CRITICAL VALUES FOR THE CONTROL PARAMETERS OF DIFFERENTIAL EVOLUTION ALGORITHMS

Daniela Zaharie

Department of Computer Science, Faculty of Mathematics, West University of Timişoara
bv. V. Pârvan, no. 4, 1900 Timişoara, ROMANIA
tel./fax.: 04056194002, e-mail:dzaharie@info.uvt.ro

Abstract. *The population diversity plays an important role in the behavior of evolution strategies. This paper analyzes, both from a theoretical and an empirical viewpoint, the relationship between the control parameters of differential evolution algorithms and the evolution of population variance. Using this relationship, values of the control parameters for which premature convergence can be prevented are obtained.*

Keywords: global optimization, differential evolution, premature convergence, population variance

1. Introduction. The Differential Evolution method (DE) proposed by Storn and Price [7] proved to be a powerful global optimization technique. Besides its good convergence properties and suitability for parallelization, DE's main assets are its conceptual simplicity and ease of use. Since its invention, DE method has been applied to different optimization problems (for a review of its applications see [3]) and different variants have been proposed. Some of these variants improve the efficiency of implementation [4], others extend its applicability [1]. Although DE has been tested extensively against artificial and real world optimization problems, only a few theoretical results concerning its behavior have been obtained.

One of the main problems in evolution strategies design is to choose the control parameters such that the strategy exhibits a good behavior, i.e. it does not prematurely converge or stagnate [5] and has an acceptable rate of convergence toward the global optimum. In this paper we deal with the problem of premature convergence and with the influence of the control parameters on the ability of DE to avoid such a situation. Although the premature convergence can occur in any evolution strategy, the DE is more likely to be affected by such a problem due to the fact that it does not use external information during the evolution. More specifically, the mutation operator used in classical evolution strategy is based on a random variable while the perturbation used in DE is based on differences between population elements. When the population lost completely its diversity (it contains identical elements) it remains unchanged by DE perturbation. Thus, to avoid premature convergence, it seems natural to keep a reasonable level of diversity in the population. Measures of the population diversity are the statistical variances computed for each component over the entire population.

The aim of this paper is to find:

- a relationship between the control parameters of DE and the population variance evolution;
- a relationship between the population variance evolution and the convergence properties.

Once we have answers to these problems we can control the DE behavior by controlling the population diversity through an appropriate choice of the parameters. In fact we will establish values for the control parameters such that the premature convergence is prevented.

The paper is organized as follows. In Section 2 the structure of the variant of DE algorithm which we analyze is presented. A theoretical analysis of the influence of variation operators (DE mutation and crossover) on the expected variance of a population of scalars is made in Section 3. The next section is dedicated to an experimental study on the influence of the population diversity and of the control parameters on the convergence properties of the algorithm. Section 5 concludes the work.

2. The structure of the analyzed algorithm. To fix the ideas, we shall consider the problem of finding a global minimum, x^*, of a function $f : D \subset \mathbb{R}^n \to \mathbb{R}$. The DE algorithm evolves a fixed size population $P = \{x_1, x_2, \ldots, x_m\}$, $x_l \in D$, $l = \overline{1,m}$. The population is initialized with elements randomly selected from D. After initialization an iterative process is started and at each iteration (called generation) a new population is produced. The iterative process cycles until a stopping criterion is satisfied (e.g. a maximum number of generations have been passed or no significant improvements are made). The structure of algorithm which we will analyze is presented in Fig. 1.

In this algorithm α_1^l, α_2^l, α_3^l are distinct elements of $\{1, 2, \ldots, m\}$ randomly selected for each l, $\widetilde{F_l^i}$ is a parameter which weights the "differential perturbation" and $p \in (0, 1]$ is a parameter which controls the proportion of perturbed elements in the new population. In the original DE [7] some restrictions are imposed: $\alpha_i^l \neq l$ for all i and at least one component of z_l is perturbed. To simplify the analysis, these restrictions are

`Initialize` $P(0) = \{x_1(0), \ldots, x_m(0)\}$, $g = 0$
`Repeat`
Mutation step:

$$y_l^i = x^i_{\alpha_1^l}(g) + \widetilde{F_l^i} \cdot (x^i_{\alpha_2^l}(g) - x^i_{\alpha_3^l}(g)), \quad l = \overline{1,m}, \quad i = \overline{1,n}$$

Crossover step:

$$z_l^i = \begin{cases} y_l^i & \text{with probability } p \\ x_l^i(g) & \text{with probability } 1-p \end{cases} \qquad l = \overline{1,m}, \quad i = \overline{1,n}$$

Selection step:

`If` $f(z_l) < f(x_l(g))$ `then` $x_l(g+1) = z_l$ `else` $x_l(g+1) = x_l(g) \qquad l = \overline{1,m}$

$g = g + 1$ (increment the generation counter)
`Until a stopping criterion is satisfied.`

Figure 1: The DE algorithm structure

ignored in our considerations without altering essentially the DE abilities. The terminology on DE [7] does not contain terms like mutation or crossover. We used them here only to identify the main steps of the algorithm. On the other hand, in the original DE the perturbation factor is constant: $\widetilde{F_l^i} = F > 0$, $l = \overline{1,m}$, $i = \overline{1,n}$. We shall consider, also, the variant with

$$\widetilde{F_l^i} = F\xi_l^i, \quad F > 0, \quad \xi_l^i \sim N(0,1) \tag{1}$$

where the random variables ξ_l^i are independent and identically distributed. Such a choice for $\widetilde{F_l^i}$ has been used in [1] for multi-objective optimization problems. As we shall see on Section 4, the numerical experiments show that the random character of $\widetilde{F_l^i}$ neither improves nor deteriorates essentially the algorithm convergence. The reason of analyzing this variant is rather theoretical. There exist results on the convergence of evolution strategies which state that if the evolution operators satisfy some properties then the algorithm converges in a probabilistic sense (see for example [6]). Roughly speaking, the result in [6] says that if the mutation operator has an ergodic character (i.e. the probability of transition, through mutation, from a given element of D to any other element of D is strictly positive) and the selection operator is elitist (i.e. the best element of the population obtained through selection is as good as the best element of the previous population) then the algorithm converges almost sure to a ϵ-vicinity, A_ϵ, of x^*. A_ϵ is defined as follows: $A_\epsilon = \{x \in D^m | f(x_*) - f(x^*) < \epsilon\}$ where x_* is the best element of the population x. These conditions are only sufficient but not necessary. Rudolph's result does not help us to find out something about the convergence of the classical DE. On the other hand, the variant with (1) satisfies the sufficient conditions for convergence except for the case when at least one component of the vectors which represent the population has the same values over the entire population. In this case, the transition to another value of this component is impossible. But the set of populations having this property is of zero measure, thus it can be neglected (in the theoretical analysis). However this property has not the same relevance in practical implementations. Due to the finite precision of floating point values, the set D^m is implemented through a discrete set, thus we can not neglect the populations having zero variance across at least one component. Hence it seems useful to control the variance of each component over the entire population.

3. Theoretical analysis of the expected population variance. We analyze the influence of the mutation and crossover operators on the population variance, at a component level. Similar analysis has been made for a class of evolution strategies in [2] in the context of analyzing their explorative power.

Since the mutation and crossover operators act similarly and independently on all components it suffices to restrict the analysis to populations of scalar elements, $x = \{x_1, \ldots, x_m\}$. The population individuals being affected by random elements we can only compute the expected value of the variance: $E(\mathrm{Var}(x))$, where E is the average with respect to all random elements involved in the population and $\mathrm{Var}(x) = \overline{x^2} - \overline{x}^2$, $\overline{x} = \sum_{l=1}^m x_l/m$. To make the difference between the deterministic elements and the random ones, the first will be denoted by small letters while the second will be denoted by capital letters. The main theoretical result of this paper is the following theorem.

Theorem. *Let $x = \{x_1, \ldots, x_m\}$ be the current population, $Y = \{Y_1, \ldots, Y_m\}$ the intermediate population obtained after applying the mutation and $Z = \{Z_1, \ldots, Z_m\}$ the population obtained by crossing over the populations x and Y. Both in case of constant $\widetilde{F}_l$ and for $\widetilde{F}_l$ given by (1) the following statements hold:*

(a) after mutation the expected population variance becomes

$$E(Var(Y)) = (2F^2 + \frac{m-1}{m})\, Var(x). \tag{2}$$

(b)after mutation and crossover the expected population variance becomes

$$E(Var(Z)) = (2F^2 p - \frac{2p}{m} + \frac{p^2}{m} + 1)\, Var(x). \tag{3}$$

Proof. (a) Each element, Y_l of the population Y is obtained as $Y_l = x_{\alpha_1^l} + \widetilde{F}_l \cdot (x_{\alpha_2^l} - x_{\alpha_3^l})$ where α_1^l, α_2^l and α_3^l are three elements of $\{1, 2, \ldots, m\}$ selected through a random sampling without replacement. This type of sampling assures that the selected indices are different one to each other. Since sampling properties do not depend on the index l, to simplify the notations we shall drop the index l. It will be specified only if it is necessary. The indices α_1, α_2 and α_3 can be viewed as random variables with values in $\{1, 2, \ldots, m\}$ and they have the property: $P(\alpha_i = k) = 1/m$ for each $i \in \{1, 2, 3\}$ and $k \in \{1, \ldots, m\}$. Thus x_{α_i} is also a random variable and it is easy to verify that $E(x_{\alpha_i}) = \overline{x}$ and $E(x_{\alpha_i}^2) = \overline{x^2}$. The variables α_1, α_2, α_3 are not mutually independent, since their values must satisfy the condition of being distinct. Thus $P(\alpha_i = k, \alpha_j = k') = 1/(m^2 - m)$ and $E(x_{\alpha_i} x_{\alpha_j}) = (m\overline{x}^2 - \overline{x^2})/(m-1)$. On the other hand it is obvious that for all l it holds: $E(\widetilde{F}_l) = 0$ and $\mathrm{Var}(\widetilde{F}_l) = E(\widetilde{F}_l{}^2) = F^2$. Also the random variables $\widetilde{F}_l$ are independent with respect to the random variables $x_{\alpha_i^l}$ for any i and l.

We compute now $E(\overline{Y^2})$.

$$E(\overline{Y^2}) = \frac{1}{m} E\left[\sum_{l=1}^{m} (x_{\alpha_1^l} + \widetilde{F}_l \cdot (x_{\alpha_2^l} - x_{\alpha_3^l}))^2\right] = \frac{1}{m} E\left[\sum_{l=1}^{m} ((x_{\alpha_1^l})^2 + 2\widetilde{F}_l x_{\alpha_1^l}(x_{\alpha_2^l} - x_{\alpha_3^l}) + \widetilde{F}_l{}^2 (x_{\alpha_2^l} - x_{\alpha_3^l})^2)\right] =$$

$$= \overline{x^2} + \frac{F^2}{m} \sum_{l=1}^{m} (\overline{x^2} - 2E(x_{\alpha_2^l} x_{\alpha_3^l}) + \overline{x^2}) = (2F^2 \frac{m}{m-1} + 1)\overline{x^2} - 2F^2 \frac{m}{m-1} \overline{x}^2. \tag{4}$$

To obtain the second equality we used the property $E(x_{\alpha_1} x_{\alpha_2}) = E(x_{\alpha_1} x_{\alpha_3})$ when $\widetilde{F}_l = F$ and the properties of $\widetilde{F}_l$ when the variant with (1) is used. The second term which we have to compute is $E(\overline{Y}^2)$. Using the properties of x_{α_1}, x_{α_2}, x_{α_3} and $\widetilde{F}_l$ we obtain:

$$E(\overline{Y}^2) = \frac{1}{m^2} E\left[\sum_{l=1}^{m} (x_{\alpha_1^l} + \widetilde{F}_l (x_{\alpha_2^l} - x_{\alpha_3^l}))\right]^2 =$$

$$= \frac{1}{m^2} E\left[\sum_{l=1}^{m} (x_{\alpha_1^l} + \widetilde{F}_l (x_{\alpha_2^l} - x_{\alpha_3^l}))^2\right] + \frac{1}{m^2} E\left[\sum_{k \neq l} (x_{\alpha_1^l} + \widetilde{F}_l (x_{\alpha_2^l} - x_{\alpha_3^l}))(x_{\alpha_1^k} + \widetilde{F_k}\, (x_{\alpha_2^k} - x_{\alpha_3^k}))\right].$$

The first term of $E(\overline{Y}^2)$ is $E(\overline{Y^2})/m$. Since for each element of Y the indices α_1, α_2 and α_3 are obtained by a new sampling, independent from the other ones, it follows that for $k \neq l$ the variables $x_{\alpha_1^l} + \widetilde{F}_l (x_{\alpha_2^l} - x_{\alpha_3^l})$ and $x_{\alpha_1^k} + \widetilde{F_k}\, (x_{\alpha_2^k} - x_{\alpha_3^k})$ are mutually independent. Thus

$$E[(x_{\alpha_1^l} + F(x_{\alpha_2^l} - x_{\alpha_3^l}))(x_{\alpha_1^k} + F(x_{\alpha_2^k} - x_{\alpha_3^k}))] = E(x_{\alpha_1^l}) \cdot E(x_{\alpha_1^k}) = \overline{x}^2.$$

Hence we obtained that $E(\overline{Y}^2) = \frac{1}{m} E(\overline{Y^2}) + \frac{m-1}{m} \overline{x}^2$. Thus

$$E(\mathrm{Var}(Y)) = E(\overline{Y^2}) - E(\overline{Y}^2) = \left(2F^2 + \frac{m-1}{m}\right) \mathrm{Var}(x). \tag{5}$$

(b) Each element of the population Z is a random variable with the structure

$$Z_l = \begin{cases} Y_l & \text{with probability } p \\ x_l & \text{with probability } 1-p \end{cases} \qquad l = \overline{1, m}.$$

To determine $E(\mathrm{Var}(Z))$ we have to compute $E(\overline{Z^2})$ and $E(\overline{Z}^2)$. Using properties of the mean of a random variables one obtains

$$E(\overline{Z^2}) = \frac{1}{m} \sum_{l=1}^{m} E(Z_l^2) = \frac{1}{m} \sum_{l=1}^{m} ((1-p)x_l^2 + pE(Y_l^2)) = (1-p)\overline{x^2} + pE(\overline{Y^2}).$$

On the other hand,

$$E(\overline{Z}^2) = \frac{1}{m^2} E\left[\left(\sum_{l=1}^{m} Z_l\right)^2\right] = \frac{1}{m^2}\left[E\left(\sum_{l=1}^{m} Z_l^2\right) + E\left(\sum_{k\neq l} Z_k Z_l\right)\right] = \frac{1}{m}E(\overline{Z^2}) + \frac{1}{m^2}\sum_{k\neq l} E(Z_k)E(Z_l) \quad (6)$$

since Z_k and Z_l are mutually independent. Using again properties of the mean the second term in the last relation can be expanded as follows:

$$\sum_{k\neq l} E(Z_k)E(Z_l) = \sum_{k\neq l}((1-p)x_k + pE(Y_k))((1-p)x_l + pE(Y_l)).$$

But $E(Y_l) = E[x_{\alpha_1^l} + \widetilde{F}_l \cdot (x_{\alpha_2^l} - x_{\alpha_3^l})] = E(x_{\alpha_1^l}) = \overline{x}$. Thus

$$\sum_{k\neq l} E(Z_k)E(Z_l) = \left[\sum_{l=1}^{m}((1-p)x_l + p\overline{x})\right]^2 - \sum_{l=1}^{m}((1-p)x_l + p\overline{x})^2 =$$

$$= (m(1-p)\overline{x} + mp\overline{x})^2 - \sum_{l=1}^{m}\left[(1-p)^2x_l^2 + 2p(1-p)\overline{x}x_l + p^2\overline{x}^2\right] = (m^2 - 2mp + mp^2)\overline{x}^2 - m(1-p)^2\overline{x^2}.$$

Replacing the last relation in (6) one obtains

$$E(\overline{Z}^2) = \frac{1}{m}E(\overline{Z^2}) + \frac{1}{m^2}(m^2 - 2mp + mp^2)\overline{x}^2 - \frac{1}{m}(1-p)^2\overline{x^2}.$$

Thus

$$E(Var(Z)) = \frac{m-1}{m}\left((1-p)\overline{x^2} + p\left(1 + 2F^2\frac{m}{m-1}\right)\overline{x^2} - 2pF^2\frac{m}{m-1}\overline{x}^2\right) - \left(1 - \frac{2p}{m} + \frac{p^2}{m}\right)\overline{x}^2 + \frac{(1-p)^2}{m}\overline{x^2} =$$

$$= \left(2F^2p - \frac{2p}{m} + \frac{p^2}{m} + 1\right)\mathrm{Var}(x). \quad (7)$$

and the theorem is proved. □

From the above result one can see that if the factor $2F^2p - 2p/m + p^2/m + 1$ is greater than one then the variation operators induce an increase of the population variance while if the factor is less than one the population diversity decreases after mutation and crossover. Since the numerical experiments suggest that the selection usually decreases the variance it follows that to prevent a too fast decrease of population diversity it would be indicate to choose the parameters which assure that the factor is greater than one. Thus, the values of the parameters which satisfy $2F^2 - 2/m + p/m = 0$ can be considered to be critical.

A similar theoretical result concerning the influence of selection on the expected population variance is difficult to obtain because the selection step depends on the objective function. Instead we will make an empirical study for a set of test functions.

4. Empirical analysis of the expected population variance.

The set of test functions. We selected the following test functions having all the properties: $n = 30$, $x_* = (0,\ldots,0)$, $f(x^*) = 0$.

1. (sphere model): $f_1(x_1,\ldots,x_n) = \sum_{i=1}^{n} x_i^2$, $x_i \in [100, 100]$;
2. (Griewank's function):$f_2(x_1,\ldots,x_n) = 1 + \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n}\frac{x_i}{\sqrt{i}}$, $x_i \in [-600, 600]$;
3. (Rastrigin's function):$f_3(x_1,\ldots,x_n) = \sum_{i=1}^{n}(x_i^2 - 10 * \cos(2\pi x_i) + 10)$, $x_i \in [-5.12, 5.12]$.

Details of the numerical experiments. We analyzed the influence of F on the algorithm behavior when m and p are fixed ($m = 50$, $p = 0.2$). The statistical analysis is based on collecting results from 50 independent runs. The empirical expected variance ($\langle \mathrm{Var}(\cdot)\rangle$) and/or standard deviation ($\langle \mathrm{Sd}(\cdot)\rangle$)are computed by averaging the variance/standard deviation over 50 runs and over all $n = 30$ components.

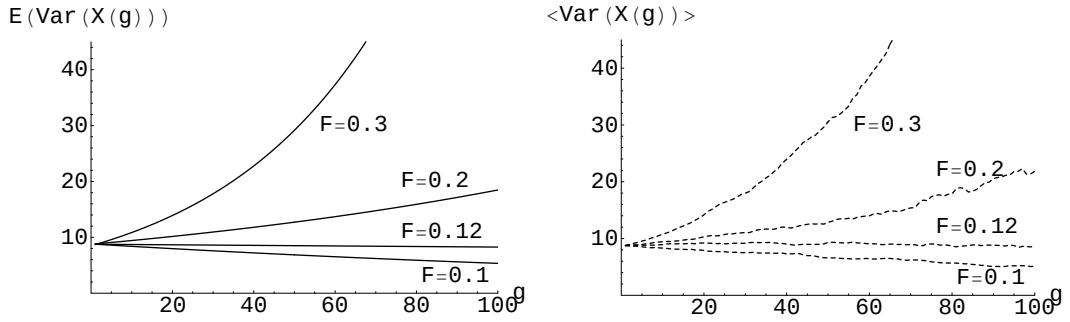
Empirical remarks. First, the experimental results confirms the theoretical result obtained in the previous section (see Fig. 2). The influence of selection on the expected population variance is illustrated in Fig.3. From Fig.3.a we can infer that the one-step dependence between the variance of population after selection and the variance of population after mutation and crossover is also linear. Since the standard deviation can be viewed as a measure of the DE perturbation, the results which we present in the following refer to the average standard deviation instead of averaged variance. So we suppose that $\langle \mathrm{Sd}(W)\rangle \simeq a \cdot \langle \mathrm{Sd}(Z)\rangle$ and, using (3) one infer that $\langle \mathrm{Sd}(W)\rangle \simeq a \cdot c \cdot \mathrm{Sd}(x)$ with $c = \sqrt{2F^2p - 2p/m + p^2/m + 1}$. We determined empirical values for a and c by

linear regression using values of the standard deviations collected during the first 100 generations. The obtained results are presented in Table 1.

The difference between the theoretical and empirical values of c can be explained as follows: (i) the theoretical result establishes the one-step evolution of the expected variance while the empirical results refer to the one-step evolution of the standard deviation (the expectation of the standard deviation is not the square root of the variance expectation); (ii) the theoretical analysis does not take into consideration the restriction that the offsprings are accepted only if they belong to D, while the implementation does this validation.

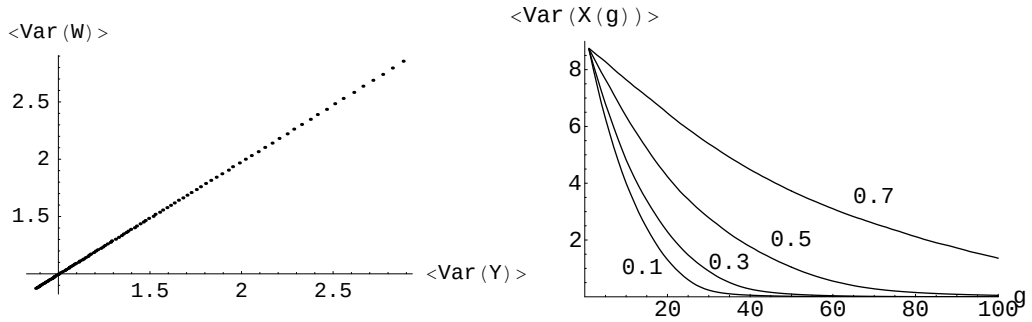
The influence of parameter F on the convergence properties of the DE algorithm is illustrated in Table 2 (the classical variant) and 3 (the variant with (1)). As stopping criterion we used: "$f_* < 10^{-3}$ or $\langle \mathrm{Sd}(X) \rangle < 10^{-6}$" ($f_*$ is the best value of the objective function found into the population). The algorithm is successful if the first part of the condition is true (in this case $\langle g^* \rangle$ denotes the averaged number of generations until the optimum is reached). If the second part of the stop condition is satisfied it is considered that the algorithm converged prematurely (the occurrence of such a situation in 50 runs is marked with PC in Tables 2 and 3).

From Tables 2 and 3 one can find that $F = 0.3$ is an empirical critical value while the theoretical critical value for $m = 50$ and $p = 0.2$ (the value which verifies $2F^2 - 2/m + p/m = 0$)is $F = 0.1341$. On the other hand, comparing the results from Tables 2 and 3 one can conclude that the convergence behavior of the two analyzed variants is almost similar.



(a) (b)

Figure 2: Evolution of the expected population variance after mutation and crossover (without selection) ((a) theoretical dependence; (b) empirical results for Rastrigin function)



(a) (b)

Figure 3: (a) Empirical dependence between the expected population variance before and after selection; (b) Evolution of expected population variance after mutation, crossover and selection for different values of F

5. Conclusions. Premature convergence can pe prevented if mutation and crossover induce an increase of the population variance. To find appropriate values for the control parameters, a simple relation ($2F^2 - 2/m + p/m = 0$) can be used. Numerical experiments for different test functions suggest that a good behavior is obtained if the factor c is greater than 1. The factor c can be computed either analytically (using (3)) or numerically by collecting information on the population standard deviation during a few generations. However, values too great for c could slow down the convergence process. The values of c computed using the analytical relation in the case of the numerical results reported in [7] are: 1.0483, 1.3332, 1.1545, 1.2196, 1.2049, 1.0139, 1.2581, 1.251, 1.4988, 1.354. As we can see all are in $(1, 1.5)$. Thus we can infer that a good convergence behavior can be obtained if c is a little greater than 1.

To conclude, we can state that similar evolution of population variance implies similar convergence properties. For instance, the empirical values for a in the case of the sphere model are very close to the values obtained for the Griewank function (see Table 1, columns four and seven) and so are their convergence properties. On the other hand the values of a obtained for the Rastrigin function are a little different from those obtained for the sphere model and the convergence behavior is also a little different.

		Sphere			Griewank			Rastrigin		
F	Theor. c	Empir. c	Empir. a	Empir. $a \cdot c$	Empir. c	Empir. a	Empir. $a \cdot c$	Empir. c	Empir. a	Empir. $a \cdot c$
0.1	0.9983	0.9932	0.9764	0.9698	0.9935	0.9760	0.9696	0.9934	0.9914	0.9849
0.2	1.0043	0.9956	0.9748	0.9705	0.9958	0.9747	0.9706	0.9970	0.9900	0.9871
0.3	1.0453	1.0009	0.9714	0.9722	1.0013	0.9711	0.9724	1.0030	0.9852	0.9882
0.5	1.1175	1.0204	0.9577	0.9772	1.0201	0.9579	0.9772	1.0204	0.9705	0.9903

Table 1: Theoretical values for c and empirical values for a and c.

	Sphere				Griewank				Rastrigin			
F	Succ.	$\langle g^* \rangle$	PC	$\langle Sd \rangle$	Succ.	$\langle g^* \rangle$	PC	$\langle Sd \rangle$	Succ.	$\langle g^* \rangle$	PC	$\langle Sd \rangle$
0.1	0	–	50	0	0	–	50	0	0	–	50	0
0.15	17	293	33	0.0019	21	323	29	0.0224	2	618	48	$8 \cdot 10^{-6}$
0.2	44	292	6	0.0061	47	321	3	0.0515	20	990	30	0.0001
0.3	50	327	0	0.0071	50	360	0	0.0433	45	2128	5	0.0004
0.5	50	477	0	0.0072	50	542	0	0.0453	50	3383	0	0.0005

Table 2: Convergence behavior of the algorithm with $\widetilde{F_l^i} = F$

	Sphere				Griewank				Rastrigin			
F	Succ.	$\langle g^* \rangle$	PC	$\langle Sd \rangle$	Succ.	$\langle g^* \rangle$	PC	$\langle Sd \rangle$	Succ.	$\langle g^* \rangle$	PC	$\langle Sd \rangle$
0.1	0	–	50	0	1	3879	49	0	0	–	50	0
0.15	10	307	40	0.0009	12	317	38	0.0100	1	483	49	$8 \cdot 10^{-6}$
0.2	39	296	11	0.0051	43	323	7	0.0424	24	556	26	0.0002
0.3	50	321	0	0.0071	50	354	0	0.0511	44	797	6	0.0004
0.5	50	425	0	0.0072	50	473	0	0.0457	50	1432	0	0.0005

Table 3: Convergence behavior of the algorithm with $\widetilde{F_l^i}$ given by (1)

References

[1] Abbass H.A., R. Sarker and C. Newton (2001). A Pareto Differential Evolution Approach to the Vector Optimisation Problems, *The Congress on Evolutionary Computation, CEC2001.*

[2] Beyer,H.G.(1999). On the Dynamics of EAs without selection, in W. Banzaf, C. Reeves (eds.), *Foundations of Genetic Algorithms*, Morgan Kaufmann, pp.5-26.

[3] Lampinen,J.(1999). A Bibliography of Differential Evolution Algorithm, *Technical Report*, Lappeenranta University of Technology, Department of Information Technology, Laboratory of Information Processing.

[4] Lampinen J.(1999), Differential Evolution-New Naturally Parallel Approach for Engineering Design Optimization, in Barry H.V. Topping(ed.), *Developments in Computational Mechanics with High Performance Computing*, Civil-Comp Press, Edinburgh, pp.217-228.

[5] Lampinen,J. and I. Zelinka (2000). On Stagnation of the Differential Evolution Algorithm, in Ošmera P. (ed), *Proc. of MENDEL 2000, 6th International Mendel Conference on Soft Computing*, pp. 76-83.

[6] Rudolph G.(1996) Convergence of Evolutionary Algorithms in General Search Spaces, *Proc. of the third IEEE Conf. on Evolutionary Computation*, IEEE Press, pp. 50-54.

[7] Storn,R. and K. Price(1995). Differential Evolution - A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces, *Technical Report TR-95-012*, ICSI, March, 1995.