Steady Success Clusters in Differential Evolution

Adam Viktorin, Roman Senkerik, Michal Pluhacek
Faculty of Applied Informatics
Tomas Bata University in Zlin
T. G. Masaryka 5555, 760 01 Zlin, Czech Republic
Email: {aviktorin, senkerik, pluhacek}@fai.utb.cz

Aleš Zamuda
Faculty of Electrical Engineering and Computer Science
University of Maribor
Smetanova 17, 2000 Maribor, Slovenia
Email: ales.zamuda@um.si

***Abstract*—This paper presents insights into the proportions between the k-means clusters of successful Differential Evolution (DE), donor generating vectors. This is demonstrated by the high certainty that these proportions are similar – and thereby, that these cluster size proportions regularly appear. A characteristic of these proportions is that they are observed at the same specific values in different test functions. It is also shown that, when varying the number of dimensions for a fitness function, the proportions are constant. However, some of the possible dynamics of these proportions are reported later on, in the situation where the optimization algorithm is changed – for instance, control parameters like the population size parameter. This parameter significantly changes, the proportions of the most frequently successful and most unsuccessful vectors. Insights like this are useful for an understanding of the inter-generational complexity that appears within evolutionary algorithms and would thus benefit future algorithm design; for example, plausible metrics for on-line control.**

***Index Terms*—Differential Evolution, Clustering, Complex Network, Success History, k-means.**

I. Introduction

This paper provides an insight into Differential Evolution (DE) [1] donor vector generation of successful indexes cluster, and choosing and testing the following hypotheses: 1) The ratios between the success of clusters sizes, (most successful, averagely successful, less successful), are statistically insignificant from one function to another, with the same DE settings; 2) The ratios between clusters are also statistically insignificant when compared to different dimensions of the same function; 3) The dependence on population size is occasionally statistically significant for all these functions, with different dimensions; and so as to obtain insights into information about inter-generational complexity that appears in evolutionary algorithms.

The motivation behind this, is the aim of implementing such reported insights into DE algorithm design in order to contribute to performance, e.g. for on-line control; or, possibly as a population size control mechanism; or, control of exploration/exploitation phases; or, historical archive management; and many more possibilities.

The following section presents work related to DE. Section 3 defines the success clusters for DE used to test the hypotheses. Section 4, presents reports and discussions of the experimental results; while Section 5 is the Conclusion, with suggestions for future work.

II. Related Work

DE [1]–[5], is a well-known evolutionary computation technique for continuous optimization purposes. DE has been modified and extended several times by means of new proposals of versions; and the performances of different DE variant instance algorithms have been widely studied and compared with other evolutionary algorithms - including in various major scientific conference competitions; where, over recent decades, DE has won almost all of the evolutionary algorithm competitions [6]–[18], as well as being applied to several applications [19], [20].

However, there are still plenty of unanswered challenges needing to be tackled - like, understanding control parameters [21] - and their on-line effects in DE [22]; and especially, due to the existing abundant room-for-improvement - and this merely in the adaptation possibilities of the search for the best DE optimization schemes [23]. An *EA behavior descriptor* (sometimes called observable or monitor) - as suggested by [21], and exemplified for DE in [22], reports insights like this and is useful for an understanding of the inter-generational complexity that appears in evolutionary algorithms, and would thus benefit future optimization algorithm design, possibly as metrics plausible for the on-line control of operating mechanisms and their parameters.

The basic concept of DE is to work with a randomly initialized population of "vectors" - also known as "candidate solutions", and which - in an evolutionary manner, produce better solutions in future generations. This is due to mutation, crossover, and elitism. In order to control the evolutionary process, DE uses four control parameters – number of generations $G_{\max}$, population size NP, crossover rate CR, and scaling factor F. Mutation, crossover, and elitism operations are repeated NP times to produce the next generation $G + 1$ of candidate solutions until the final generation $G_{\max}$, is reached.

1) Initialization: The control parameters are set by the DE user, and the initial population of NP candidate solutions is randomly generated from an objective space.

2) Mutation: The mutation strategy used in canonical DE is "rand/1" and uses three mutually different donor vectors at indexes r_1, r_2, and r_3 (w.r.t. $r_1 \neq r_2 \neq r_3 \neq i$ for vectors at $\forall i \in \{1, 2, ..., NP\}$) from the current generation G population. The three vectors at those indexes, using scaling factor F, are combined to produce mutated vector $\boldsymbol{v}_{i,G}$:

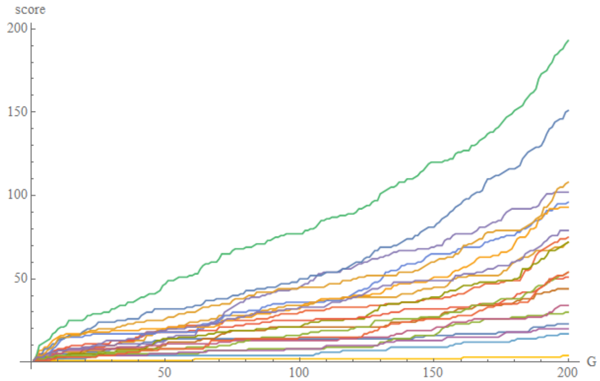


Fig. 1: Population score history for Schwefel's function, $D = 10$, $NP = 20$.

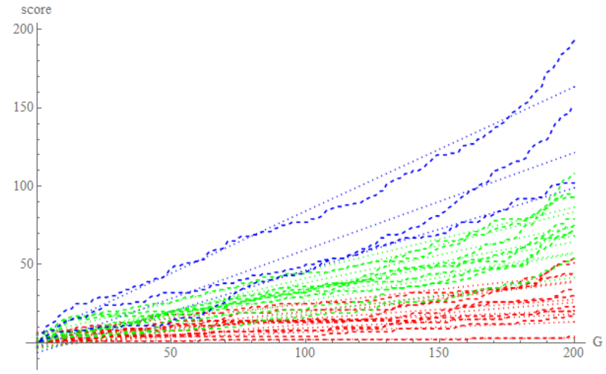


Fig. 2: Clustering of population (colors denote clusters) for Schwefel's function with $D = 10$, for the accumulating success scores at $NP = 20$ population vector indexes.

$$\boldsymbol{v}_{i,G} = \boldsymbol{x}_{r_1,G} + F\left(\boldsymbol{x}_{r_2,G} - \boldsymbol{x}_{r_3,G}\right). \tag{1}$$

3) Crossover and Elitism: Binomial crossover with the help of the CR value is used to produce trial vector $\boldsymbol{u}_{i,G}$ out of the original vector $\boldsymbol{x}_{i,G}$ and trial vector $\boldsymbol{v}_{i,G}$, where index j is an index of a component of a D-dimensional vector, $\forall j \in \{1, 2, ..., D\}$, $U[0,1]$ and a floating point uniform random number generator between 0 and 1, and where j_{rand} is a randomly-generated index of a component which has to be selected from the mutated vector:

$$u_{j,i,G} = \begin{cases} v_{j,i,G} & \text{if } U[0,1] \leq CR \text{ or } j = j_{\text{rand}} \\ x_{j,i,G} & \text{otherwise} \end{cases}. \tag{2}$$

The objective function value of the trial vector, $f(\boldsymbol{u}_{i,G})$, is then compared with the objective function value of original vector, $f(\boldsymbol{x}_{i,G})$; and if it is lower (in case of minimization), the trial vector demonstrates elitism and is placed in the next generation $G + 1$; otherwise, the original vector $\boldsymbol{x}_{i,G}$ survives to the next generation:

$$\boldsymbol{x}_{i,G+1} = \begin{cases} \boldsymbol{u}_{i,G} & \text{if } f(\boldsymbol{u}_{i,G}) < f(\boldsymbol{x}_{i,G}) \\ \boldsymbol{x}_{i,G} & \text{otherwise} \end{cases}. \tag{3}$$

III. Success Clusters in Differential Evolution

In order to implement success-based clustering, the score had to be added to the canonical DE algorithm. Each individual in the population was extended by a score value and this score value was set at 0 in the initialization phase. Whenever trial vector $\boldsymbol{u}_{i,G}$ demonstrated elitism, the score of three donor generating vectors $\boldsymbol{x}_{r_1,G}$, $\boldsymbol{x}_{r_2,G}$, and $\boldsymbol{x}_{r_3,G}$ was increased by 1 point. The score of each individual in the population was recorded in each generation, thus creating a population score history of the generations. An example of the score history can be seen in Fig. 1.

A. Data preparation

Simple linear regression was performed for each individual score history, so as to obtain its slope value; which was then used as an input to the clustering algorithm. The example vector of 20 slope values for the data depicted in Fig. 1, is as follows: (0.0702333, 0.311772, 0.139396, 0.274322, 0.526086, 0.1921, 0.0659274, 0.0115435, 0.0891675, 0.304226, 0.185942, 0.423253, 0.358715, 0.12338, 0.792863, 0.620304, 0.405396, 0.224345, 0.210787, 0.336079).

B. Clustering and the k-means algorithm

Clustering of the population was performed using the k-means algorithm and linear regression slopes, (using Wolfram language `NonlinearModelFit` function), of the score history of each individual in a population and then used as an input.

In an attempt to avoid any unnecessary correlations between the clustering mechanism and the resulting sizes of clusters, one of the simplest clustering methods was used – the k-means algorithm, proposed by Lloyd in 1957 – but only published in 1982 [24]. For a recent survey on this topic see [25].

As previously mentioned, the score history slopes of each individual in the population were used as a set of observations $(x_1, x_2, ..., x_{NP})$, and then divided into three clusters:

- **Successful** – The cluster with the higher slope values (most successful individuals in the population)
- **Average** – The cluster with average slope values
- **Unsuccessful** – The cluster with the lower slope values (least successful individuals in population).

The pseudo-code of the implementation of the k-means algorithm used in this paper is depicted in Algorithm 1. An example of the clustering results for the data from Fig. 1 is depicted in Fig. 2. All score values of an individual in the first 200 generations for the linear regression were also used. So the result of the linear regression is:

$$score = a * G + b\,, \tag{4}$$

Where, a is the slope value used in clustering (as described in the previous subsection); G is the generation number; and b is the line offset - which was neglected for this clustering (Note: In Fig. 2, the regressed lines are drawn with the offset

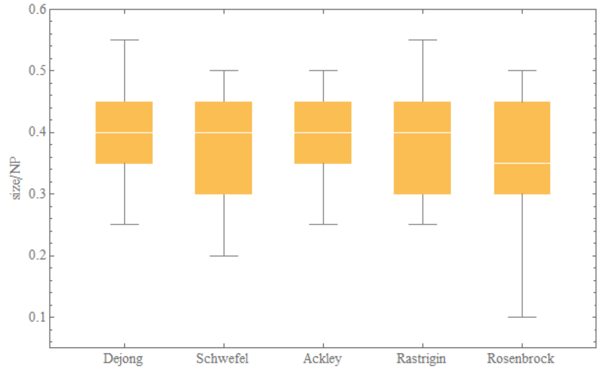


Fig. 3: Unsuccessful cluster size statistics for 5 test functions with $D = 5$ and $NP = 20$

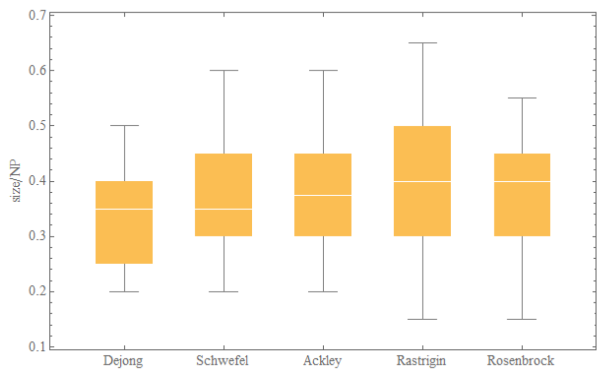


Fig. 4: Average cluster size statistics for 5 test functions with $D = 5$ and $NP = 20$.

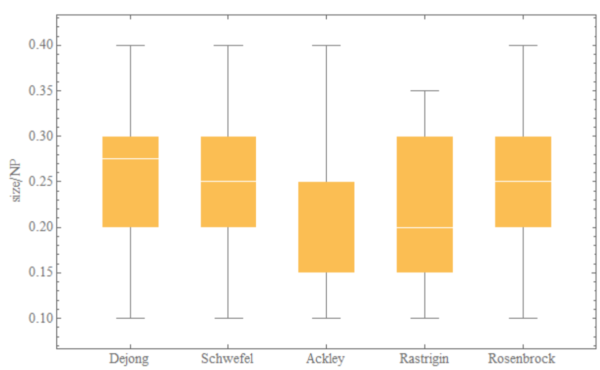


Fig. 5: Successful cluster size statistics for 5 test functions with $D = 5$ and $NP = 20$.

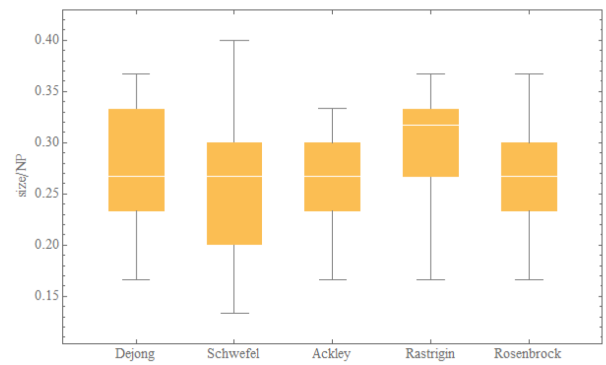


Fig. 6: Unsuccessful cluster size statistics for 5 test functions with $D = 20$ and $NP = 30$.

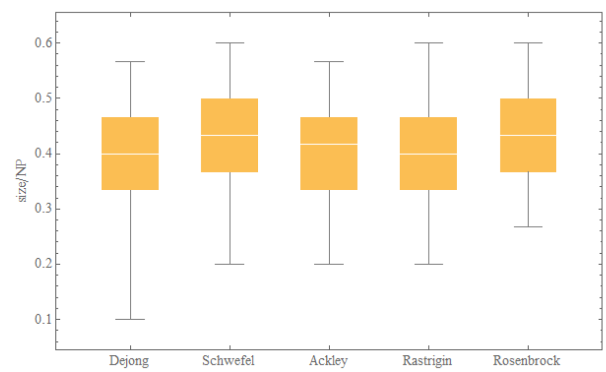


Fig. 7: Average cluster size statistics for 5 test functions with $D = 20$ and $NP = 30$.

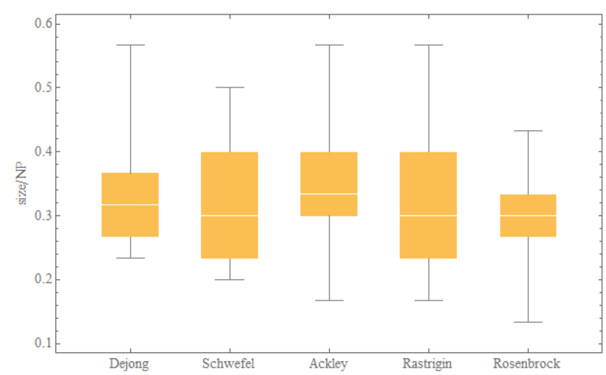


Fig. 8: Successful cluster size statistics for 5 test functions with $D = 20$ and $NP = 30$.

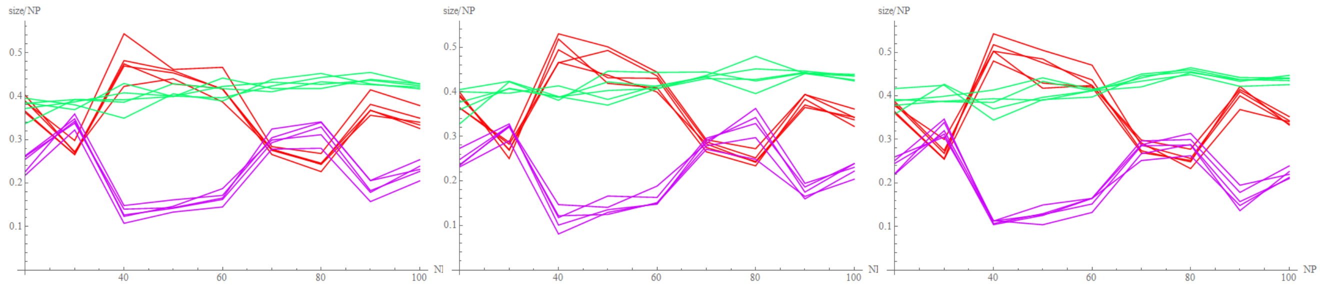
b, hence the signs of minor overlapping on the left-hand side of the image).

In order to obtain basic statistics, clustering was performed for 30 independent runs of 5 test functions – (De Jong's, Ackley's, Schwefel's, Rastrigin's, and Rosenbrock's); in three dimensional settings ($D = 5$, $D = 10$, or $D = 20$); and for two sizes of population: ($NP = 20$ or $NP = 30$).

While the experiment setup might seem simple, the results obtained in this way are valuable for further analysis and for a deeper understanding of population clustering in DE. Perhaps, what one would initially intuit without conducting experiments, is that due to the uniform selection of r_1, r_2, r_3, the donor generated successful indices would be distributed evenly (uniformly) through the index space. Or, on the basis of a more educated guess, one would perhaps argue that most improvements might be clustered around the current best vectors (that is why they would form clusters in the index space). But these guesses would be inaccurate, since it would be hard to imagine that these cluster proportions would be non-uniform and stable with specific values per function (their proportions re-appear steadily) – as reported in the next section.

IV. Results

For demonstrative purposes, and in order to support the above-mentioned hypotheses, analyses and conclusions, two



(a) $D = 5$. (b) $D = 10$. (c) $D = 20$.

Fig. 9: Successful cluster size ratios with varying NP over values $\{20, 30, 40, ..., 100\}$ for 5 test functions with $D = \{5, 10, 20\}$. The lines colors denote successful (purple color), average (green color), and unsuccessful (red color) clusters.

TABLE I: Wilcoxon signed-rank test p-values, $D = 5$, $NP = 20$.

	De Jong	Schwefel	Ackley	Rastrigin	Rosenbrock
1. cluster - Unsuccessful					
De Jong	**1.**	**0.076**	**0.508**	**0.516**	**0.136**
Schwefel	-	**1.**	**0.217**	**0.147**	**0.773**
Ackley	-	-	**1.**	**0.807**	**0.36**
Rastrigin	-	-	-	**1.**	**0.76**
Rosenbrock	-	-	-	-	**1.**
2. cluster - Average					
De Jong	**1.**	**0.111**	**0.13**	0.025	**0.174**
Schwefel	-	**1.**	**1.**	**0.873**	**0.764**
Ackley	-	-	**1.**	**0.962**	**0.689**
Rastrigin	-	-	-	**1.**	**0.258**
Rosenbrock	-	-	-	-	**1.**
3. cluster - Successful					
De Jong	**1.**	**0.516**	**0.091**	0.003	**0.676**
Schwefel	-	**1.**	**0.113**	**0.057**	**0.409**
Ackley	-	-	**1.**	**0.797**	0.048
Rastrigin	-	-	-	**1.**	0.016
Rosenbrock	-	-	-	-	**1.**

TABLE II: Wilcoxon signed-rank test p-values, $D = 5$, $NP = 30$.

	De Jong	Schwefel	Ackley	Rastrigin	Rosenbrock
1. cluster - Unsuccessful					
De Jong	**1.**	**0.817**	**0.952**	0.036	**0.837**
Schwefel	-	**1.**	**0.712**	0.012	**0.777**
Ackley	-	-	**1.**	**0.064**	**0.796**
Rastrigin	-	-	-	**1.**	0.033
Rosenbrock	-	-	-	-	**1.**
2. cluster - Average					
De Jong	**1.**	**0.918**	**0.269**	**0.966**	**0.973**
Schwefel	-	**1.**	**0.321**	**0.741**	**0.793**
Ackley	-	-	**1.**	**0.736**	**0.434**
Rastrigin	-	-	-	**1.**	**0.778**
Rosenbrock	-	-	-	-	**1.**
3. cluster - Successful					
De Jong	**1.**	**0.876**	**0.228**	**0.3**	**0.676**
Schwefel	-	**1.**	**0.516**	**0.466**	**0.904**
Ackley	-	-	**1.**	**0.139**	**0.31**
Rastrigin	-	-	-	**1.**	**0.374**
Rosenbrock	-	-	-	-	**1.**

TABLE III: Wilcoxon signed-rank test p-values, $D = 10$, $NP = 20$.

	De Jong	Schwefel	Ackley	Rastrigin	Rosenbrock
1. cluster - Unsuccessful					
De Jong	**1.**	**0.29**	**0.509**	**0.116**	**0.932**
Schwefel	-	**1.**	**0.492**	**0.92**	**0.501**
Ackley	-	-	**1.**	**0.277**	**0.99**
Rastrigin	-	-	-	**1.**	**0.205**
Rosenbrock	-	-	-	-	**1.**
2. cluster - Average					
De Jong	**1.**	0.024	**0.065**	0.025	**0.242**
Schwefel	-	**1.**	**0.597**	**0.949**	**0.208**
Ackley	-	-	**1.**	**0.609**	**0.492**
Rastrigin	-	-	-	**1.**	**0.166**
Rosenbrock	-	-	-	-	**1.**
3. cluster - Successful					
De Jong	**1.**	0.032	0.03	**0.054**	**0.158**
Schwefel	-	**1.**	**0.936**	**1.**	**0.385**
Ackley	-	-	**1.**	**0.951**	**0.337**
Rastrigin	-	-	-	**1.**	**0.563**
Rosenbrock	-	-	-	-	**1.**

TABLE IV: Wilcoxon signed-rank test p-values, $D = 10$, $NP = 30$.

	De Jong	Schwefel	Ackley	Rastrigin	Rosenbrock
1. cluster - Unsuccessful					
De Jong	**1.**	**0.141**	**0.274**	0.044	**0.331**
Schwefel	-	**1.**	**0.402**	**0.86**	**0.368**
Ackley	-	-	**1.**	**0.51**	**0.909**
Rastrigin	-	-	-	**1.**	**0.428**
Rosenbrock	-	-	-	-	**1.**
2. cluster - Average					
De Jong	**1.**	**1.**	**0.399**	**0.232**	**0.569**
Schwefel	-	**1.**	**0.608**	**0.247**	**0.837**
Ackley	-	-	**1.**	**0.847**	**0.726**
Rastrigin	-	-	-	**1.**	**0.691**
Rosenbrock	-	-	-	-	**1.**
3. cluster - Successful					
De Jong	**1.**	**0.167**	**0.727**	**0.871**	**0.781**
Schwefel	-	**1.**	**0.106**	**0.131**	**0.123**
Ackley	-	-	**1.**	**0.97**	**0.936**
Rastrigin	-	-	-	**1.**	**0.981**
Rosenbrock	-	-	-	-	**1.**

box-plot triplets are depicted in Figs. 3–8. These box-plot triplets show the proportional clusters sizes – relative to the NP for 30 independent heuristics runs; all 5 test functions; and 2 different experiment types involving opposite boundary

TABLE V: Wilcoxon signed-rank test p-values, $D = 20$, $NP = 20$.

	De Jong	Schwefel	Ackley	Rastrigin	Rosenbrock
1. cluster - Unsuccessful					
De Jong	**1.**	**0.273**	**0.262**	**0.903**	**0.689**
Schwefel	-	**1.**	**1.**	**0.579**	**0.352**
Ackley	-	-	**1.**	**0.416**	**0.394**
Rastrigin	-	-	-	**1.**	**1.**
Rosenbrock	-	-	-	-	**1.**
2. cluster - Average					
De Jong	**1.**	**0.406**	**0.855**	**0.567**	**0.289**
Schwefel	-	**1.**	**0.314**	**0.254**	**0.076**
Ackley	-	-	**1.**	**0.637**	**0.432**
Rastrigin	-	-	-	**1.**	**0.627**
Rosenbrock	-	-	-	-	**1.**
3. cluster - Successful					
De Jong	**1.**	**0.869**	**0.196**	**0.319**	**0.111**
Schwefel	-	**1.**	**0.068**	**0.136**	**0.061**
Ackley	-	-	**1.**	**0.613**	**0.673**
Rastrigin	-	-	-	**1.**	**0.546**
Rosenbrock	-	-	-	-	**1.**

TABLE VI: Wilcoxon signed-rank test p-values, $D = 20$, $NP = 30$.

	De Jong	Schwefel	Ackley	Rastrigin	Rosenbrock
1. cluster - Unsuccessful					
De Jong	**1.**	**0.182**	**0.119**	0.031	**0.397**
Schwefel	-	**1.**	**0.832**	0.003	**0.361**
Ackley	-	-	**1.**	0.005	**0.344**
Rastrigin	-	-	-	**1.**	0.022
Rosenbrock	-	-	-	-	**1.**
2. cluster - Average					
De Jong	**1.**	**0.174**	**0.979**	**0.862**	**0.056**
Schwefel	-	**1.**	**0.414**	**0.183**	**0.982**
Ackley	-	-	**1.**	**0.614**	**0.138**
Rastrigin	-	-	-	**1.**	**0.151**
Rosenbrock	-	-	-	-	**1.**
3. cluster - Successful					
De Jong	**1.**	**0.303**	**0.289**	**0.117**	**0.104**
Schwefel	-	**1.**	**0.31**	**0.673**	**0.523**
Ackley	-	-	**1.**	**0.292**	0.034
Rastrigin	-	-	-	**1.**	**0.76**
Rosenbrock	-	-	-	-	**1.**

scenarios. Figs. 3–5 are related to the case of $D = 5$, and $NP = 20$; whereas Figs. 6–8 depict $D = 20$, and $NP = 30$. The ordering of the image triplets is as follows: Unsuccessful, Average, and Successful clusters. The differences between mean values and standard deviations are mostly observable in the case of the successful cluster (Figs. 5 and 8); but, for the other two types, the statistical characteristics clearly overlap. Therefore, it was decided to test the per-function stability hypothesis.

Hypothesis 1 can be accepted: (Tabs. I–VI, bold font, above 5%), in almost all function case comparisons, when using the Wilcoxon signed-rank test [26] in Tabs. I–IX. The Wilcoxon signed-rank test yields a p-value; which needs to be under significance level, in order to reject a null hypothesis (in this case, a value above $\alpha = 0.05$ was chosen for bold-facing

Algorithm 1 The k-means algorithm with parameters applied.

Input: $x_1, x_2, ..., x_{NP}$
Cluster count: $k = 3$
Clusters: $S_1 = S_2 = S_3 = \emptyset$
For each x from input
  Assign x to randomly (uniform dist) selected cluster (S_1, S_2, S_3)
End
Do
  Count centroids of each cluster $c_i = \frac{1}{|S_i|} \sum_{x_j \in S_i} x_j$
  For each x from input
    Assign each x to a cluster based on the distance to centroids
    Calculate distance to i-th cluster $d_i = (x - c_i)^2$
  End
While no further change in clusters S_1, S_2, S_3

the non-rejected null hypotheses – and, if rejected ... $p \leq \alpha$; then, the cluster sizes' values in such cases would be marked as different). In Tabs. I and II, the p-value is less than 5% in only 4/30 and 3/30 cases to reject the null hypothesis on a function-to-function basis (thus, 88.333% cases in these tables confirm the hypothesis). An example setting is, $D = 5$ with $NP = 20$ (i.e. Tab. I); here, the values of p are only below 5% in Rastrigin vs. De Jong for the average and successful clusters; or for Rosenbrock vs. Ackley or vs. Rastrigin for the successful cluster. For the remaining tables in this set, the rejection rate is 4/30 (Tab. III); 1/30 (Tab. IV); 0/30 (Tab. V); 5/30 (Tab. VI).

Similar conclusion rates can be observed for per-dimension cluster differences (that is to say hypothesis 2) comparisons, as reported in Tabs. VIIa–VIIj. The null hypothesis 2 is rejected only in 4/9 cases for Tab. VIIa (for $NP = 20$ and $NP = 30$); and in no other case (0/9 cases, all Tabs. VIIa– VIIj) in our per-dimension comparisons analysis.

More often however, the hypotheses are rejected when varying the population size ($NP = 20$ to $NP = 100$). Here, reported values for the De Jong function outcomes: 28/36, 10/36, and 29/36 – (together: 67/108); or 28/36, 14/36, and 32/36 (together: 74/108); for $D = 5$ and $D = 10$, respectively.

As the first two hypotheses are well confirmed, we would currently suggest the plausibility of the application these three hypotheses outcomes as an insight metric (*EA behavior descriptor*, as suggested by [21] and exemplified for DE in [22]).

With regard to hypothesis 3, this is rejected in several cases (see Tabs. VIII–IX). As such, it might need further analysis – but this hypothesis might nevertheless be useful as an insight metric in an on-line run: for the classification of the fitness space or of the optimization function challenge class and its features – in order to control the optimization algorithm components, like variation mechanisms and their control parameters – e.g., population sizing or structuring in population-based optimization algorithms – including DE.

TABLE VII: Wilcoxon signed-rank test p-values for different functions and their dimensions.

(a) De Jong, $NP = 20$.

	5	10	20
1. cluster - Unsuccessful			
5	**1.**	**0.772**	**0.692**
10	-	**1.**	**0.516**
20	-	-	**1.**
2. cluster - Average			
5	**1.**	**0.871**	0.049
10	-	**1.**	0.021
20	-	-	**1.**
3. cluster - Successful			
5	**1.**	**0.746**	0.012
10	-	**1.**	0.004
20	-	-	**1.**

(b) De Jong, $NP = 30$.

	5	10	20
1. cluster - Unsuccessful			
5	**1.**	**0.239**	**0.768**
10	-	**1.**	**0.067**
20	-	-	**1.**
2. cluster - Average			
5	**1.**	**0.105**	**0.955**
10	-	**1.**	**0.108**
20	-	-	**1.**
3. cluster - Successful			
5	**1.**	**0.847**	**0.798**
10	-	**1.**	**0.666**
20	-	-	**1.**

(c) Schwefel, $NP = 20$.

	5	10	20
1. cluster - Unsuccessful			
5	**1.**	**0.932**	**0.946**
10	-	**1.**	**0.899**
20	-	-	**1.**
2. cluster - Average			
5	**1.**	**0.523**	**0.355**
10	-	**1.**	**0.819**
20	-	-	**1.**
3. cluster - Successful			
5	**1.**	**0.184**	**0.114**
10	-	**1.**	**0.673**
20	-	-	**1.**

(d) Schwefel, $NP = 30$.

	5	10	20
1. cluster - Unsuccessful			
5	**1.**	**0.125**	**0.323**
10	-	**1.**	**0.076**
20	-	-	**1.**
2. cluster - Average			
5	**1.**	**0.356**	**0.195**
10	-	**1.**	**0.791**
20	-	-	**1.**
3. cluster - Successful			
5	**1.**	**0.087**	**0.365**
10	-	**1.**	**0.303**
20	-	-	**1.**

(e) Ackley, $NP = 20$.

	5	10	20
1. cluster - Unsuccessful			
5	**1.**	**1.**	**0.235**
10	-	**1.**	**0.194**
20	-	-	**1.**
2. cluster - Average			
5	**1.**	**0.801**	**0.819**
10	-	**1.**	**0.919**
20	-	-	**1.**
3. cluster - Successful			
5	**1.**	**0.808**	**0.36**
10	-	**1.**	**0.585**
20	-	-	**1.**

(f) Ackley, $NP = 30$.

	5	10	20
1. cluster - Unsuccessful			
5	**1.**	**0.762**	**0.265**
10	-	**1.**	**0.319**
20	-	-	**1.**
2. cluster - Average			
5	**1.**	**0.124**	**0.202**
10	-	**1.**	**0.863**
20	-	-	**1.**
3. cluster - Successful			
5	**1.**	**0.125**	**0.673**
10	-	**1.**	**0.458**
20	-	-	**1.**

(g) Rastrigin, $NP = 20$.

	5	10	20
1. cluster - Unsuccessful			
5	**1.**	**0.4**	**0.927**
10	-	**1.**	**0.472**
20	-	-	**1.**
2. cluster - Average			
5	**1.**	**0.837**	**0.642**
10	-	**1.**	**0.486**
20	-	-	**1.**
3. cluster - Successful			
5	**1.**	**0.342**	**0.079**
10	-	**1.**	**0.746**
20	-	-	**1.**

(h) Rastrigin, $NP = 30$.

	5	10	20
1. cluster - Unsuccessful			
5	**1.**	**0.172**	**0.929**
10	-	**1.**	**0.091**
20	-	-	**1.**
2. cluster - Average			
5	**1.**	**0.461**	**0.775**
10	-	**1.**	**0.683**
20	-	-	**1.**
3. cluster - Successful			
5	**1.**	**0.854**	**0.658**
10	-	**1.**	**0.541**
20	-	-	**1.**

(i) Rosenbrock, $NP = 20$.

	5	10	20
1. cluster - Unsuccessful			
5	**1.**	**0.314**	**0.85**
10	-	**1.**	**0.311**
20	-	-	**1.**
2. cluster - Average			
5	**1.**	**0.603**	**0.627**
10	-	**1.**	**0.819**
20	-	-	**1.**
3. cluster - Successful			
5	**1.**	**0.375**	**0.922**
10	-	**1.**	**0.483**
20	-	-	**1.**

(j) Rosenbrock, $NP = 30$.

	5	10	20
1. cluster - Unsuccessful			
5	**1.**	**0.871**	**0.99**
10	-	**1.**	**0.852**
20	-	-	**1.**
2. cluster - Average			
5	**1.**	**0.399**	**0.154**
10	-	**1.**	**0.321**
20	-	-	**1.**
3. cluster - Successful			
5	**1.**	**0.211**	**0.058**
10	-	**1.**	**0.187**
20	-	-	**1.**

Nevertheless, when the population size was changed, the proportional sizes of average cluster remained comparable (for Tab. VIII: 10 out of 36; and for Tab. IX: 14 out of 36, respectively are rejections). This might suggest that a linear increase in population size also linearly increases the size of average clusters, but the rest of the individuals is unevenly divided between the successful and unsuccessful clusters. Such phenomena are also supported by Fig. 9, where the proportions of clusters for 5 test functions are plotted against population size. As reported in Tabs.I–IX, the hypotheses results are largely invariant – regardless of the change in test function or dimension.

V. Conclusion

This paper presented an insight into proportions between k-means clusters of successful Differential Evolution (DE) donor generating vectors. It also demonstrated that the probability of these proportions being similar is high – and thereby, that these cluster proportions appear regularly and are observed in different test functions and their dimensions. The algorithm was run, in several independent runs, and the success clusters proportions remained similar. It was also shown that the

TABLE VIII: Wilcoxon signed-rank test p-values, function De Jong, $D = 5$.

NP	20	30	40	50	60	70	80	90	100
1. cluster - Unsuccessful									
20	**1.**	0.	0.012	**0.159**	**0.561**	0.	0.	**0.271**	0.001
30	-	**1.**	0.	0.	0.	**0.94**	0.002	0.	0.004
40	-	-	**1.**	**0.08**	0.038	0.	0.	0.006	0.
50	-	-	-	**1.**	**0.719**	0.	0.	0.01	0.
60	-	-	-	-	**1.**	0.	0.	**0.068**	0.
70	-	-	-	-	-	**1.**	0.008	0.	0.002
80	-	-	-	-	-	-	**1.**	0.	0.
90	-	-	-	-	-	-	-	**1.**	**0.061**
100	-	-	-	-	-	-	-	-	**1.**
2. cluster - Average									
20	**1.**	0.039	**0.078**	0.	0.003	0.001	0.	0.	0.001
30	-	**1.**	**0.992**	**0.053**	**0.142**	**0.411**	0.015	**0.061**	**0.13**
40	-	-	**1.**	**0.053**	**0.102**	**0.299**	0.021	**0.104**	0.024
50	-	-	-	**1.**	**0.381**	**0.285**	**0.926**	**0.727**	**0.393**
60	-	-	-	-	**1.**	**0.689**	**0.276**	**0.974**	**0.975**
70	-	-	-	-	-	**1.**	**0.086**	**0.267**	**0.589**
80	-	-	-	-	-	-	**1.**	**0.673**	**0.279**
90	-	-	-	-	-	-	-	**1.**	**0.853**
100	-	-	-	-	-	-	-	-	**1.**
3. cluster - Successful									
20	**1.**	0.003	0.	0.	0.	0.003	0.	0.004	**0.405**
30	-	**1.**	0.	0.	0.	**0.509**	**0.758**	0.	0.
40	-	-	**1.**	**0.957**	**0.108**	0.	0.	0.002	0.
50	-	-	-	**1.**	**0.122**	0.	0.	0.001	0.
60	-	-	-	-	**1.**	0.	0.	0.008	0.
70	-	-	-	-	-	**1.**	**0.249**	0.	0.
80	-	-	-	-	-	-	**1.**	0.	0.
90	-	-	-	-	-	-	-	**1.**	0.003
100	-	-	-	-	-	-	-	-	**1.**

TABLE IX: Wilcoxon signed-rank test p-values, function De Jong, $D = 10$.

NP	20	30	40	50	60	70	80	90	100
1. cluster - Unsuccessful									
20	**1.**	0.	0.	**0.829**	**0.483**	0.	0.	**0.299**	0.
30	-	**1.**	0.	0.	0.	**0.237**	**0.119**	0.	0.
40	-	-	**1.**	0.004	0.	0.	0.	0.	0.
50	-	-	-	**1.**	**0.902**	0.	0.	**0.262**	0.003
60	-	-	-	-	**1.**	0.	0.	**0.153**	0.001
70	-	-	-	-	-	**1.**	0.01	0.	0.
80	-	-	-	-	-	-	**1.**	0.	0.
90	-	-	-	-	-	-	-	**1.**	0.001
100	-	-	-	-	-	-	-	-	**1.**
2. cluster - Average									
20	**1.**	0.002	**0.058**	0.	0.	0.	0.	0.	0.
30	-	**1.**	0.032	**0.315**	**0.812**	**0.381**	**0.94**	**0.294**	**0.299**
40	-	-	**1.**	0.032	0.001	0.003	0.021	0.004	0.007
50	-	-	-	**1.**	**0.845**	**0.853**	**0.233**	**0.829**	**0.812**
60	-	-	-	-	**1.**	**0.797**	**0.299**	**0.864**	**0.846**
70	-	-	-	-	-	**1.**	**0.188**	**0.727**	**0.918**
80	-	-	-	-	-	-	**1.**	**0.168**	**0.276**
90	-	-	-	-	-	-	-	**1.**	**0.984**
100	-	-	-	-	-	-	-	-	**1.**
3. cluster - Successful									
20	**1.**	0.009	0.	0.	0.	**0.349**	0.001	0.	0.028
30	-	**1.**	0.	0.	0.	**0.07**	**0.294**	0.	0.
40	-	-	**1.**	0.003	0.001	0.	0.	0.	0.
50	-	-	-	**1.**	**0.565**	0.	0.	0.008	0.
60	-	-	-	-	**1.**	0.	0.	0.021	0.
70	-	-	-	-	-	**1.**	0.002	0.	0.001
80	-	-	-	-	-	-	**1.**	0.	0.
90	-	-	-	-	-	-	-	**1.**	0.
100	-	-	-	-	-	-	-	-	**1.**

proportions are steady, even when varying the number of dimensions for a fitness function. The possible dynamics of these proportions was however, reported later – in the case where changing the optimization algorithm for instance – control parameters like population size: here, the proportions of the most successful and most unsuccessful vectors changed more significantly.

The behaviors of these two proportions were presented as two new metrics, plausible for use with DE algorithms and online control. Therefore, in future work, it would be interesting to apply these metrics during an evolutionary DE run to control their behavior, like for instance – with variation operators or population structuring. The control of population size with an emphasis on pruning out unsuccessful individuals, would be especially interesting. Based on the score development of individuals, one may detect possible scenarios with scores developments and proportions between cluster sizes, or for the exploration/exploitation phases of heuristics [27]. From the insight perspective, it would also be interesting to study the effects of the fitness improvement occurrences; crossover (parent vector) variation influences; or per-component and fitness function parameters separability characteristics' effects on successful cluster formation, coupled with per-component propagation.

ACKNOWLEDGMENT

This work was supported by the GACR Grant Agency of the Czech Republic Project No.: P103/15/06700S; further, by the Ministry of Education, Youth and Sports of the Czech Republic within the National Sustainability Programme, Project No.: LO1303 (MSMT-7778/2014) and the RDF – CEBIA-Tech Project No.: CZ.1.05/2.1.00/03.0089; and by teh Tomas Bata University Internal Grant Agency Project No.: IGA/CebiaTech/2016/007. This work was also funded in part by the Slovenian Research Agency, Project No.: P2-0041; and supported by COST Action 15140.

REFERENCES

[1] R. Storn and K. Price, "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," *Journal of Global Optimization*, vol. 11, pp. 341–359, 1997.

[2] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*, ser. Natural Computing Series. Berlin, Germany: Springer-Verlag, 2005.

[3] F. Neri and V. Tirronen, "Recent Advances in Differential Evolution: A Survey and Experimental Analysis," *Artificial Intelligence Review*, vol. 33, no. 1–2, pp. 61–106, 2010.

[4] S. Das and P. N. Suganthan, "Differential Evolution: A Survey of the State-of-the-art," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 4–31, 2011.

[5] S. Das, S. S. Mullick, and P. Suganthan, "Recent advances in differential evolution – An updated survey," *Swarm and Evolutionary Computation*, vol. 27, pp. 1–30, 2016.

[6] J. Brest, S. Greiner, B. Bošković, M. Mernik, and V. Žumer, "Self-Adapting Control Parameters in Differential Evolution: A Comparative Study on Numerical Benchmark Problems," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 6, pp. 646–657, 2006.

[7] A. K. Qin, V. L. Huang, and P. N. Suganthan, "Differential evolution algorithm with strategy adaptation for global numerical optimization," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 398–417, 2009.

[8] J. Zhang and A. C. Sanderson, "JADE: adaptive differential evolution with optional external archive," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 5, pp. 945–958, 2009.

[9] S. Das, A. Abraham, U. Chakraborty, and A. Konar, "Differential Evolution Using a Neighborhood-based Mutation Operator," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 3, pp. 526–553, 2009.

[10] E. Mininno, F. Neri, F. Cupertino, and D. Naso, "Compact Differential Evolution," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 32–54, 2011.

[11] R. Mallipeddi, P. N. Suganthan, Q. K. Pan, and M. F. Tasgetiren, "Differential evolution algorithm with ensemble of parameters and mutation strategies," *Applied Soft Computing*, vol. 11, no. 2, pp. 1679–1696, 2011.

[12] J. Brest, P. Korošec, J. Šilc, A. Zamuda, B. Bošković, and M. S. Maučec, "Differential evolution and differential ant-stigmergy on dynamic optimisation problems," *International Journal of Systems Science*, vol. 44, no. 4, pp. 663–679, 2013.

[13] R. Tanabe and A. Fukunaga, "Evaluating the performance of SHADE on CEC 2013 benchmark problems," in *Proceedings of the 2013 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2013, pp. 1952–1959.

[14] R. Tanabe and A. S. Fukunaga, "Improving the search performance of SHADE using linear population size reduction," in *2014 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2014, pp. 1658–1665.

[15] J. Brest, M. S. Maučec, and B. Bošković, "iL-SHADE: Improved L-SHADE Algorithm for Single Objective Real-Parameter Optimization," in *2016 IEEE World Congress on Computational Intelligence (IEEE WCCI 2016), Vancouver, Canada*, 2016, pp. 1188–1195.

[16] A. Viktorin, M. Pluhacek, and R. Senkerik, "Success-History Based Adaptive Differential Evolution Algorithm with Multi-Chaotic Framework for Parent Selection Performance on CEC2014 Benchmark Set," in *2016 IEEE World Congress on Computational Intelligence (IEEE WCCI 2016), Vancouver, Canada*, 2016, pp. 4797–4803.

[17] R. Polakova, J. Tvrdík, and P. Bujok, "L-SHADE with Competing Strategies Applied to CEC2015 Learning-based Test Suite," in *2016 IEEE World Congress on Computational Intelligence (IEEE WCCI 2016), Vancouver, Canada*, 2016, pp. 4790–4796.

[18] N. H. Award, M. Z. Ali, P. N. Suganthan, and R. G. Reynolds, "An Ensemble Sinusoidal Parameter Adaptation incorporated with L-SHADE for Solving CEC2014 Benchmark Problems," in *2016 IEEE World Congress on Computational Intelligence (IEEE WCCI 2016), Vancouver, Canada*, 2016, pp. 2958–2965.

[19] A. Zamuda, J. D. H. Sosa, and L. Adler, "Constrained Differential Evolution Optimization for Underwater Glider Path Planning in Sub-mesoscale Eddy Sampling," *Applied Soft Computing*, vol. 42, pp. 93–118, 2016.

[20] A. Zamuda, "Differential Evolution and Large-Scale Optimization Applications," *IGI Global, InfoSci-Videos*, April 2016.

[21] G. Karafotias, M. Hoogendoorn, and A. Eiben, "Parameter control in evolutionary algorithms: Trends and challenges," *Evolutionary Computation, IEEE Transactions on*, vol. 19, no. 2, pp. 167–187, April 2015.

[22] A. Zamuda and J. Brest, "Self-adaptive control parameters' randomization frequency and propagations in differential evolution," *Swarm and Evolutionary Computation*, vol. 25, pp. 72–99, 2015, (Special Issue on Recent Advances in Modern Nature-Inspired Algorithms, RAMONA).

[23] R. Tanabe and A. S. Fukunaga, "How Far Are We From an Optimal, Adaptive DE?" in *Proceedings of the 14th International Conference on Parallel Problem Solving from Nature (PP SN XIV)*. IEEE, 2016, p. accepted.

[24] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[25] S. J. Nanda and G. Panda, "A survey on nature inspired metaheuristic algorithms for partitional clustering," *Swarm and Evolutionary computation*, vol. 16, pp. 1–18, 2014.

[26] S. García, D. Molina, M. Lozano, and F. Herrera, "A study on the use of non-parametric tests for analyzing the evolutionary algorithms behaviour: a case study on the CEC2005 Special Session on Real Parameter Optimization," *Journal of Heuristics*, vol. 15, no. 6, pp. 617–644, 2009.

[27] M. Črepinšek, S.-H. Liu, and M. Mernik, "Exploration and exploitation in evolutionary algorithms: A survey," *ACM Computing Surveys (CSUR)*, vol. 45, no. 3, p. 35, 2013.