DIFFERENTIAL EVOLUTION: FROM THEORETICAL ANALYSIS TO PRACTICAL INSIGHTS

Daniela Zaharie

West University of Timişoara
Department of Computer Science
blvd. Vasile Pârvan, no. 4, 300223 Timişoara
Romania
dzaharie@info.uvt.ro

Abstract: *Differential Evolution (DE) is one of the most popular optimization metaheuristics, with a significant number of variants and a large area of applications for real-world problems. The current knowledge on the DE behavior is mainly based on experiments involving benchmark test suites and less on theoretical results. However, in the last years some encouraging results concerning the DE dynamics have been reported. This paper reviews the main theoretical results concerning DE convergence, stability and probability distribution of the populations generated during evolution. A particular emphasis is placed on the theoretical results which lead to practical insights or explain empirical rules concerning the choice of DE control parameters. A new theoretical result concerning the population variance evolution in the case of DE/either-or variant is also presented.*

Keywords: *differential evolution, DE/rand/*/*, DE/current-to-rand/*/*, DE/either-or, binomial crossover, exponential crossover, convergence properties, population variance*

1 Introduction

Differential Evolution (DE) is currently one of the most used population based stochastic metaheuristics. Its popularity is mainly due to its simplicity and effectiveness in solving various types of problems, including multi-objective, multi-modal, dynamic and constrained optimization problems. Since R. Storn and K. Price proposed the first DE variant [11], more than fifteen years ago, dozens of differential evolution flavors, involving changes in the main operators, hybridization with other optimization methods, automated parameters' tuning, (self)-adaptation schemes, structured populations and so on, have been proposed. Excellent surveys containing current status of research in the field of differential evolution were recently published [5], [9]. Despite the large number of reported applications of DE and the huge volume of experimental results it is still difficult to give answers to questions like "How could we control the DE behavior?" or "Why is DE successful for a class of problems and why does it fail for other problems?". The theoretical analysis of DE is still well behind the experimental results, most of the current knowledge on differential evolution being based on empirical observations.

The aim of this paper is to review the existing theoretical results concerning the convergence properties of DE and the influence of the choice of DE parameters on the population evolution, focusing on the usage of these results in deriving practical insights for designing effective and efficient optimization tools. The paper is structured as follows. The first section presents the overall structure of the DE algorithm and the traditional DE mutation and crossover variants. Section 2 reviews the existing theoretical results and presents a new result concerning the population variance evolution in the case of DE/either-or variant. In Section 3 are presented some examples when from theoretical results one can extract suggestions for appropriate choice of DE parameters or one can just explain some rules of thumb already used be practitioners. Section 4 concludes the paper.

2 The general structure of Differential Evolution

The overall structure of differential evolution is typical for evolutionary algorithms consisting of two main steps: initialization and an iterative transformation of a population of candidate solutions belonging to the search space $D \subset \mathbb{R}^n$. Each iteration corresponds to an evolutionary generation and consists of constructing new elements by mutation and crossover, evaluating the new elements and selecting those which are included in the next generation. In the following we shall consider the case of single-objective minimization problems, i.e. problems of finding $x^* \in D$ satisfying $f(x^*) \leq f(x)$ for all $x \in D$, where $f : D \subset \mathbb{R}^n \to \mathbb{R}$. The population of candidate solutions corresponding to generation g will be denoted by $\{x_1(g), x_2(g), \ldots, x_m(g)\}$ and the components of a vector x_i will be denoted by $(x_i^1, x_i^2, \ldots, x_i^n)$. For each element x_i from the current population it is constructed

a mutant element denoted by y_i and from x_i and y_i is constructed by crossover a so-called trial element denoted by z_i. By using these notations the general structure of the differential evolution is described in Algorithm 1, the particularities of a given DE instance being related with the particularities of the mutation and crossover operators. In almost all DE implementations, the selection operator is working as follows: the trial element is compared with the current element and the best of them is transferred in the new population.

Algorithm 1 The general structure of a generational DE

```
Population initialization X(0) ← {x_1(0), ..., x_m(0)}
g ← 0
Compute {f(x_1(g)), ..., f(x_m(g))}
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)
    Compute f(z_i)
    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
```

The DE variants mainly differ with respect to the variation (mutation and crossover) operators. The most used operators are briefly described in the following.

2.1 DE/rand/D/* variant

In the classical DE/rand/1/* variant, in order to construct a mutant vector, y_i, three population elements should be randomly selected. One of these is called base vector, x_{r_1}, and the other two (x_{r_2} and x_{r_3}) are used to construct a difference term to be added to the base vector (see Eq. (1)).

$$y_i = x_{r_1} + F \cdot (x_{r_2} - x_{r_3}), \qquad F \in (0, 2) \tag{1}$$

The constant F in Eq. (1) is a scale factor which influences the diversity of the set of mutant vectors. The indices r_1, r_2, r_3 are usually randomly selected from $\{1, \ldots, m\}$ but such they satisfy the restriction of being distinct and different of i. The influence of this restriction on the DE behavior is analyzed in [8] where is stated that it enables DE to achieve a good convergence. The simplest extension of the DE/rand/1/* mutation variant is that involving several difference terms instead of just one term, as it is illustrated in Eq. (2), where $r_2(d)$ and $r_3(d)$ denote distinct randomly selected indices. The most used variant is that corresponding to $D = 2$ and $F_1 = F_2$, as using more than two differences have not proved to be beneficial in practice.

$$y_i = x_{r_1} + \sum_{d=1}^{D} F_d \cdot (x_{r_2(d)} - x_{r_3(d)}) \tag{2}$$

The trial element, z_i, can be obtained by mixing the elements of the mutant y_i with the elements of the current element x_i. There are two main crossover variants for DE: binomial and exponential. The binomial crossover constructs the trial vector by taking, in a random manner, elements either from the mutant vector or from the current element, as it is described in Eq. (3).

$$z_i^j = \begin{cases} y_i^j & \text{if } U_j < CR \text{ or } j = k \\ x_i^j & \text{otherwise} \end{cases}, j = \overline{1, n} \tag{3}$$

In Eq. (3), U_j denotes a random value generated for each j in accordance to a uniform distribution over $[0, 1]$, $CR \in [0, 1]$ is the crossover rate and k is a randomly selected index from $\{1, \ldots, n\}$. The use of k ensures that at least one component is taken from the mutant vector, even if CR is zero. The name of this crossover type comes from the property that the number of components taken from the mutant vector has a binomial distribution [8]. This binomial crossover is very similar to the so-called uniform crossover used in evolutionary algorithms. The exponential crossover was designed to be similar to one point and two points crossover variants used in

genetic algorithms. The trial vector is obtained by taking from the mutant vector a sequence of consecutive components (in a circular manner), as described in Eq. (4) where $k \in \{1, 2, \ldots, n\}$ is a random index, L is a random value in $\{1, 2, \ldots, n\}$ such that $Prob(L = h) = (1 - CR)CR^{h-1}$ if $h < n$ and $Prob(L = n) = CR^{n-1}$ [16]. In the same equation, $\langle j \rangle_n$ denotes j if $j \leq n$ and $j - n$ if $j > n$.

$$z_i^j = \begin{cases} y_i^j & \text{if } j \in \{k, \langle k+1 \rangle_n, \ldots \langle k+L-1 \rangle_n\} \\ x_i^j & \text{otherwise} \end{cases}, j = \overline{1,n} \tag{4}$$

An obvious difference between binomial and exponential crossover is the fact that in the binomial case the components inherited from the mutant vector are arbitrarily selected, while in the case of exponential crossover they form one or two compact subsequences. Another significant difference concerns the relationship between the parameter CR and the probability of taking an element from the mutant vector (i.e. the mutation probability, p_m) which is linear for the binomial variant and nonlinear for the exponential crossover [16].

2.2 DE/*best/1/* variant

Other common DE variants are those based on changing the randomly selected base vector with linear combinations involving the best element (x_*) and/or the target element (x_i) from the current population. Some of the most popular variants are DE/rand-to-best/1/* (eq. (5)) and DE/current-to-best/1/* (eq. (6)).

$$y_i = \lambda x_* + (1 - \lambda)x_{r_1} + F \cdot (x_{r_2} - x_{r_3}) \tag{5}$$

$$y_i = \lambda x_* + (1 - \lambda)x_i + F \cdot (x_{r_1} - x_{r_2}) \tag{6}$$

In all cases $\lambda \in (0, 1)$ controls the influence of the best or of the current element on the base vector. The trial element, z_i, is obtained by using one of the previously described crossover variants.

2.3 DE/current-to-rand/1 variant

This variant constructs the trial element by using an arithmetic recombination between the target element (x_i) and a mutant element generated using a DE/rand/1 mutation (eq. (7)). The parameter $q \in (0, 1)$ controls the relative influence of the target and mutant elements, respectively, playing a role similar to that of CR in binomial and exponential crossover.

$$z_i = qx_i + (1 - q)x_{r_1} + F \cdot (x_{r_2} - x_{r_3}) \tag{7}$$

2.4 DE/either-or variant

Both binomial and exponential crossover variants have as main disadvantage the fact that they are not rotationally invariant processes making differential evolution less effective for rotated functions. On the other hand, completely eliminating the crossover leads to a poor behavior of DE for multimodal problems. In order to solve this problem, in [8] it is proposed to replace the crossover operator with another way of recombining the population elements, similar to arithmetic recombination, obtaining the so-called DE/either-or variant (eq. (8)). F and K are scaling factors. Some comments on the choice of K are presented later in the paper.

$$z_i = \begin{cases} x_{r_1} + F \cdot (x_{r_2} - x_{r_3}) & \text{with probability } p_F \\ x_{r_1} + K \cdot (x_{r_2} - x_{r_1}) + K \cdot (x_{r_3} - x_{r_1}) & \text{with probability } 1 - p_F \end{cases} \tag{8}$$

3 Theoretical analysis of Differential Evolution

The theoretical results concerning DE behavior are still significantly fewer than the reported experimental studies and real-world applications. One possible explanation is the fact that the theory developed for evolution strategies involving perturbations based on given probability distribution functions (e.g. normal distribution) cannot be directly applied in the case of DE. More specifically, the typical condition of ergodicity of the mutation operator which (together with the elitism of the selection operator) is sufficient to prove the convergence of most of evolution strategies [10] is not necessarily satisfied by the standard DE variants, which do not involve a mutation based on a random variable with probability distribution having an unbounded support. If the scale factor, F, is replaced with a random variable having a normal distribution [14] or if a stochastic term sampled from a distribution with unbounded support is added to the mutant term as in [3], then the sufficient conditions for stochastic convergence are satisfied. In the general case, proving the convergence of DE is not obvious, since

the usage of sampled differences make the study of the DE dynamics more difficult. However, in the last years several important results concerning the distribution of the DE populations, convergence and stability were obtained, even if some of them are under restrictive assumptions [1], [6], [7], [13], [14].

All these results were obtained at a component level, i.e. the population elements have just one component ($n = 1$). This is a natural consequence of the fact that the DE operators act independently on the components of population elements. In the following we shortly review these results, emphasizing, when it is possible, their practical implications.

3.1 Distribution of the trial population and stability results

A first analysis of the distribution of the trial population is presented in [13] for a DE/current-to-best/D variant ($y_i = \lambda x_* + (1-\lambda)x_i + F \cdot \sum_{d=1}^{D}(x_{r_1(d)} - x_{r_2(d)})$) used in the context of multi-objective optimization. The analysis is conducted under the assumption that the initial population has a normal distribution with identity as covariance matrix. The specific properties of DE/current-to-best/D mutation ensure that the normal distribution is conserved also for the population of trial vectors. By analyzing the evolution of the population mean and of the covariance matrix, the authors of [13] extracted two main practical guidelines in choosing the parameters: (i) $2DF^2 + (1-\lambda)^2$ should be larger than one, in order to ensure a proper exploration of the search space; (ii) the greediness coefficient $\lambda \in [0, 1]$ should be relatively large, in order to improve the convergence behavior.

A detailed analysis of the probability distribution of the population of mutant vectors obtained by DE/rand/1 variant was conducted in [1] under the more plausible assumption of an uniform distribution of the initial population. Even if the uniformity of the distribution is not conserved during the intermediate stages of the evolution, the distribution of the population in the last stage of the evolution can be approximated with a uniform distribution in a neighborhood of the global optimum. Based on this assumption, Alli and Fatti succeeded to derive the probability density functions for the mutant vectors and to propose a computationally tractable variant of directly sampling mutants (without using explicit differences), based on an approximation involving a beta distribution. The main practical implications of the variant which does not use explicit differences variant (fde) proposed in [1] are: (i) it does not generate mutants outside the search domain; (ii) it does not use a scale factor, F (however it uses another parameter to estimate the parameters of the beta distribution). Despite these advantages, the fde variant is less used in practice, mainly because its implementation is not so simple as that of classical DE variants.

In [6] is proposed a simple mathematical model for the dynamics of a DE/rand/1/bin population of scalar elements assuming that the objective function is sufficiently smooth and uni-modal in the analyzed region. By using the Lyapunov function method, it is proved that the DE dynamics is stable and with no oscillatory behaviour once that the neighborhood of the optimum has been reached. The main findings are: (i) the DE dynamics is similar with a gradient descent strategy with a descent step value influenced by CR and F; (ii) in the final stage of the evolution, the convergence rate is mainly determined by CR.

The most recent and comprehensive analysis of the evolution of the probability distribution function (PDF) corresponding to DE/rand/1/* populations is presented in [7]. The authors of [7] derived a recurrence relating the PDF corresponding to the current generation to the PDF corresponding to the previous generation. Moreover they have proven that in the case of an objective function with an unique global optimum (but without restrictions on the number of local optima) the PDF converges to a Dirac distribution function concentrated in the global optimum. Therefore, this Dirac function is the equilibrium state of the dynamical system associated to the PDF recurrence relation. In the same paper it is also proved, by identifying a Lyapunov function, that the equilibrium distribution is asymptotically stable. The main practical insight, illustrated in the case of one-dimensional objective functions is: the convergence time (related to the settling time of the Lyapunov function) decreases as CR increases and has a fluctuating, but with an increasing trend, evolution when F increases.

3.2 Influence of the DE parameters on the variance of the trial population

One of the DE particularities is that the population diversity is crucial in avoiding premature convergence and in ensuring the exploration of the search space. This particularity motivated the interest on analyzing the influence of the variation operators and of their parameters on the population variance. Such an analysis for DE was presented in [14] for DE/rand/D/bin and later extended for DE/best-to-rand/D/* and DE/current-to-rand/1 in [15]. The starting idea in computing the component-based expected variance, $E(Var(Z))$, of the trial population, $\{z_1, z_2, \ldots, z_m\}$ is based on Beyer's approach [2] of analyzing the explorative power of evolution strategies in the absence of selection (i.e. for flat objective functions). For all analyzed DE variants the expected variance of the trial population depends linearly on the variance of the current population and the coefficient(s) of this dependence is(are) determined by the parameters of the DE variant. The free term of

the linear dependence is zero (i.e. $E(Var(Z)) = c \cdot Var(X)$), except for variants involving the currently best element.

In the simplest case (DE/rand/1/*), the dependence described by Eqs. (9) and (10) involves in an interrelated manner all parameters (m, F, CR and n) suggesting that each of these parameters has an influence on the diversity of the population of trial elements.

$$E(Var(Z)) = \left(1 + 2p_m F^2 - \frac{p_m(2-p_m)}{m}\right) Var(X) \tag{9}$$

In Eq. (9) the value of the mutation probability, p_m, depends on the crossover rate parameter, CR, as follows [16]:

$$p_m = \begin{cases} CR(1-1/n) + 1/n & \text{for binomial crossover} \\ \frac{1-CR^n}{n(1-CR)} & \text{for exponential crossover} \end{cases} \tag{10}$$

When the currently best element is used (e.g. DE/rand-to-best/*) then the dependence of $E(Var(Z))$ on $Var(X)$ (see Eq. (11)) involves a free term expressing how far the elements of the current population are spread around the best element [15]. If x^* is close to the population mean then one obtains again a simple linear dependence based on a coefficient involving the parameters: λ, m, F, CR and n. In Eq.(11) p_m depends on CR and n according to eq. (10).

$$E(Var(Z)) = \left(1 + 2p_m F^2 - \frac{p_m(2-p_m)}{m} - \lambda p_m^2 \frac{m-1}{m}\right) E(Var(X)) + \lambda^2 p_m(1-p_m)\frac{1}{m}\sum_{i=1}^{m}(x_* - x_i)^2 \tag{11}$$

In the case of the usage of arithmetic recombination, the relationship between $E(Var(Z))$ and $Var(X)$ is rather simple (Eq. (12)) and the multiplicative factor involves only the parameters m, F, q, being independent of the problem size [15].

$$E(Var(Z)) = \left(1 + 2F^2 - 2(1-q) + \frac{2m-1}{m}(1-q)^2\right) Var(X) \tag{12}$$

The most complicated expression arises in the case of DE/either-or variant (Eq. (13)) where the multiplicative coefficient depends on m, F, K and p_F, being again independent of the problem size. This is a new result, as the behaviour of the DE/either-or population variance has not been analyzed before. A sketch of the proof of Eq. (13) is presented in the Appendix.

$$\begin{aligned} E(Var(Y)) \quad = \quad & \left(p_F^2(1 + 2F^2 - \tfrac{1}{m}) + 2p_F(1-p_F)(\tfrac{m-1}{m} + F^2 + 3K^2 - 2K)\right. \\ & \left. +(1-p_F)^2\left(\tfrac{m-1}{m} + 2\tfrac{m-2}{m}(3K^2 - 2K)\right)\right) Var(X) \end{aligned} \tag{13}$$

All these results, providing explicit relationships between the DE control parameters and the evolution of the population variance (in the absence of selection) can be used to derive values for the DE parameters which induces a desired evolution of the population variance (e.g. avoid premature convergence).

4 From DE theory to practical insights

Theoretical results are valuable for practitioners as long as they can lead to rules to be applied in practice. In this section we will see how can we use some of the theoretical results presented in the previous section in order to explain some empirical rules concerning the choice of DE parameters and also to infer some other practical remarks.

4.1 Avoiding premature convergence

The main cause of DE premature convergence is the lost of population diversity, thus ensuring that the variance of the population does not decrease in the absence of selective pressure is a key element in avoiding premature convergence. Except for DE/best/* variants, the trial population variance can be controlled by choosing parameters' values which ensure that the coefficient c (in $E(Var(Z)) = c \cdot Var(X)$) is never smaller than 1. Since the scale factor, F, is usually related to the diversity of the trial elements, one can find lower bounds for F under which the population variance decreases even in the absence of the selective pressure. Such lower bounds can be computed in the case of DE/rand/1/*, DE/current-to-rand and DE/either-or by solving the equations $c(m, F, CR, n) = 1$, $c(m, F, q) = 1$ and $c(m, F, K, p_F) = 1$, respectively. Figure 1 illustrates the influence of the

population size (m) and of the crossover parameter (CR for DE/rand/1/*, q for DE/current-to-rand and p_F for DE/either-or) on the lower bound, F_{low}. In the case of DE/rand/1/* variants the increase of population size significantly decreases the value of F_{low}, meaning that for large enough populations (e.g. $m \geq 100$) the variance does not decrease in the absence of selective pressure, unless for small values of F (e.g. $F < 0.1$). In the case of DE/current-to-rand and DE/either-or the increase of population size does not ensure (unless the crossover parameters, q and p_F, are large) the decrease of F_{low} (see Figure 1). On the other hand, for DE/rand/1/* and DE/either-or the lower bound F_{low} decreases as the crossover parameter increases while in the case of DE/current-to-rand the dependence of F_{low} on q is nonmonotonous (with a maximal value for F_{low} when $q = 0.5$). From a practical point of view the knowledge of F_{low} is more important in the case of small populations. A succesfull practical usage of the lower bound F_{low} has been reported in [4] in the context of dynamic optimization. Figure 2 presents a comparative image of the influence of the control parameters on the trial population variance suggesting that DE/rand/1/* variants are less sensitive to the value of the scale factor, F, than DE/current-to-rand and DE/either-or.

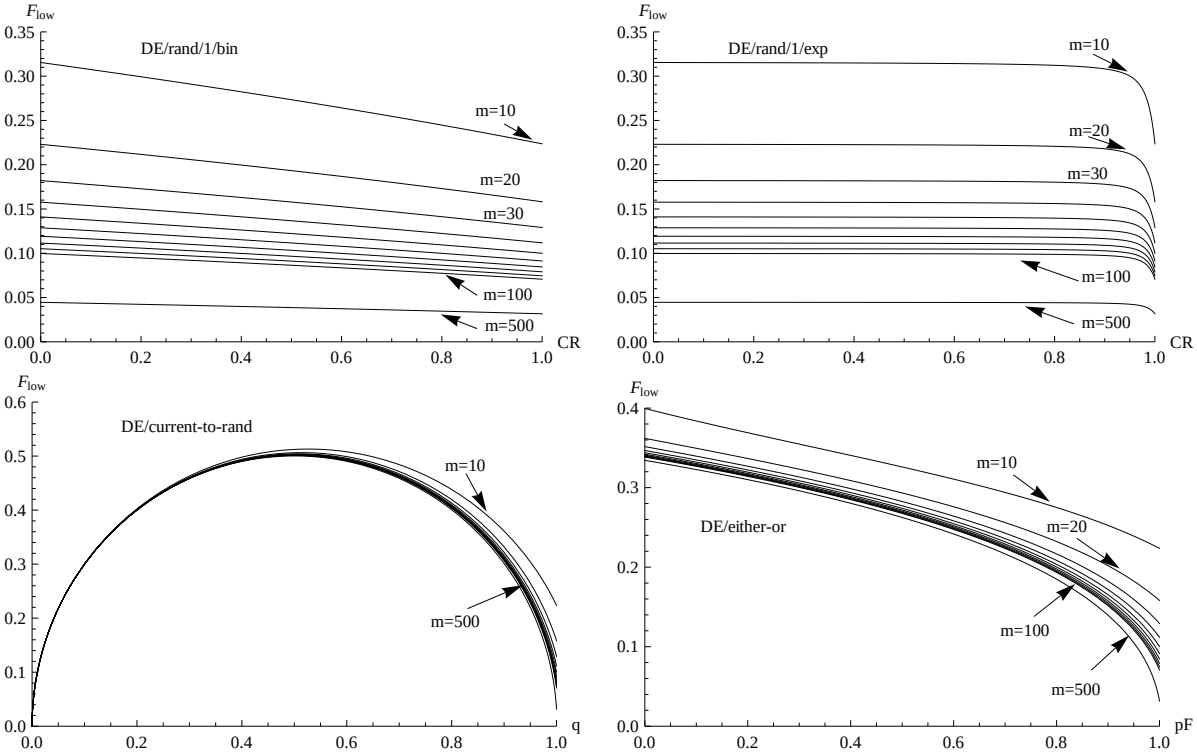

Figure 1: Influence of the population size (m) and of the crossover parameter (CR, q, p_F) on the lower bound for F.

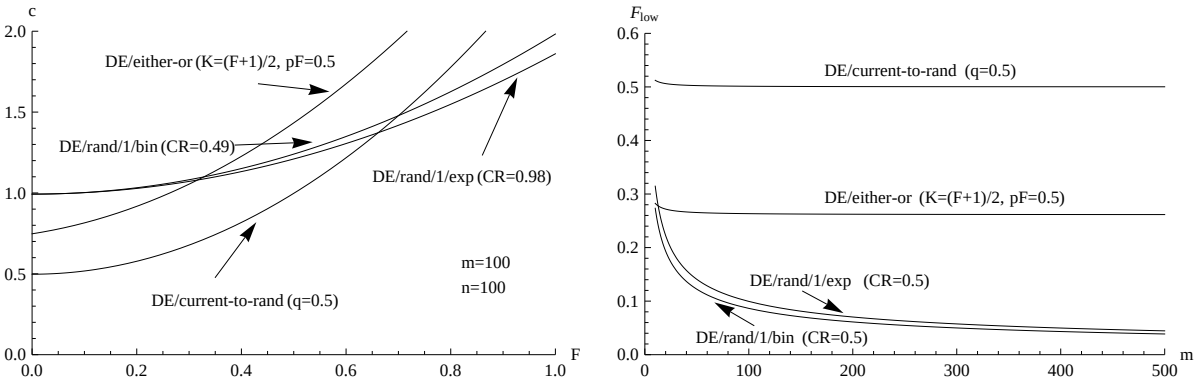

Figure 2: Comparison between DE variants ($p_m = q = p_F = 0.5$): dependence of the multiplicative coefficient of the trial population variance on F (left); dependence of the lower bound for F on m (right).

4.2 Choosing values for the crossover rate

In the case of DE/rand/*/* variants, the crossover rate (CR) controls the number of components taken from the mutant vector but it is not identical to the mutation probability (p_m). It is obvious that the DE behaviour is influenced by p_m but we can control it only through CR (based on the Eq. 10). Therefore it is important to know which value should have CR in order to obtain a desired value for $p_m \in [1/n, 1]$. In the case of binomial crossover, CR can be easily computed from p_m since $CR = (p_m - 1/n)/(1 - 1/n)$. In the case of exponential crossover one have to solve a polynomial equation ($CR^n - np_mCR + np_m - 1 = 0$). This approach was used in [12] in order to design adaptive DE involving a selection between binomial and exponential crossover. Another practical issue is related to the influence of the problem size on the choice of appropriate CR. While the binomial crossover is only slightly influenced by the problem size, the exponential crossover is more sensitive to the problem size. For high-dimensional problems the range of values for CR for which p_m is significant is very narrow, therefore in the case of exponential crossover it could be difficult to identify appropriate CR values.

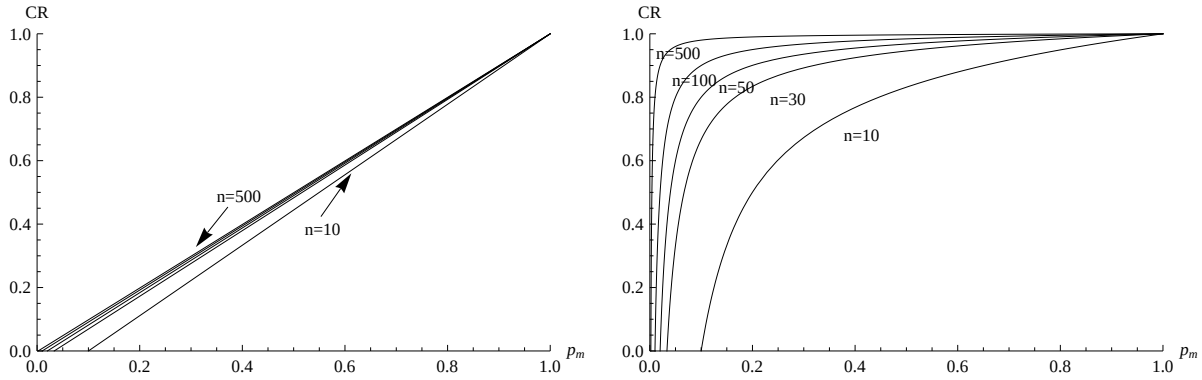

Figure 3: The relationship between CR and p_m (for $n \in \{10, 30, 50, 100, 500\}$) for binomial crossover (left) and exponential crossover (right).

4.3 DE empirical rules vs. theoretical results

In the DE field it is common to extract from numerical experiments some empirical knowledge on the algorithm behaviour, leading to rules used in implementations. It would be ideal if such empirical rules would be in accordance with some theoretical results. One such situation is presented in [16] where the theoretical results on exponential crossover explains why appropriate values for CR should be in the range $[0.9, 1]$.

Here we present another situation where an empirically induced rule has some theoretical explanation behind. In the case of DE/either-or variant it is stated in [8] that *"From experience $K = 0.5 \cdot (F+1)$ can be recommended as a good first choice for K given F."*. A natural question is if such a choice of K has some particular properties. Well, if we look to Figure 4 (left) we can see that for $F = 0.8$ and $K = 0.9$ the coefficient $c(m, F, K, p_F)$ involved in the variance relationship (eq. (13)) has almost the same value for all values of p_F. Same behaviour can be remarked for $F = 0.5$ and $K = 0.75$, for $F = 0.6$ and $K = 0.8$ and so on. In fact, by solving the equation $c(m, F, K, p_F) = c(m, F, K, p'_F)$ with respect to K for some given values for m, p_F, p'_F and for $F \in \{0.1, 0.2, \ldots, 2\}$ we obtained values for K corresponding to the points marked in Figure 4 (right) which are close to the line corresponding to the dependence $K = (F+1)/2$. Therefore, we can say that for $K = (F+1)/2$ the variance of the trial population is not significantly influenced by p_F. Thus, in this case DE/either-or could be less sensitive to p_F, therefore being easier to choose an appropriate value for it.

5 Conclusions

Despite the fact that the DE theory is still behind the numerical studies there are some encouraging results. Moreover most of these results provide rules in choosing the DE parameters which allow to control the algorithm behavior. On the other hand, most of these results were obtained for panmictic populations involving classical mutation and crossover operators. There are a lot of DE variants for which the theoretical analysis is still an open issue (e.g. DE with structured populations, neighborhood based selection of parents, opposition-based DE, binary differential evolution etc).

Acknowledgement: This work was partially supported by the strategic grant POSDRU/21/1.5/G/13798,

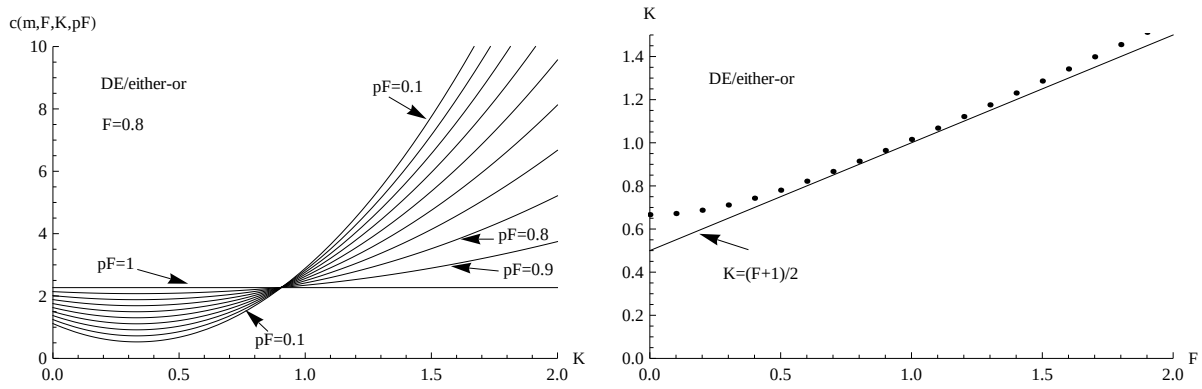

Figure 4: The relationship between the parameters of DE/either-or (F, K and p_F) and the multiplicative coefficient which controls the variance of the trial population.

inside POSDRU Romania 2007-2013, co-financed by the European Social Fund Investing in People.

References

[1] Ali, M.M., Fatti, L.P., A Differential Free Point Generation Scheme in the Differential Evolution Algorithm, *Journal of Global Optimization*, Vol. 35, 2006, pp. 551-572.

[2] Beyer, H.G., On the Dynamics of EAs without Selection, *Foundations of Genetic Algorithms*, Eds: Banzaf, W. and Reeves, C., 1999, pp. 5-26.

[3] ter Braak, C.J.F, A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces, *Stat. Comput.*, Vol. 16, 2006, pp. 239-249.

[4] Brest, J., Zamuda, A., Boskovic, B., Maucec, M. S., Zumer, V., Dynamic optimization using self-adaptive differential evolution, *Proc. IEEE Congr. Evol. Comput.*, 2009, pp. 415-422.

[5] Das, S., Suganthan, P.N., A Survey of the State-of-the-Art, *IEEE Trans. on Evolutionary Computation*, Vol. 15, No. 1, 2011, pp. 4-31.

[6] Dasgupta, S., Das, S., Biswas, A., Abraham, A., On stability and convergence of the population-dynamics in differential evolution, *AI Commun.*, Vol. 22, No. 1, 2009, pp. 1-20.

[7] Ghosh, S., Das, S., Vasilakos, A.V., Suresh, K., On Convergence of Differential Evolution over a Class of Continuous Functions with Unique Global Optimum, *IEEE Trans. on Systems, Man, and Cybernetics - part B*, Vol. 42, No. 1, 2012, pp. 107-124.

[8] Price, K.V., Storn, R.M., Lampinen, J.A., *Differential Evolution. A Practical Approach to Global Optimization*, Springer, 2005.

[9] Neri F., Tirronen, V, Recent Advances in Differential Evolution: a Review and Experimental Analysis, *Artificial Intelligence Review*, Vol.33, No. 1, 2010, pp. 61-106.

[10] Rudolph, G., Convergence of Evolutionary Algorithms in General Search Spaces, *Proceedings of the 3rd Congress of Evolutionary Computation, (CEC 1996)*, 1996, pp. 50-54.

[11] Storn R., Price, K., Differential Evolution a simple and efficient adaptive scheme for global optimization over continuous spaces. International Computer Science Institute, Berkeley, TR-95-012, 1995.

[12] Tvrdik, J., Adaptive Differential Evolution and Exponential Crossover, *Proceedings of International Multiconference on Computer Science and Information Technology (IMCSIT 2008)*, 2008, pp. 927-931.

[13] Xue, F., Sanderson, A.C., Graves, R.J., Multi-objective differential evolution - algorithm, convergence analysis and applications, *Proceedings of Congress on Evolutionary Computation (CEC 2005)*, 2005, pp. 743-750.

[14] Zaharie, D., Critical values for the control parameters of differential evolution algorithms, *Proceedings of 8th International Conference on Soft Computing (Mendel 2002)*, Eds: R. Matousek, P.Osmera, 2002, pp. 62-67.

[15] Zaharie, D., Statistical properties of differential evolution and related random search algorithms. *Proceedings in Computational Statistics (Compstat 2008)*, Ed: P. Brito, 2008, pp. 473-485.

[16] Zaharie, D., Influence of Crossover on the Behavior of Differential Evolution Algorithms,*Applied Soft Computing*, Vol.9, No.3, 2009, pp. 1126-1138.

Appendix: Sketch of the proof for Eq. (13) In order to simplify the notations let us denote by I_i, J_i and K_i the random variables with uniform distribution on $\{1, 2, \ldots, m\}$ used to sample the indices r_1, r_2 and r_3 involved in the description of the DE/either-or variant (see Eq. (8)). Therefore the equation describing the dependence of the random variable corresponding to the trial element, Z_i, on the elements from the current population can be rewritten as in Eq. (14).

$$Z_i = \begin{cases} V_i = x_{I_i} + F \cdot (x_{J_i} - x_{K_i}) & \text{with probability } p_F \\ W_i = x_{I_i} + K \cdot (x_{J_i} + x_{K_i} - 2x_{I_i}) & \text{with probability } 1 - p_F \end{cases} \tag{14}$$

If we denote with 1_{M_i} the mutation event (the event of taking V_i) then the random variable Z_i can be written as

$$Z_i = V_i \cdot 1_{M_i} + W_i \cdot 1_{\overline{M_i}}. \tag{15}$$

Our aim is to compute the expected variance of the trial population, i.e. $E(Var(Z))$. Since

$$Var(Z) = \frac{1}{m^2} \sum_{i<j} (Z_i - Z_j)^2 \tag{16}$$

and the distribution of $(Z_1, Z_2, \ldots, Z_m)$ is the same as the distribution of any permutation, $(Z_{\pi(1)}, Z_{\pi(2)}, \ldots, Z_{\pi(m)})$, of the population ($\pi$ being a permutation of $\{1, 2, \ldots, m\}$) it follows that:

$$E[Var(Z)] = \frac{m-1}{2m} E[(Z_i - Z_j)^2] \tag{17}$$

i and j being two arbitrary but distinct indices. Using Eq. (15) and the fact that $Prob(1_{M_i}) = p_F$ and $Prob(1_{\overline{M_i}}) = 1 - p_F$ it follows that:

$$E[Var(Z)] = \frac{m-1}{2m} \left(p_F^2 E[(V_i - V_j)^2] + 2p_F(1-p_F) E[(V_i - W_j)^2] + (1-p_F)^2 E[(W_i - W_j)^2] \right). \tag{18}$$

The first term of the sum appearing in the right hand part of Eq. (18) is also involved in the DE/rand/1 mutation (without crossover), thus according to results in [15] it satisfies:

$$E[(V_i - V_j)^2] = \frac{2m}{m-1}(1 + 2F^2 - 1/m) E[Var(X)]. \tag{19}$$

Let us analyze the second term:

$$\begin{aligned} E[(V_i - W_j)^2] &= E[((x_{I_i} - x_{I_j}) + F \cdot (x_{J_i} - x_{K_i}) - K \cdot (x_{J_j} + x_{K_j} - 2x_{I_j}))^2] \\ &= E[(x_{I_i} - x_{I_j})^2] + F^2 \cdot E[(x_{J_i} - x_{K_i})^2] + K^2 \cdot E[(x_{J_j} + x_{K_j} - 2x_{I_j})^2] \\ &\quad + 2F \cdot E[(x_{I_i} - x_{I_j}) \cdot (x_{J_i} - x_{K_i})] - 2K \cdot E[(x_{I_i} - x_{I_j}) \cdot (x_{J_j} + x_{K_j} - 2x_{I_j})] \\ &\quad - 2K \cdot F \cdot E[(x_{J_i} - x_{K_i}) \cdot (x_{J_j} + x_{K_j} - 2x_{I_j})] \end{aligned} \tag{20}$$

In order to conduct all computations involved in Eq. (20) the following identities will be used (for I and J random variables over $\{1, 2, \ldots, m\}$):

$$\begin{aligned} &E(x_I - x_J) = 0 \qquad E[(x_I - x_J)^2] = 2E[Var(X)] \qquad \text{(if } I \text{ and } J \text{ can take the same values)} \\ &\qquad\qquad\quad E[(x_I - x_J)^2] = \tfrac{2m}{m-1} E[Var(X)] \qquad \text{(if } I \text{ and } J \text{ take always distinct values)} \end{aligned} \tag{21}$$

Since I_i and I_j are independent they can take the same values, while according to DE rule, J_i and K_i should always take distinct values. Therefore one obtains:

$$E[(x_{I_i} - x_{I_j})^2] = 2E[Var(X)] \quad \text{and} \quad E[(x_{J_i} - x_{K_i})^2] = \frac{2m}{m-1} E[Var(X)] \tag{22}$$

On the other hand, using the independence between the set of random variables $\{I_i, J_i, K_i\}$ and the set $\{I_j, J_j, K_j\}$, it follows:

$$E[(x_{I_i} - x_{I_j})(x_{J_i} - x_{K_i})] = 0 \quad \text{and} \quad E[(x_{J_i} - x_{K_i})(x_{J_j} + x_{K_j} - 2x_{I_j})] = 0 \tag{23}$$

On the other hand, the third term in Eq. (20) can be written as follows:

$$\begin{aligned} E[(x_{J_j} + x_{K_j} - 2x_{I_j})^2] &= E[(x_{J_j} - x_{I_j})^2] + E[(x_{K_j} - x_{I_j})^2] + 2E[(x_{J_j} - x_{I_j})(x_{K_j} - x_{I_j})] \\ &= \tfrac{4m}{m-1}E[Var(X)] + 2E[x_{J_j}(x_{K_j} - x_{I_j})] - 2E[x_{I_j} \cdot x_{K_j} - x_{I_j}^2] \\ &= \tfrac{4m}{m-1}E[Var(X)] + \tfrac{2m}{m-1}E[Var(X)] = \tfrac{6m}{m-1}E[Var(X)] \end{aligned} \quad (24)$$

The last computations in the above equation are based on the following (easy to check) equalities: $E[x_{J_j}(x_{K_j} - x_{I_j})] = 0$ and $E[x_{I_j} \cdot x_{K_j} - x_{I_j}^2] = -m/(m-1) \cdot E[Var(X)]$. On the other hand, since I_i and I_j are independent random variables it follows that $E[(x_{I_i} - x_{I_j})(x_{J_i} + x_{K_j} - 2x_{I_j})] = -2m/(m-1) \cdot E[Var(X)]$. By adding all these terms one obtains:

$$E[(V_i - W_j)^2] = \left(2 + \frac{2m}{m-1}F^2 + \frac{6m}{m-1}K^2 - \frac{4m}{m-1}K\right) E[Var(X)] \quad (25)$$

The last term in Eq. (18) is:

$$\begin{aligned} E[(W_i - W_j)^2] &= E[((1-2K)(x_{I_i} - x_{I_j}) + K \cdot (x_{J_i} - x_{J_j}) + K \cdot (x_{K_i} - x_{K_j}))^2] \\ &= (1-2K)^2 \cdot E[(x_{I_i} - x_{I_j})^2] + K^2 \cdot E[(x_{J_i} - x_{J_j})^2] + K^2 \cdot E[(x_{K_i} - x_{K_j})^2] \\ &\quad +2K(1-2K) \cdot E[(x_{I_i} - x_{I_j}) \cdot (x_{J_i} - x_{J_j})] + 2K(1-2K) \cdot E[(x_{I_i} - x_{I_j}) \cdot (x_{K_i} - x_{K_j})] \\ &\quad +2K^2 \cdot E[(x_{J_i} - x_{J_j}) \cdot (x_{K_i} - x_{K_j})] \end{aligned} \quad (26)$$

Since the pairs (I_i, I_j),(J_i, J_j) and (K_i, K_j) contain independent random variables it follows that:

$$E[(x_{I_i} - x_{I_j})^2] = E[(x_{J_i} - x_{J_j})^2] = E[(x_{K_i} - x_{K_j})^2] = 2E[Var(X)] \quad (27)$$

and

$$E[(x_{I_i} - x_{I_j})(x_{J_i} - x_{J_j})] = E[(x_{I_i} - x_{I_j})(x_{K_i} - x_{K_j})] = E[(x_{J_i} - x_{J_j})(x_{K_i} - x_{K_j})] = \frac{2}{m-1}E[Var(X)] \quad (28)$$

Thus by adding all terms in Eq. (26) it follows:

$$E[(W_i - W_j)^2] = \left(\left(4 - \frac{4}{m-1}\right)(3K^2 - 2K) + 2\right) E(Var(X)) \quad (29)$$

Using Eqs. (19),(20) and (26) in Eq. (18) it follows Eq. (13). Figure 5 illustrates the similarities between the theoretical evolution of the trial population variance (in the absence of selection pressure) and the empirically estimated variance. Also the theoretical critical value $F = 0.274$ is close to the empirically estimated one.

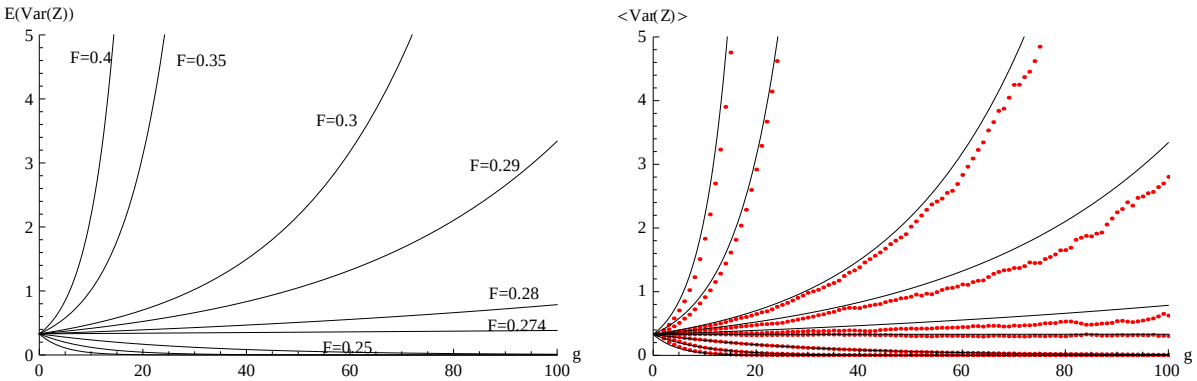

Figure 5: The evolution of the trial population variance (in the absence of selection) for DE/either-or ($m = 50$, $K = (F+1)/2$, $p_F = 0.5$): theoretical evolution (left) and empirical plus theoretical evolution (right).