Revisiting the Analysis of Population Variance in Differential Evolution Algorithms

Daniela Zaharie
Department of Computer Science
West University of Timisoara
blvd. Vasile Pârvan, 300223 Timişoara, Romania
Email: daniela.zaharie@e-uvt.ro

Flavia Micota
Department of Computer Science
West University of Timisoara
blvd. Vasile Pârvan, 300223 Timişoara, Romania
Email: flavia.micota@e-uvt.ro

***Abstract*—The performance of Differential Evolution (DE) algorithms is highly dependent on the trial population diversity and on the way the control parameter space is sampled. Therefore, identifying critical regions containing control parameters (e.g. scale factor, crossover rate) which can induce undesired behaviour (e.g. premature convergence) is useful. In this context, the aim of the paper is twofold. On one hand, the paper revisits some existing theoretical results on the expected variance of the trial population aiming to provide a comparative image on critical regions in the control parameter space for several DE variants: DE/rand/1/*, DE/best/1/*, DE/rand-to-best/*, DE/either-or. On the other hand, a new theoretical result on DE/rand/1/* population variance evolution is obtained under the assumption that the bound constraints are handled by random reinitialization of infeasible components. The relationship between the probability of violating the bound constraints and the value of the scale factor, F, is theoretically derived for DE/rand/1/* and empirically analyzed for other DE mutation operators.**

I. Introduction

Differential Evolution (DE) is undoubtfully one of the most popular stochastic population-based metaheuristic, the amount of papers devoted to it being currently impressive (see some recent surveys [5],[10],[6]). Through its search mechanism particularities (e.g. finite set of search directions and specific combination of deterministic and random components) DE found its own niche in the evolutionary computation field. Since its proposal by R. Storn and K. Price in [14], the family of DE algorithms has been extended by the introduction of new mutation and crossover operators, control parameter adaptation rules, hybridization with other local or global search methods etc.

Most of the DE related work focuses on analyzing the behaviour on benchmark test functions or on various real-world applications. From a pragmatical point of view the empirical analysis of the DE performance is very useful but it should be complemented with insights on the DE behaviour extracted through a theoretical analysis. The DE theoretical results are still scarce and can be grouped in several categories: (i) analysis of the probability distribution of the mutant or trial population [1], [16], [17] and of its evolution, under some specific assumptions, towards a distribution concentrated on the global optimum [8]; (ii) analysis of the DE dynamics stability by using the Lyapunov function method [7]; (iii) analysis of the expected variance of the trial population aiming to obtain insights on the evolution of the population diversity and to control the risk of premature convergence [15], [18], [20], [21].

The aim of this paper is to revisit some of the existing results on the expected variance of the DE trial population under some practical assumptions, as it is the issue of handling bounding box constraints. The interest in analyzing the trial population variance is motivated by the fact that it is a measure of population diversity and for DE algorithms, the population diversity (which is related to the differences between the population elements) has a direct impact on the population dynamics. In fact, in the absence of a mutation based on a perturbation which is independent of the current population, the population diversity is the main driving force of the evolution. Therefore, maintaining the DE population diversity is of paramount importance. Moreover, the DE behavior is influenced by the strategy of sampling the control parameter space, which could contain regions which should not be systematically sampled, as they might induce premature convergence (the population variance is decreasing even without selection pressure, i.e. for flat functions). Therefore, Sect. II provides an overview, for most of the DE mutation variants, on the shapes and sizes of control parameter regions which should be avoided as they induce premature convergence even for flat functions. On the other hand, the influence of the methods for handling bounding-box constraints on DE behavior has been investigated mainly experimentally [2], [9], [11]. The main remarks following these experimental analyses are that the method of handling the bound constraints has a significant influence on the performance of Differential Evolution [2] and Particle Swarm Optimization [9]. The interesting flat landscape analysis conducted in [9] suggests that some bound handling techniques introduce a significant bias in the search of the feasible region favouring either the middle or the boundaries. However, this bias should be interpreted modulo the probability of violating the boundaries (for low violation probabilities the bias could be less effective). In order to gather a more in-depth knowledge on these aspects, particularly in the case of DE, Sect. III addresses questions such as: (i) how frequently do the DE/rand/1 trial vectors violate the bounding box constraint? (ii) what is the impact of the random reinitialization of trial elements violating the bounding box

978-1-5090-4601-0/17/$ 31.00 2017 European Union

constraints on the population diversity?

II. Summary of existing results on expected variance of DE trial population

Most of DE algorithms are structured following the general template of population-based metaheuristics. Algorithm 1 illustrates the main steps of a generational DE. At each generation, a trial population $Z = \{z_1, \ldots, z_m\}$ is constructed based on a difference-based mutation and a crossover. In order to ensure that each trial element is in the feasible region (e.g. a bounding box as $[a_1, b_1] \times \ldots \times [a_n, b_n]$) it is repaired as soon as it is detected that it violates the bound conditions. The bounding box constraints are checked and handled independently for each component of the trial solution, thus the analysis can be conducted component-wise.

Algorithm 1 The general structure of a generational DE

```
Population initialization X(0) ← {x_1(0), ..., x_m(0)}
g ← 0
while the stopping condition is false do
    for i = 1,m do
        Y_i ← generateMutant(X(g))
        Z_i ←crossover(x_i(g),Y_i)
        if Z_i violates the bound conditions then
            Z_i ← repair(Z_i)
        end if
    end for
    for i = 1,m do
        if f(Z_i) < f(x_i(g)) then
            x_i(g+1) ← Z_i
        else
            x_i(g+1) ← x_i(g)
        end if
    end for
    g ← g + 1
end while
```

A. DE mutation and crossover

Most of the DE mutation operators belong to one of the categories described below, where I_i, J_i and K_i denotes random indices uniformly generated (under some non-equality constraints) from the set $\{1, 2, \ldots, m\}$, ξ denotes random values corresponding to the scale factor and $\lambda \in [0, 1]$ is a convex recombination parameter.

DE/rand-to-best/L. The mutation described in Eq. (1), which constructs the mutant element Y_i, incorporates several well-known variants: DE/rand/1 ($\lambda = 0$, $L = 1$), DE/rand/2 ($\lambda = 0$, $L = 2$), DE/best/1 ($\lambda = 1$, $L = 1$), DE/best/2 ($\lambda = 1$, $L = 2$).

$$Y_i = \lambda x_* + (1 - \lambda) x_{I_i} + \sum_{l=1}^{L} \xi_l \cdot (x_{J_{il}} - x_{K_{il}}) \qquad (1)$$

DE/current-to-rand/1. It is characterized by the usage as base element of a convex recombination between the current element and a random one. This change in the base vector is expected to reduce the risk of generating mutants which violate the bound constraints.

$$Y_i = \lambda x_i + (1 - \lambda) x_{I_i} + \xi \cdot (x_{J_i} - x_{K_i})$$

DE/either-or. It is a variant, proposed in [12], which combines the role of mutation and crossover in one operator (by using the selection probability p_F) aiming to ensure rotational invariance. In Eq. 2 the coefficients F and K correspond to scaling factors.

$$Z_i = \begin{cases} x_{I_i} + F \cdot (x_{J_i} - x_{K_i}) & \text{with prob. } p_F \\ x_{I_i} + K \cdot (x_{J_i} + x_{K_i} - 2x_{I_i}) & \text{with prob. } 1 - p_F \end{cases} \qquad (2)$$

Except for DE/either-or and the variants based on arithmetical recombination, the DE trial vectors, Z_i, are constructing by mixing the components of the current element (x_i) with those of the mutant (Y_i). The amount of components taken from the mutant is controlled by the crossover ratio parameter (CR) which therefore determines the mutation probability, p_m. The relationship between p_m, CR and the problem size (n) depends on the type of crossover [19]: $p_m = CR(1 - 1/n) + 1/n$ (in the case of binomial crossover) and $p_m = (1 - CR^n)/(n(1 - CR))$ (in the case of exponential crossover).

B. Estimation of the expected variance of the trial population

The idea of using the expected population variance as predictor of the explorative power of an evolutionary algorithm has been introduced by Beyer [3] who conducted a component-wise analysis of the variance in the case of evolution strategies applied to flat landscapes. By conducting a similar analysis in the case of no-selection DE, the dependence between the expected variance of the trial population, $\mathbb{E}[\text{Var}(Z)]$, and the variance of the current population, $\text{Var}(X)$, has been derived for several mutation operators, as illustrated in Table I. As the analysis is conducted at the component level, the difference between binomial and exponential crossover is reflected only by different value of p_m for same value of CR.

In all cases the expected variance of trial population depends linearly on the current population variance ($\mathbb{E}[\text{Var}(Z)] = c \cdot \text{Var}(X) + d$). The slope of this linear dependence depends on the control parameters (p_m, $F^2 = \sum_{l=1}^{L} \mathbb{E}[\xi_l^2]$, λ, K, p_F etc), on the population size (m) and, indirectly (through the relationship between p_m and CR) on the problem size, n. The free term, d, is non-zero only for mutations which involves the best element in the population, x_*, and depends on the distance between x_* and the population average $\overline{X}$. When d is zero or very small and c is less than 1 then the variance of the trial population is smaller than the variance of the current population, even in the absence of selection pressure. In such a case the DE population will lose diversity quickly leading to premature convergence. This means that regions in the control parameter space which are characterized by a corresponding coefficient $c(p_m, F, m, ...)$ less than one should be avoided as they would favor premature convergence. Such regions could be as well used to induce premature convergence, when a quick population convergence is desired. In practice, one way to

DE mutation	$\mathbb{E}[\text{Var}(Z)]$ (expected variance of the trial population)	
DE/rand-to-best/L/*	$\left(1+2p_m\sum_{l=1}^{L}\mathbb{E}[\xi_l^2]-\frac{p_m(2-p_m)}{m}-p_m\lambda(2-\lambda)\frac{m-1}{m}\right)\text{Var}(X)+\lambda^2p_m(1-p_m)\frac{m-1}{m}(x_*-\overline{X})^2$	
DE/rand/1/*	$\left(1+2p_mF^2-\frac{p_m(2-p_m)}{m}\right)\text{Var}(X)$	[18]
DE/rand/2/*	$\left(1+4p_mF^2-\frac{p_m(2-p_m)}{m}\right)\text{Var}(X)$	[18]
DE/best/1/*	$\left(1+2p_mF^2-p_m-\frac{p_m(1-p_m)}{m}\right)\text{Var}(X)+p_m(1-p_m)\frac{m-1}{m}(x_*-\overline{X})^2$	[21]
DE/current-to-rand/1/*	$\left(1+2p_m\mathbb{E}[\xi^2]-p_m(2-p_m)(1-\lambda)\left(2\lambda+\frac{1-\lambda}{m}\right)\right)\text{Var}(X)$	[20]
DE/either-or	$\left(p_F^2(1+2F^2-\frac{1}{m})+2p_F(1-p_F)(\frac{m-1}{m}+F^2+3K^2-2K)+(1-p_F)^2\left(\frac{m-1}{m}+2\frac{m-2}{m}(3K^2-2K)\right)\right)\text{Var}(X)$	[20]

TABLE I
DEPENDENCE OF THE EXPECTED VARIANCE OF THE TRIAL POPULATION (Z) ON THE VARIANCE OF THE CURRENT POPULATION (X) FOR VARIOUS DE MUTATIONS.

reach the desired balance between exploration and exploitation is to sample the control parameter space around the border of the premature convergence region.

C. Critical regions

Figures 1-5 present regions in the space of two control parameters (e.g. F versus CR, F versus p_F) characterized by $c(m, F, ...) < 1$. All plots contain overlapped regions corresponding to population sizes of $m = 10$ (light red) and $m = 100$ (light blue) and a problem size of $n = 50$. In all cases the critical region for $m = 10$ is larger than for $m = 100$. In the case of DE/current-to-rand/1, the largest critical area (in the (CR, F) space) corresponds to values of λ in $[0.4, 0.6]$. In these cases, for values of F smaller than 0.5 the algorithm is prone to premature convergence, disregarding the value of CR. The smallest critical region corresponds to $\lambda = 0$, i.e. to DE/rand/1/bin. If binomial and exponential crossovers are analyzed comparatively (Figs. 1 and 2), it follows that in the case of binomial crossover the lower bound of effective F decreases linearly with CR, while in the case of exponential crossover a decrease in the lower bound can be noticed only for values of CR very close to 1.

An interesting behaviour can be noticed in the case of DE/either-or (Fig. 3) where for some values of K the critical region is rather large, with values of CR (e.g. $CR < 0.3$) for which premature convergence may be induced disregarding the value of F. It should be also noticed that, when K depends on F, a small region is obtained for $K = (F + 1)/2$ which is the recommended value in [12] (and for which the trial population variance is less dependent on p_F, as illustrated in [20]). On the other hand if K is chosen independent of F, then for $K > 0.7$ the premature convergence region becomes small especially for large values of m.

Finally, in the case of mutations involving the best element in the population, x_*, (Figs. 4 and 5), the critical regions area depends on the distance between x_* and the population average, being the largest when they are very close. As the ratio $(x_*-\overline{X})^2/\text{Var}(X)$ increases, the region becomes smaller, particularly in the case of exponential crossover. Thus, a bias of the population average with respect to the best element in the population might act as a diversity promoter even for small values of F (except for the cases when CR is close to 1).

III. INCLUDING BOUND CONSTRAINT HANDLING INTO ANALYSIS

All results presented in the previous section have been obtained under the simplifying assumption that no bound constraints are involved. However, in practice the design variables are bounded, thus values exceeding the bounds should be considered infeasible. The aim of this section is to derive the expected variance of the DE/rand/1/* trial population under the more realistic assumption that the bound constraints are properly handled.

A. Methods for bound constraint handling

There are several approaches in handling the bound constraints and the most frequently used (also included in the experimental analysis presented in [2]) are described in the following.

- *Random* reinitialization. The trial components which violate the bound constraints are replaced with values generated uniformly in the bounding box, i.e. if $z_i^j \notin [a_j, b_j]$ then z_i^j is replaced with a random value uniformly generated in $[a_j, b_j]$.
- *Projection* on the closest bound. The infeasible values are replaced with the closest bounds, i.e. if $z_i^j < a_j$ then it is replaced with a_j and if $z_i^j > b_j$ it is replaced with b_j.
- *Average* between the current element and the bound. An infeasible component is replaced with the average between the closest bound and the corresponding component of the current element, i.e. if $z_i^j < a_j$ then it is replaced with $(x_i^j + a_j)/2$ and if $z_i^j > b_j$ it is replaced with $(x_i^j + b_j)/2$.
- *Repeated* reflection. The repaired value is computed by repeatedly reflecting the infeasible value with respect to the closest bound, i.e. if $z_i^j < a_j$ then it is replaced with $2a_j - z_i^j$ and if $z_i^j > b_j$ it is replaced with $2b_j - z_i^j$.
- *Resampling*. A new trial element is generated, using the mutation operator on newly selected parents, until a feasible element is obtained. As such an iterated sampling

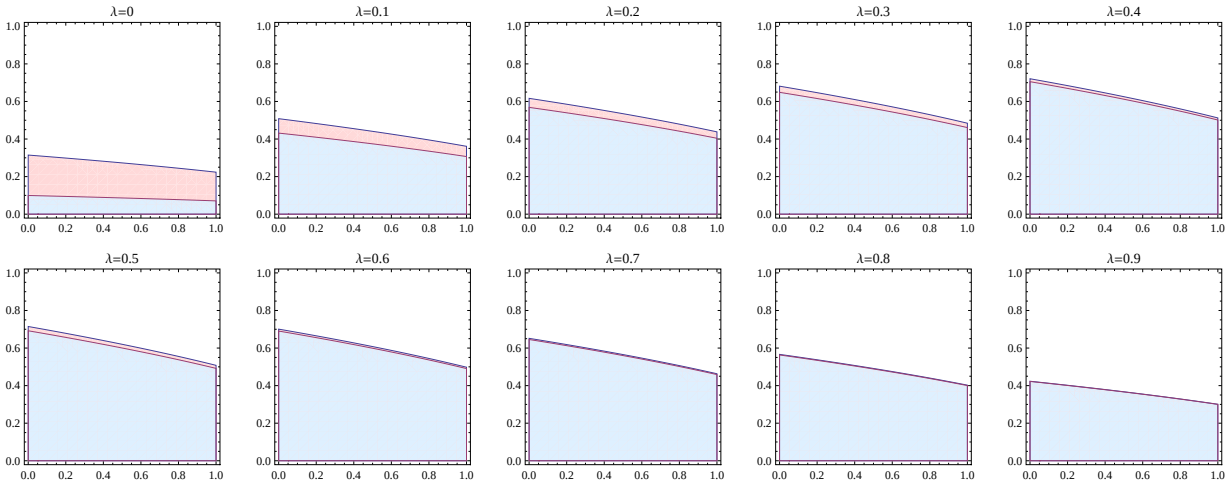

Fig. 1. Influence of λ on the critical region for DE/current-to-rand/1/bin (CR on Ox, F on Oy; overlapped regions for $m = 10$ - light red, $m = 100$ - light blue)

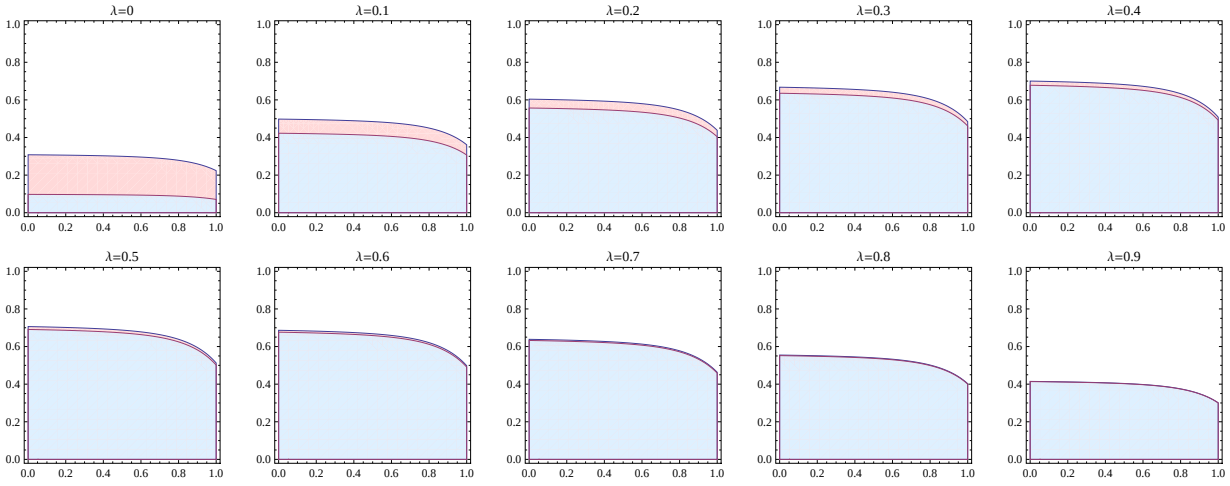

Fig. 2. Influence of λ on the critical region for DE/current-to-rand/1/exp (CR on Ox, F on Oy; overlapped regions for $m = 10$ - light red, $m = 100$ - light blue)

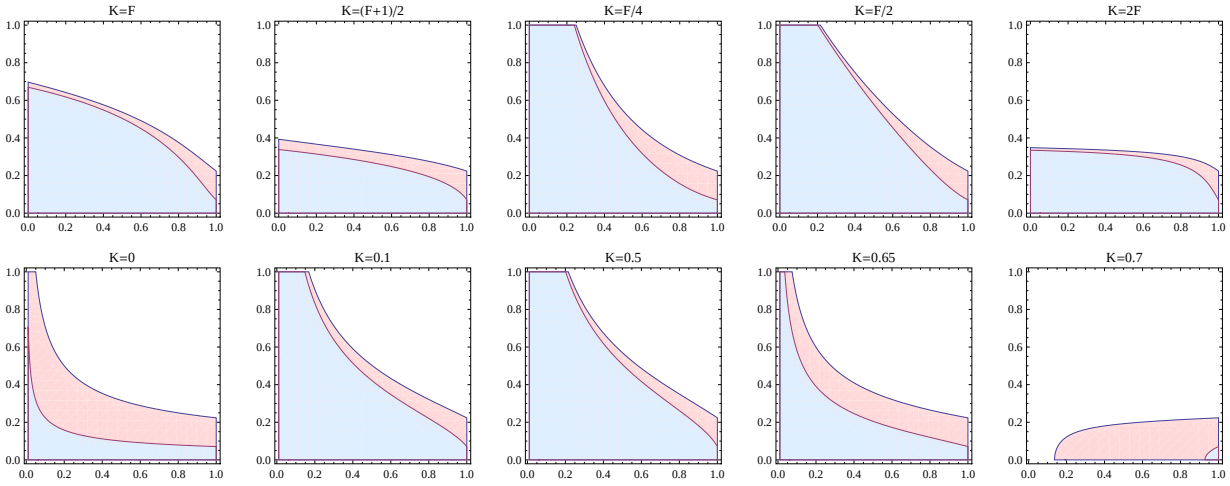

Fig. 3. Influence of K on the critical region for DE/either-or (p_F on Ox, F on Oy; overlapped regions for $m = 10$ - light red, $m = 100$ - light blue)

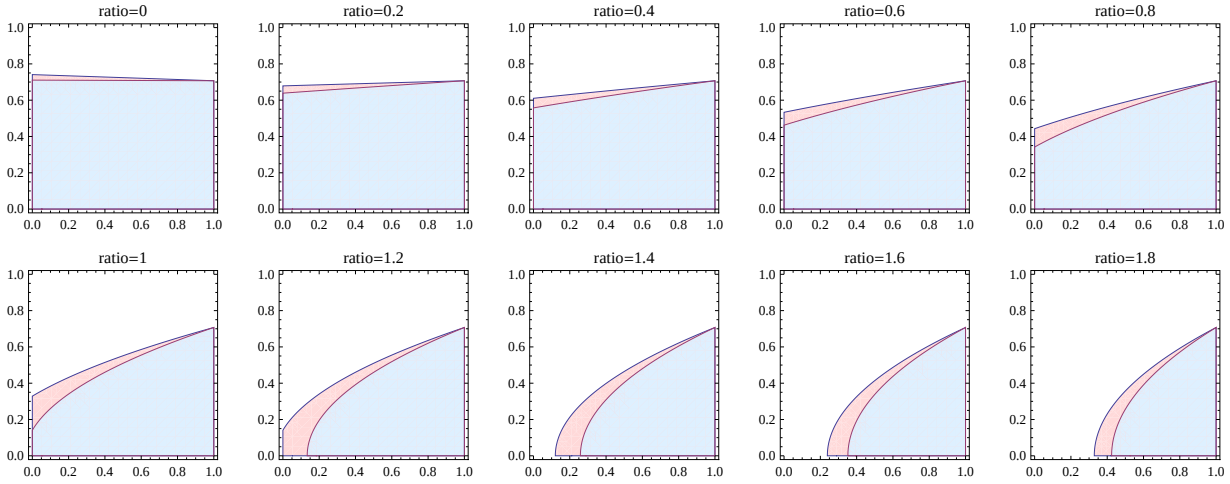

Fig. 4. Influence of the ratio $(x_* - \overline{X})^2/\mathrm{Var}(X)$ on the critical region for DE/best/1/bin (CR on Ox, F on Oy; overlapped regions for $m = 10$ - light red, $m = 100$ - light blue)

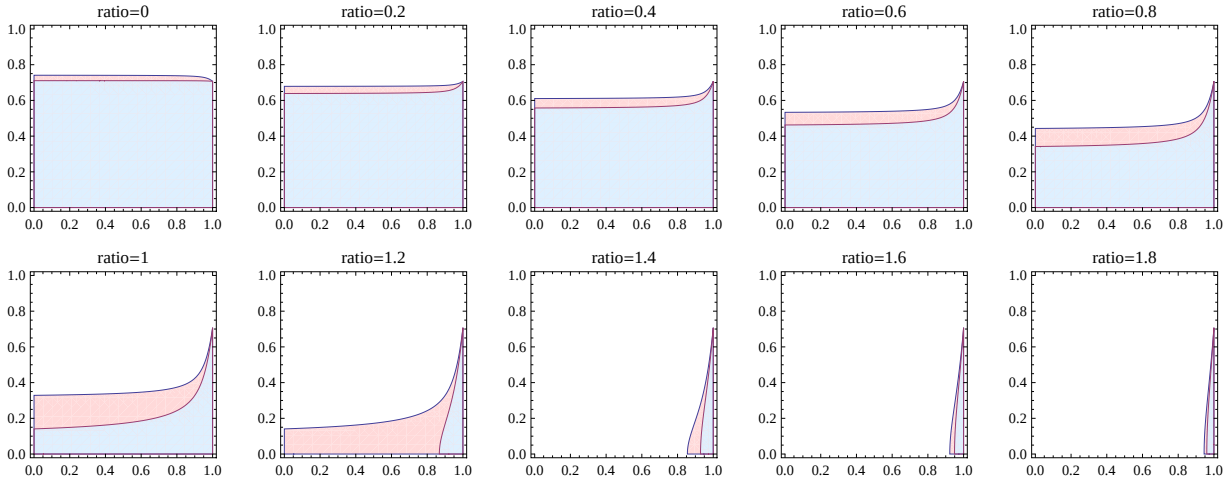

Fig. 5. Influence of the ratio $(x_* - \overline{X})^2/\mathrm{Var}(X)$ on the critical region (CR, F) for DE/best/1/exp (CR on Ox, F on Oy; overlapped regions for $m = 10$ - light red, $m = 100$ - light blue)

from the DE pool might by computationally costly, particularly for high-dimensional problems, another approach is a decoupled resampling (the repairing strategy is applied independently for each component violating the bounds).

The influence of a bound constraint method on the DE behaviour depends on its ability to generate trial elements which cannot be created by the DE operators. Two extreme cases corresponds to *resampling* which always generate elements from the DE pool (except for the case of component-wise resampling) and *random reinitialization* which can sample any element in the search space. The other methods have an intermediate behavior. Two of them (*projection* and *average*) preserve the change direction (increase or decrease of the components' values) but favors elements on or near the boundary. However, the motivation of choosing one method is rarely presented in DE papers, thus obtaining some insights on their characteristics might be useful.

B. Estimation of the bound violation probability

As the mutation-induced perturbation on a population element depends on the scale factor, F, it is to be expected that the probability of generating an element which violates the bound constraints increases as F increases. The question is how depends the violation probability on F. In the following we estimate the bound violation probability for DE/rand/1 mutation by using results on the probability distribution function (pdf) of DE/rand/1 mutants proved by Ali and Fatti in [1]. The analysis is valid for the first stages of the evolution when the elements of the population are almost uniformly distributed on the search space.

Let us consider a population of m scalar values uniformly distributed in an interval $[a, b]$ on which the DE/rand/1 mutation is applied, i.e. a mutant y is constructed as $x_I + F \cdot (x_J - x_K)$ (with I, J and K distinct random values uniformly

distributed in $\{1,\ldots,m\}$). As in [1] one can assume, without losing generality, that $a=0$ and $b=1$. Moreover we will consider that $F \in (0,1]$, thus any mutant will belong to $[-F,1+F]$ and the mutants violating the bounds will belong to $[-F,0)\cup(1,1+F]$. If f_Y denotes the pdf of Y then the probability of violating the bounds is $P(Y\in[-F,0)\cup(1,1+F]) = \int_{-F}^{0} f_Y(y)dy + \int_{1}^{1+F} f_Y(y)dy$. Since the pdf of Y, as derived in [1], satisfies $f_Y(y) = (F+y)^2/(2F^2)$ for $-F \le y \le 0$ and $f_Y(y) = (F-y+1)^2/(2F^2)$ for $1 \le y \le 1+F$ it follows that the bound violation probability $p_v = P(Y\in[-F,0)\cup(1,1+F]) = F/6+F/6 = F/3$.

Thus in the first stage of evolution, when one can consider that the population is (almost) uniformly distributed on the search space, the probability for a mutant to violate the bounds is close to $F/3$. In the case of flat functions this property remains true also for further generations (see Figure 6). In the case of non-flat functions it is expected that the violation probability decreases in time, as the population may be concentrated in a smaller region (see Fig. 7 for the case of the Sphere function, which also illustrates the influence of position of the global optimum in the search space and of the bound handling method).

A further question is related to the influence of the mutation type on the bound violation probability. Even if the pdf corresponding to other DE mutation can be computed, the computation is more intricate than for DE/rand/1. Therefore we conducted an empirical analysis and estimated the violation probability as average of the ratio of components violating the bounds in the context of constructing a mutant population of $m=100$ elements (each one with $n=100$ components) for 1000 iterations (thus the averages have been estimated based on 10^7 instances). The components violating the bounds in one iteration have been replaced with values obtained by applying one of the repairing rules: *random*, *resampling*, *projection*, *average*, *reflection*. As the results in Table II and Figure 6 illustrate there are some differences in the violation probability corresponding to different mutations. As expected, the violation probability is smaller in the case of DE/current-to-rand/1 (as the base element is a convex recombination of two population elements, thus farther from the bounds than at least one of these elements). On the other hand, in the case of DE/rand/2 the empirically estimated violation probability is close to $F\sqrt{2}/3$ which is in agreement with the remark that in case of two differences the impact of the scale factor is multiplied by $\sqrt{2}$ (see the results in Table I on the expected trial population variance where the factor $2F^2$ is replaced with $4F^2$). For DE/either-or (with $p_F=0.5$ and $K=(F+1)/2$) the violation probability is again close to $F/3$. Finally the results reported in columns 2-5 of Table II suggest that the violation probability, in the case of flat functions, is not influenced by the method used to repair the infeasible elements.

C. Evolution of the population variance

Let us analyze now the influence of the random reinitialization of infeasible elements (values outside $[a,b]$) on the expected variance of the trial population. By applying the same approach as in [20] (see Appendix) one can obtain that $\mathbb{E}[\text{Var}(Z)]$ depends linearly on $\text{Var}(X)$:

$$\mathbb{E}[\text{Var}(Z)] = c(p_m,p_v,m,F)\cdot\text{Var}(X) + d(p_m,p_v,m,a,b) \tag{3}$$

where the slope coefficient $c(p_m,p_v,m,F)$ depends on the control parameters as follows:

$$c(p_m,p_v,m,F) = (1-p_m)^2 + p_m p_v(1-p_m)\frac{m-1}{m} + p_m^2(1-p_v)^2 B\left[\frac{m-1}{m}+2F^2\right] + 2p_m(1-p_m)B\left[\frac{m-1}{m}+F^2\right] + 2p_m^2 p_v(1-p_v)B\left[\frac{(m-1)^2}{2m^2}+\frac{m-1}{m}F^2\right] \tag{4}$$

with B denoting a bounding function, i.e. $B(u)=u$ if $u\le 1$ and $B(u)=1$ if $u>1$. The second coefficient depends on the population average, $\overline{X}$, and on the characteristics of the probability distribution used to generate feasible elements (e.g. uniform distribution):

$$d(p_m,p_v,a,b) = p_m p_v(1-p_m p_v)\frac{m-1}{m}\left(\overline{X}-\frac{a+b}{2}\right)^2 + p_m p_v\left(1-\frac{1-p_m p_v}{m}\right)(b-a)^2/12 \tag{5}$$

Figure 8 illustrates the difference between the theoretical expected variance of the trial population in the case when the bound constraints are not handled (red dashed line) and in the case when random reinitialization is used (black solid line). The red curves correspond to $(1+2p_mF^2 - pm(2-pm)/m)^g\text{Var}(X(0))$. The black curves correspond to values of the variance computed by $\mathbb{E}[\text{Var}(Z(g))] = c(p_m,p_v,m,F)\cdot\mathbb{E}[\text{Var}(Z(g-1))] + d(p_m,p_v,m,a,b)$. Based on the results from the previous subsection the bound violation probability is set to $p_v=F/3$. These results confirm the intuition that random reinitialization of infeasible elements can slow down the decrease of variance (in case of small value of the scale factor F) but also avoid the unlimited increase of the variance (in case of larger values of F). However, a closer inspection of the evolution of variance shows that the theoretical expression provides an over-estimation. Fig. 9 illustrates the comparison between the expected variance evolution and that estimated experimentally based on 10 independent runs. This time the expected variance has been computed using only the one-step dependence, i.e. $\mathbb{E}[\text{Var}(Z(g))] = c(p_m,p_v,m,F)\cdot\mathbb{E}[\text{Var}(X(g))] + d(p_m,p_v,m,a,b)$, meaning that it is computed based on the variance of the current population not on that estimated at the previous step. The expected variance computed using the theoretical result is closer to empirical variance but slightly over-estimate it (especially for large values of F).

IV. Conclusions and further work

Knowledge of the regions in the control parameter space containing parameter values which induce decrease of the population variance even for flat landscapes might be useful in the context of adaptive algorithms, as a systematic sampling of

F	$F/3$	DE/rand/1				DE/either-or	DE/current -to-rand/1	DE/rand/2	$F\sqrt{2}/3$
		Random	Resampling	Projection	Reflection				
0.2	0.067	0.069 ± 0.002	0.067 ± 0.002	0.065 ± 0.002	0.069 ± 0.003	0.050 ± 0.002	0.013 ± 0.001	0.095 ± 0.003	0.094
0.5	0.167	0.165 ± 0.004	0.169 ± 0.004	0.165 ± 0.004	0.169 ± 0.004	0.155 ± 0.004	0.082 ± 0.003	0.232 ± 0.004	0.236
0.8	0.267	0.272 ± 0.004	0.266 ± 0.004	0.270 ± 0.004	0.266 ± 0.004	0.269 ± 0.004	0.204 ± 0.004	0.366 ± 0.005	0.377
1	0.330	0.330 ± 0.005	0.331 ± 0.005	0.333 ± 0.005	0.333 ± 0.005	0.330 ± 0.005	0.291 ± 0.005	0.449 ± 0.005	0.471

TABLE II

ESTIMATION OF THE BOUND VIOLATION PROBABILITY, p_v (FLAT LANDSCAPE ANALYSIS)

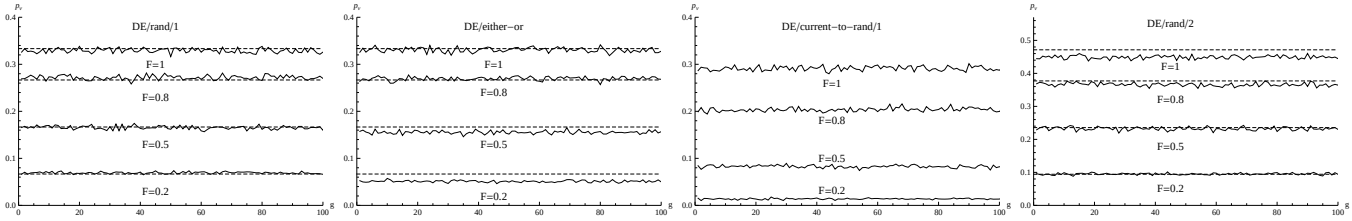

Fig. 6. Probability of bound violation (flat landscape analysis).Infeasible elements are replaced with random elements in the search space

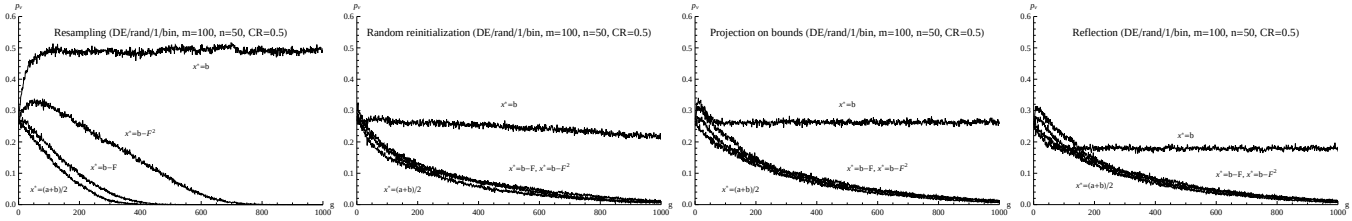

Fig. 7. Evolution of bound violation probability for a sphere function with different locations of the optimum and different bound handling methods.

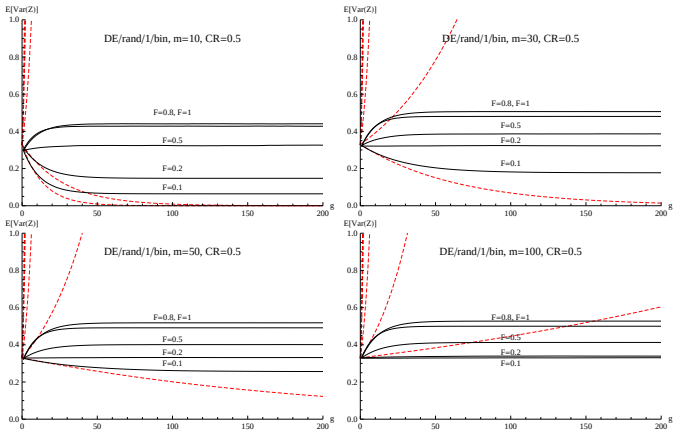

Fig. 8. Theoretical evolution of the expected variance of the trial population for DE/rand/1/bin without constraint handling (red dashed line) and with random initialization of infeasible elements (black solid line)

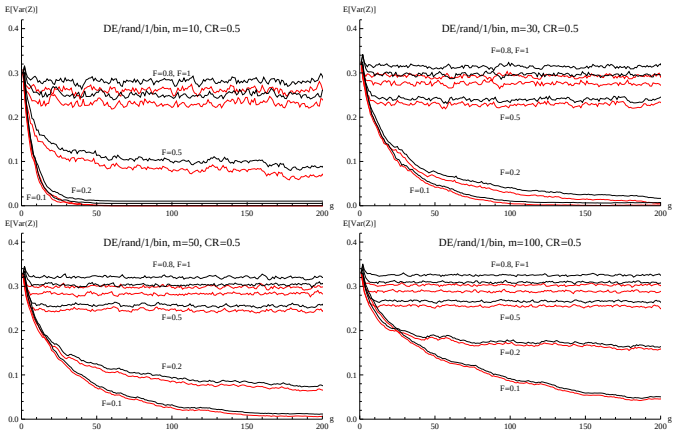

Fig. 9. Theoretical (black solid line) vs. empirical evolution (red line) of the expected variance of the trial population (flat landscape)

these regions would lead to premature convergence. This can be particularly relevant for small population size as in μDE [4], as the premature convergence regions usually increases as the population size decreases. Both theoretical and empirical analyses suggest that in the first evolution stages the probability of violating the bounds by mutants created using DE/rand/1 is close to $F/3$. By including the bound constraint handling in the analysis of the expected variance an upper bound of the trial population variance has been obtained. Further work will address the analysis of DE/current-to-pbest/* variants and of other bound constraint handling methods.

APPENDIX: SKETCH OF THE PROOF OF EQS.(3)-(5)

When a random reinitialization repairing strategy is applied, a trial element can be described as a random variable satisfying

$$Z_i = x_i \cdot \mathbf{1}_{\overline{M}_i} + (Y_i \cdot \mathbf{1}_{\overline{V}_i} + R_i \cdot \mathbf{1}_{V_i}) \cdot \mathbf{1}_{M_i}$$

with $\mathbf{1}_A$ denoting the indicator function corresponding to a probabilistic event, A. The events involved in the above equations are: M_i (during the crossover the component from the

mutant vector is selected), V_i (the mutant vector component violates the bound constraints) and their corresponding complement events, $\overline{M}_i$ and $\overline{V}_i$. The probabilities corresponding to these events are $Prob(M_i) = p_m$, $Prob(\overline{M}_i) = 1 - p_m$, $Prob(V_i) = p_v$, $Prob(\overline{V}_i) = 1 - p_v$ where p_m is the mutation probability and p_v is the bound violation probability. Since $\mathbb{E}[\text{Var}(Z)] = \frac{m-1}{2m}\mathbb{E}[(Z_i - Z_j)^2]$, where Z_i and Z_j are random but distinct elements from the trial population, it follows that it is enough to compute:

$$\begin{array}{l}\mathbb{E}[(Z_i - Z_j)^2] = (1-p_m)^2\mathbb{E}[(x_i - x_j)^2] + \\ p_m^2(1-p_v)^2\mathbb{E}[(Y_i - Y_j)^2] + p_m^2 p_v^2\mathbb{E}[(R_i - R_j)^2] + \\ 2p_m(1-p_m)(1-p_v)\mathbb{E}[(x_i - Y_j)^2] + \\ 2p_m(1-p_m)p_v)\mathbb{E}[(x_i - R_j)^2] + 2p_m^2 p_v(1-p_v)\mathbb{E}[(Y_i - R_j)^2]\end{array} \quad (6)$$

Lemma 1: Let $X = (x_1, x_2, \ldots, x_m)$ be a population of scalars and I and J two random variables taking values in the set of indices $\{1, 2, \ldots, m\}$. The random variables x_I and x_J have the following properties:
(i) if I and J are uniformly distributed then $\mathbb{E}(x_I) = \mathbb{E}(x_J) = \overline{X}$ ($\overline{X} = \frac{1}{m}\sum_{i=1}^{m} x_i$) and $\mathbb{E}(x_I^2) = \mathbb{E}(x_J^2) = \overline{X^2}$ ($\overline{X^2} = \frac{1}{m}\sum_{i=1}^{m} x_i^2$);
(ii) if I and J have distinct values then $\mathbb{E}[(x_I - x_J)^2] = \frac{2m}{m-1}\mathbb{E}[\text{Var}(X)]$;
(iii) if I and J are independent then $\mathbb{E}[(x_I - x_J)^2] = 2\mathbb{E}[\text{Var}(X)]$.

By using this lemma and properties of random variables uniformly distributed in $[a, b]$ one obtains:

$$\begin{array}{lcl}\mathbb{E}[(x_i - x_j)^2] & = & \frac{2m}{m-1}\text{Var}(X) \\ \mathbb{E}[(Y_i - Y_j)^2] & = & 2\left(1 + \frac{2m}{m-1}F^2\right)\text{Var}(X) \\ \mathbb{E}[(R_i - R_j)^2] & = & 2(b-a)^2/12 \\ \mathbb{E}[(x_i - Y_j)^2] & = & 2\left(1 + \frac{m}{m-1}F^2\right)\text{Var}(X) \\ \mathbb{E}[(x_i - R_j)^2] & = & \text{Var}(X) + \\ & & (\overline{X} - (a+b)/2)^2 + (b-a)^2/12 \\ \mathbb{E}[(Y_i - R_j)^2] & = & \left(\frac{m-1}{m} + F^2\right)\text{Var}(X) + \\ & & (\overline{X} - (a+b)/2)^2 + (b-a)^2/12\end{array} \quad (7)$$

By including these terms in Eq. 6 one obtains:

$$\begin{array}{l}\mathbb{E}[\text{Var}(Z)] = c(p_m, p_v, m, F)\cdot\text{Var}(X) + \\ p_m p_v(1 - p_m p_v)\frac{m-1}{m}\left(\overline{X} - \frac{a+b}{2}\right)^2 + \\ p_m p_v\left(1 - \frac{1-p_m p_v}{m}\right)(b-a)^2/12\end{array} \quad (8)$$

where

$$\begin{array}{l}c(p_m, p_v, m, F) = 2p_m(1-p_v)\left(1 - \frac{p_m p_v}{m}\right)F^2 + \\ (1-p_m)^2 + p_m(2 - p_m p_v)\frac{m-1}{m} + p_m^2 p_v(1-p_v)\frac{(m-1)^2}{m^2}\end{array} \quad (9)$$

However, as the mutant vectors Y_i are conditioned to belong to $[a, b]$ (as otherwise they would be replaced with random elements R_i) the contribution of $\mathbb{E}[(Y_i - Y_j)^2]$, $\mathbb{E}[(x_i - Y_j)^2]$, $\mathbb{E}[(Y_i - R_j)^2]$ is not fully given by Eqs. 7 but should be bounded. We consider that the contribution of each of these terms is such that it is not larger than the corresponding fraction of $\text{Var}(X)$. Consequently the slope $c(p_m, p_v, m, F)$ is adjusted as follows:

$$\begin{array}{l}c(p_m, p_v, m, F) = (1-p_m)^2 + p_m p_v(1-p_m)\frac{m-1}{m} + \\ p_m^2(1-p_v)^2 B\left(\frac{m-1}{m} + 2F^2\right) + 2p_m(1-p_m)B\left(\frac{m-1}{m} + F^2\right) + \\ 2p_m^2 p_v(1-p_v)B\left(\frac{(m-1)^2}{2m^2} + \frac{m-1}{m}F^2\right)\end{array} \quad (10)$$

with B denoting a bounding function, i.e. $B(u) = u$ if $u \leq 1$ and $B(u) = 1$ if $u > 1$.

REFERENCES

[1] M.M. Ali, L.P. Fatti, A Differential Free Point Generation Scheme in the Differential Evolution Algorithm, *Journal of Global Optimization*, 35, pp. 551-572, 2006.
[2] J. Arabas, A. Szczepankiewicz, T. Wroniak, Experimental Comparison of Methods to Handle Boundary Constraints in Differential Evolution, *Proc. of PPSN XI*, LNCS 6239, pp. 411-420, 2010.
[3] H.G. Beyer, On the Dynamics of EAs without Selection, *Foundations of Genetic Algorithms*, Eds: W. Banzaf, and C. Reeves, pp. 5-26, 1999.
[4] C. Brown, Y. Jin, M. Leach, M. Hodgson, μJADE: Adaptive Differential Evolution with a Small Population, *Soft Computing*, 20(10), pp. 4111-4120, 2016.
[5] S. Das, P.N. Suganthan, A Survey of the State-of-the-Art, *IEEE Trans on Evol Comput*, 15(1), pp. 4-31, 2011.
[6] S. Das, S.S. Mullick, P.N. Suganthan, Recent advances in differential evolution.An updated survey, *Swarm and Evol Comput*, 27, pp. 130, 2016.
[7] S. Dasgupta, S. Das, A. Biswas, A. Abraham, On stability and convergence of the population-dynamics in differential evolution, *AI Commun.*, 22(1), pp. 1-20, 2009.
[8] S. Ghosh, S. Das, A.V. Vasilakos, K. Suresh, On Convergence of Differential Evolution over a Class of Continuous Functions with Unique Global Optimum, *IEEE Trans. on SMC-B*, 42(1), pp. 107-124, 2012.
[9] S. Helwig, J. Branke, S. Mostaghim, Experimental Analysis of Bound Handling Techniques in Particle Swarm Optimization, *IEEE Trans on Evol Comput*, 17(2), pp. 259-271, 2013.
[10] F. Neri, V. Tirronen, Recent Advances in Differential Evolution: a Review and Experimental Analysis, *Artificial Intelligence Review*, 33(1), pp. 61-106, 2010.
[11] N. Padhye, P. Mittal, K. Deb, Feasibility Preserving Constraint-Handling Strategies for Real Parameter Evolutionary Optimization, arXiv:1504.00442[v] [cs.NE] 17 Apr 2015.
[12] K.V. Price, R.M.Storn, J.A. Lampinen, *Differential Evolution. A Practical Approach to Global Optimization*, Springer, 2005.
[13] C. Segura, C.A. Coello Coello, A.G. Hernandez Diaz, *Improving the Vector Generation Strategy of Differential Evolution for Large-Scale Optimization*, Information Sciences, 323, pp. 106-129, 2015.
[14] R.M. Storn, K.V. Price, Differential Evolution a simple and efficient adaptive scheme for global optimization over continuous spaces, *International Computer Science Institute*, Berkeley, TR-95-012, 1995.
[15] S. Thangavelu, G. Jeyakumar, R.M. Balakrishnan, C.S. Velayutham, Theoretical Analysis of Expected Population Variance Evolution for a Differential Evolution Variant, in L.C. Jain et al. (eds.), *Computational Intelligence in Data Mining*, vol 2, pp. 403-416, 2015.
[16] F. Xue, A.C. Sanderson, R.J. Graves, Multi-objective differential evolution - algorithm, convergence analysis and applications, *Proc. of CEC 2005*, pp. 743-750, 2005.
[17] F. Xue, A.C. Sanderson, Theoretical Analysis of Differential Evolution, in *Adaptive Differential Evolution*, Springer, pp. 15-38, 2009.
[18] D. Zaharie, Critical values for the control parameters of differential evolution algorithms, *Proc. of 8th International Conf on Soft Computing (Mendel)*, Eds: R. Matousek, P.Osmera, pp. 62-67, 2002.
[19] D. Zaharie, Influence of Crossover on the Behavior of Differential Evolution Algorithms,*Applied Soft Computing*, Vol.9, No.3, pp. 1126-1138, 2009.
[20] D. Zaharie, Differential Evolution: from Theoretical Analysis to Practical Insights, *Proc. of 19th International Conf on Soft Computing (Mendel)*, pp.126-131, 2012.
[21] E. Zhabitskaya, Constraints on Control Parameters of Asynchronous Differential Evolution, LNCS 7125, pp. 322-327, 2012.