successful Exploration or Failed Exploration shown in Fig. 3, the selected DE trial for $d = 10$ (Fig. 3e) exhibits a "stall" pattern more common to PSO [7]: a band of solutions with "average" relative fitness that lead to Failed Exploration, and a decreasingly small number of solutions able to achieve Successful Exploration.

However, an alternate pattern also exists for DE in which the likelihood of a highly relatively fit exploratory solution appears to *increase* as the DE search progresses. This shift can be observed in the longer runs for $d \geq 5$ (Figs. 2d–2f and also Figs. 3d and 3f) and is notable because it means the DE search is generating solutions in different attraction basins (i.e. performing exploration) that are relatively near each basin's local optimum. Two aspects of DE's search behaviour likely explain this result. The first is that DE may be exploiting the search space's separability and regularity, with the population becoming aligned along a subset of axes, allowing (relatively large) difference vectors to generate points in better attraction basins in the same axis. However, for any scale factor $F < 1$ this mechanism will eventually lead to an increase in Failed Exploration as new solutions fall short of the region within each basin of attraction where the selection mechanism would accept them.

The second aspect of DE's behaviour that can explain exploratory solutions approaching the value of local optima is that, as exploitative moves cause clusters of solutions to contract around their respective local optima, very short difference vectors can be generated and applied to *base* solutions (i.e., r_1 in Equation 2) that are in different basins to the *target* for replacement. This phenomenon, first documented by Montgomery [8], leads to solutions "moving" a large distance in the search space as they are replaced by *new* solutions that are both near other population members *and* near to a local optimum. This explains the observed good performance in the selected trial for $d = 30$ (Fig. 3f).

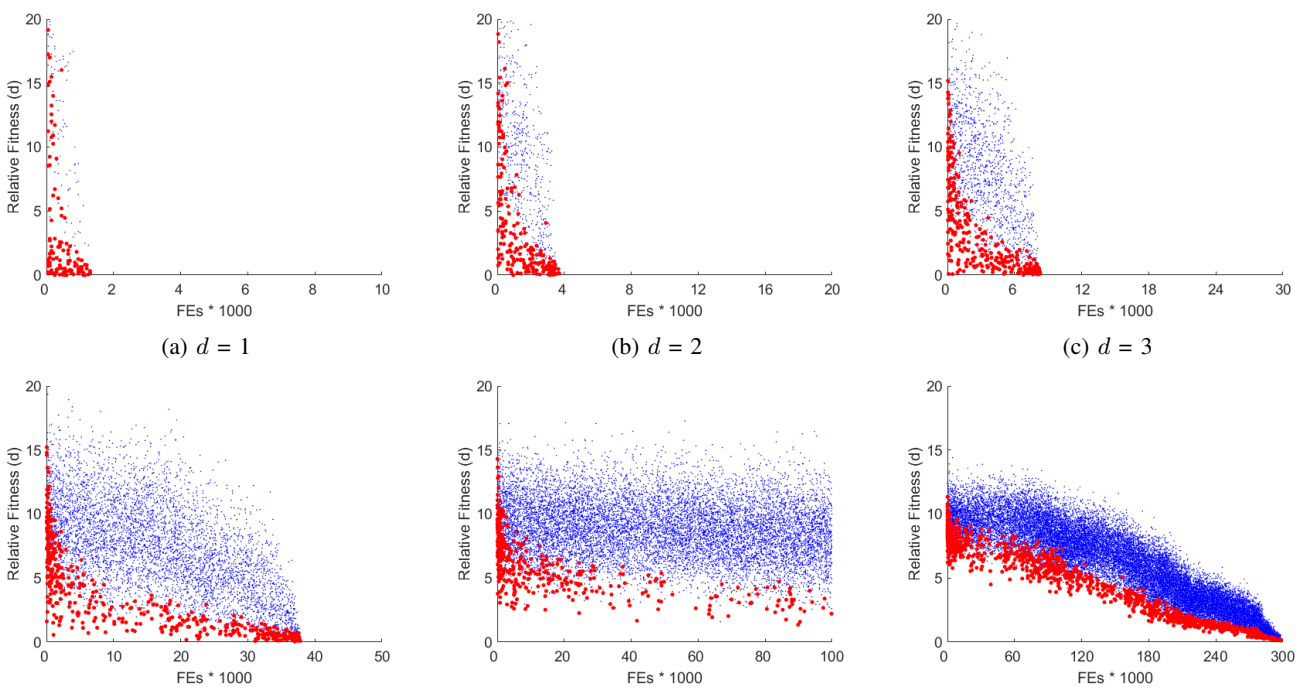
Regardless of the mechanisms involved, in both the cases of poor search performance ($d = 1, 2, 3, 5, 10$) and the selected example of good performance ($d = 30$), as the search progresses the quality of exploratory solutions relative to that of their local optimum needs to be exceptional if they are to survive selection. In the chosen example of good search performance, near the end of the search even exploratory search solutions with exceptional relative fitness were rejected



(a) $d = 1$ (b) $d = 2$ (c) $d = 3$

(d) $d = 5$ (e) $d = 10$ (f) $d = 30$

Fig. 2: CDFs of exploratory (solid) and exploitative (dashed) search solutions for DE on Rastrigin in $d = 1, 2, 3, 5, 10, 30$ (those for $d < 30$ are the slowest converging trials, while $d = 30$ is the best performing trial). Both sets of lines tend to get fitter (closer to the x-axis) as search progresses. The CDF for uniform random points (shown in bold) is replicated from Fig. 1



(a) $d = 1$ (b) $d = 2$ (c) $d = 3$

(d) $d = 5$ (e) $d = 10$ (f) $d = 30$

Fig. 3: Relative fitness of exploratory search solutions for DE on Rastrigin in $d = 1, 2, 3, 5, 10, 30$. Small blue dots represent Failed Exploration and large red dots represent Successful Exploration. As search progresses, the red dots occupy a smaller range at the bottom of the cloud of blue dots.

by the algorithm. Consequently, these results suggest that the fitness profile of attraction basins in a multi-modal search space can make it easier or harder for DE to identify and keep solutions in promising regions of the search space.

V. Discussion

In low dimensions, a sinusoidal attraction basin can allow exploratory search solutions with exceptional relative fitness to occur frequently enough to allow sufficiently high rates of Successful Exploration to reach the global optimum – even when interim reference solutions reach local optima with zero relative fitness. However, it is shown that the relative fitness of exploratory search solutions in sinusoidal attraction basins can become more and more tightly clustered around the mean height of the attraction basin as dimensionality increases. A greatly reduced probability to have exploratory search solutions of exceptional relative fitness can become an issue if two factors are present: (1) rejection of search solutions is based on a comparison of their current fitness with a stored reference solution, and (2) the stored reference solutions can have exceptional relative fitness.

These two factors are present in many metaheuristics. Fitness-based selection (i.e. "survival of the fittest" is specifically used in DE. Due to this fitness-based selection, the fitness of the stored reference solutions will increase over time, and much of this fitness improvement is due to improvements in relative fitness – i.e. exploitation of existing reference solutions leads them to approach their local optima. The second factor is thus also present in DE, and the combination of these two factors can lead to increasing rates of Failed Exploration with increasing search space dimensionality.

References

[1] J. H. Holland, "Adaptation in natural and artificial systems. an introductory analysis with applications to biology, control and artificial intelligence," *Ann Arbor: University of Michigan Press, 1975*, 1975.

[2] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 4. IEEE, 1995, pp. 1942–1948.

[3] Y. Gonzalez-Fernandez and S. Chen, "Leaders and followers — a new metaheuristic to avoid the bias of accumulated information," in *Evolutionary Computation (CEC), 2015 IEEE Congress on*. IEEE, 2015, pp. 776–783.

[4] S. Chen, A. Bolufé-Röhler, J. Montgomery, and T. Hendtlass, "An analysis on the effect of selection on exploration in particle swarm optimization and differential evolution," in *Evolutionary Computation (CEC), 2019 IEEE Congress on*. IEEE, 2019, pp. 3038–3045.

[5] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *Journal of global optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[6] J. Montgomery and S. Chen, "An analysis of the operation of differential evolution at high and low crossover rates," in *Evolutionary Computation (CEC), 2010 IEEE Congress on*. IEEE, 2010, pp. 1–8.

[7] S. Chen, I. Abdulselam, N. Yadollahpour, and Y. Gonzalez-Fernandez, "Particle swarm optimization with pbest perturbations," in *2020 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2020, pp. 1–8.

[8] J. Montgomery, "Differential evolution: Difference vectors and movement in solution space," in *Evolutionary Computation, 2009. CEC'09. IEEE Congress on*. IEEE, 2009, pp. 2833–2840.