Parameter adaption in differential evolution by controlling the population diversity

Article · January 2002

CITATIONS
50

READS
192

1 author:

Daniela Zaharie
West University of Timisoara
108 PUBLICATIONS **1,265** CITATIONS

SEE PROFILE

Analele Universităţii din Timişoara
Vol. XXXX, Fasc. 2, 2002
Seria Matematică–Informatică

Parameter Adaptation in Differential Evolution by Controlling the Population Diversity

Presented at 4rd Int. Workshop on Symbolic and Numeric Algorithms for Scientific Computing
9-12 October 2002, Timisoara, Romania

Daniela ZAHARIE

Abstract. In this paper an adaptive parameter control in differential evolution algorithms is proposed. The adaptation strategy is based on theoretical and empirical results concerning the population diversity evolution. The main idea is to adaptively adjust the parameters such that a fast decrease of the population variance is avoided.

Key words: differential evolution, global optimization, population diversity, adaptive parameter control

AMS Subject Classification: *(2000)*:90C56, 68W20

1. Introduction. The Differential Evolution algorithm (DE) introduced by Storn and Price in (Storn, 1995) is a stochastic population-based search method successfully applied in global optimization problems. As in other evolution strategies, two main processes drive the evolution: the *variation* (perturbation) process which ensures the exploration of the search space and the *selection* process which ensures the exploitation abilities of the algorithm. Both the perturbation and the selection processes are simpler than those used in other evolution strategies. The perturbation of a population element is done by probabilistically replacing it with an offspring obtained by adding to a randomly selected element a perturbation proportional with the difference between other two randomly selected elements. The selection is also of a particular type being a one to one competition between the parent and its offspring.

There exist at least three types of convergence behavior in DE:

(i) *Good convergence:* the algorithm approaches the global optimum in a "reasonable" number of generations. In this case a good trade-off between the exploration and the exploitation processes exists.

(ii) *Premature convergence:* the algorithm is stuck into a non-optimal state, primarily due to the loss of population diversity. In this case the exploration of the search space is too weak.

(iii) *Slow convergence:* the algorithm does not approach the global optimum in a reasonable number of generations. The perturbation process is more "powerful" than the selection process.

An essential role in influencing the convergence behavior of a given DE is played by its strategy parameters (e.g. the population size, the magnitude of the perturbation term and the probability of applying the perturbation). As many experimental results indicate ((Storn, 1995), (Liu, 2002), (Wei, 2002)), the performance of DE is very sensitive to the choice of strategy parameters. The empirical results also suggest that the parameters values space could be decomposed in three "regions", each one for each convergence behavior type. If we would know such a decomposition we would choose adequate values for the parameters. Unfortunately such a decomposition is difficult to be obtained because: (i) it is highly dependent on the objective function; (ii) the borders of the regions could have intricate shapes or could be fuzzy; (iii) the decomposition could vary in time being dependent on the current population. Another difficulty is that for some objective functions the good behavior region in parameters space could be small, so difficult to be attained. Instead of trying to find parameters in the good behavior region we could try to avoid the regions with clear undesirable behavior (for example premature convergence). A partial result in this sense has been obtained in (Zaharie, 2002) where, based on a theoretical result on the population variance evolution after the perturbation step, we obtained a surface which split the 3-dimensional parameter space (involved parameters are: population size - m, the perturbation factor - F and the perturbation probability - p) into two regions: one for which the variation operators induce an increase of the population variance and another one where they induce a decrease of the population variance. Empirical results (Zaharie, 2002) suggest that the parameter region where the variation operators induces a decrease of the population variance is almost similar with the parameter region corresponding to premature convergence. Approximation of the other separation surface (that

between the good convergence region and the slow convergence region) is more difficult to be obtained.

It is unanimous accepted that maintaining adequate population diversity is an important issue in evolution strategies. The role of population diversity is even greater in DE due to the absence of an external infusion of diversity (the perturbation is internally induced). We believe that the evolution of the population diversity is highly responsible for the DE behavior even if it is obviously that it is not the only influential factor. Empirical results on a set of test functions (Zaharie, 2001), (Zaharie, 2002) suggest that:

(i) rapid decreasing of population diversity induces premature convergence due to the fact that the population loses too soon its variability and the internal perturbation loses its power;

(ii) too slow decreasing (or even increasing) of population diversity induces a slow convergence due to the spread of population on uninteresting regions of the search space.

There exist at least two strategies of keeping the population diversity at an acceptable level: (a) injecting external diversity into population by replacing some of the non-optimal elements of the population with randomly selected elements from the search space (Šmuc, 2002); (b) choosing the strategy parameters such that the population diversity does not decrease too fast (Zaharie, 2002).

The population diversity evolution depends not only on the strategy parameters but also on the objective function. The theoretical results obtained in (Zaharie, 2001) and (Zaharie, 2002) refer only to the variation operators (which do not depend on the objective function) thus they do not offer us sufficient information to assure a direct choice of the parameters, a parameter tuning being necessary, which is a time-consuming task. An alternative to parameter tuning is the parameter control (Eiben, 1999) which can be of one of the following types:

(i) *deterministic parameter control*: the parameters are modified during the evolution, using a rule which depend on the generation number but which does not use any feedback from the search;

(ii) *adaptive parameter control*: the changes on the parameters values are based on feedback information from the search process;

(iii) *self-adaptive parameter control*: the strategy parameters are encoded into the population elements and transformed via perturbation operators during the evolution. The selection process ensures that the most adequate parameters are kept for the next generation.

Recently, in DE field have been proposed techniques for controlling the strategy parameters. In (Liu, 2002) a fuzzy adaptive parameter control strategy is proposed. The membership functions and the fuzzy rules used in parameters modifications are established based on human knowledge and expertise gathered from previous experiments on benchmark test functions. A self-adaptive approach to DE is presented in (Abbas, 2002) in the context of multi-objective optimization problems.

In this paper an adaptive parameter control is proposed based on theoretical and empirical insights on the relationship between the parameters, the population diversity evolution and the convergence behavior of the algorithm. The problems which we tackle in this paper are: (i) how influence the parameters of a general DE the population diversity evolution; (ii) how could we adapt the parameters to obtain good convergence properties. The first question has been partially answered in (Zaharie, 2001) and (Zaharie, 2002). Here we extend the previous results for a general DE algorithm which include the particular cases already analyzed. Using theoretical and empirical insights we develop an adaptive parameter control which keeps the parameters values in the good convergence region.

2. The general structure of the differential algorithm We consider the problem of finding the minimum $x^* \in D$ of an objective function $f : D \subset R^n \to R$ on which we do not impose any restriction. The DE algorithm which we shall analyze is a general one (see Figure 1) which incorporates as particular cases the DE1 and DE2 schemes introduced in (Storn, 1995). The algorithm evolves a fixed size population, $P = \{x_1, \ldots, x_m\}$ which is randomly initialized with elements from D. After population initialization an iterative process is started and at each iteration (generation) a new population is produced until a stopping condition is satisfied.

At each generation, each population element (x_l) could be replaced (with probability p) with a newly generated element. The new element is a linear combination between the best element of the population ($x_*(g)$), a randomly selected element (x_{α^l}) and Q differences between pairs of randomly selected elements. For each l, the indices α^l, β_q^l and γ_q^l (for $q = \overline{1,Q}$) are randomly selected without replacement from $\{1, \ldots, m\}$. Besides the

---

`Initialize` $P(0) = \{x_1(0), \ldots, x_m(0)\}$, $g = 0$
`Repeat`
Mutation step:
$$y_l^i = \lambda x_*^i(g) + (1-\lambda)x_{\alpha^l}^i(g) + \sum_{q=1}^{Q} F_q \cdot (x_{\beta_q^l}^i(g) - x_{\gamma_q^l}^i(g)),$$
$$l = \overline{1,m}, \quad i = \overline{1,n}$$
Crossover step:
$$z_l^i = \begin{cases} y_l^i & \text{with probability } p \\ x_l^i(g) & \text{with probability } 1-p \end{cases}$$
$$l = \overline{1,m}, \quad i = \overline{1,n}$$
Selection step:
`If` $f(z_l) < f(x_l(g))$ `then` $x_l(g+1) = z_l$ `else` $x_l(g+1) = x_l(g)$
$$l = \overline{1,m}$$
$g = g + 1$ (increment the generation counter)
`Until a stopping criterion is satisfied`.

Fig. 1. The general DE algorithm

population size (m), the parameters of the algorithm are: $\lambda \in [0,1]$ (the coefficient of the convex combination between the best element of the population and a randomly selected element); $p \in [0,1]$ (the probability of replacing an element with the newly generated one); $\{F_q\}_{q=\overline{1,Q}}$ (the factors which amplify the "differential" terms) and Q (the number of "differential" terms).

Some frequently used particular variants of the general algorithm are:

- DE/rand/1/bin (DE1 in (Storn, 1995)): $\lambda = 0$ and $Q = 1$;
- DE/best/1/bin (DE2 in (Storn, 1995)): $\lambda \in (0,1]$ and $Q = 1$. The original DE2 scheme is characterized by the fact that $\alpha_1^l = l$;
- DE/best/2/bin ((Gämperle, 2002)): $\lambda = 1$ and $Q = 2$, $F_1 = F_2 = F$.

We remark here that, due to the particular form of the DE selection, the variant with $\lambda > 0$ can be viewed as a combination between DE and particle swarm algorithms without inertial term (for a similar hybridization between fast evolutionary programming (Yao, 1999) and particle swarm algorithms see (Wei, 2002)).

3. Population variance evolution As we noted in the first section, the population diversity is an important element in assuring a good convergence behavior. Different measures of the population diversity can be used: computed either on the search space or on the objective function values space. Due to its good theoretical properties we choose as diversity measure the variance computed at the component level of the population elements. Since in DE algorithm the perturbations are made independently for each component it will not be a loss of generality if we will conduct our computations at components level. To do this we will consider a population of scalars $x = \{x_1, \ldots, x_m\}$ with $x_l \in R$, $l = \overline{1,m}$ and we shall measure its diversity by $Var(x) = \overline{x^2} - \overline{x}^2$ with $\overline{x} = \sum_{l=1}^{m} x_l/m$.

To analyze the population variance evolution we introduce the following notations: x denotes the current population, Y denotes the population after mutation, Z the population after crossover and W the population after that selection has been applied. Since random elements are involved in variation steps Y, Z and W are populations of random variables. Thus for these populations we can only compute the expected value of variance either theoretical (the theoretical mean of a random variable is denoted by $E(\cdot)$) or empirical (the empirical mean of a random variable is denoted by $\langle\cdot\rangle$). The empirical expected value of the population variance is computed by averaging the variance over all components.

In the following we shall theoretically analyze the influence of the *variation operators* and of their parameters on the expected population variance. The theoretical result extends those presented in (Zaharie, 2001) and (Zaharie, 2002) being valid for the general DE (Figure 1) not only for particular cases.

Before stating the main theoretical result of this paper we present some useful preliminary remarks. If α_i and α_j are two values randomly selected without replacement from $\{1, \ldots, m\}$ then the following relations hold:

$$E(x_{\alpha_i}) = \overline{x} \qquad E(x_{\alpha_i}^2) = \overline{x^2} \qquad E(x_{\alpha_i} x_{\alpha_j}) = (m\overline{x}^2 - \overline{x^2})/(m-1). \quad (1)$$

Theorem 1. *The expected variance of population obtained after mutation and crossing over is:*

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

$$\frac{m-1}{m}p\lambda^2(1-p)(\overline{x}-x_*)^2 \tag{2}$$

where $F^2=\sum_{q=1}^{Q}F_q^2$.

Proof. First we shall analyze the influence of the mutation step:

$$E(Var(Y))=E(\overline{Y^2})-E(\overline{Y}^2).$$

The first term of this difference is:

$$E(\overline{Y^2})=\frac{1}{m}\sum_{l=1}^{m}E(Y_l^2)=\frac{1}{m}\sum_{l=1}^{m}E[(\lambda x_*+(1-\lambda)x_{\alpha^l}+\sum_{q=1}^{Q}F_q(x_{\beta_q^l}-x_{\gamma_q^l}))^2]=$$

$$=\frac{1}{m}\sum_{l=1}^{m}\Big[\lambda^2x_*^2+2\lambda(1-\lambda)x_*\overline{x}+(1-\lambda)^2\overline{x^2}+$$

$$+2\sum_{q=1}^{Q}F_q^2\left(\overline{x^2}-\frac{m\overline{x}^2-\overline{x^2}}{m-1}\right)\Big]=$$

$$=\lambda^2x_*^2+2\lambda(1-\lambda)x_*\overline{x}+(1-\lambda)^2\overline{x^2}+2F^2\frac{m}{m-1}Var(x).$$

After some simple algebraic computations one obtains:

$$E(\overline{Y}^2)=\frac{1}{m}+\frac{m-1}{m}(\lambda x_*+(1-\lambda)\overline{x})^2.$$

Thus

$$E(Var(Y))=\left(2F^2+\frac{m-1}{m}(1-\lambda)^2\right)Var(x). \tag{3}$$

In a similar manner we compute the expected variance of population Z:

$$E(Var(Z))=\frac{m-1}{m}E(\overline{Z^2})-\frac{1}{m^2}\sum_{k\neq l}E(Z_k)E(Z_l).$$

The first term of the above difference is:

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

$$=\frac{m-1}{m}(1-p)\overline{x^2}+p\frac{m-1}{m}(\lambda^2x_*^2+2\lambda(1-\lambda)x_*\overline{x}+(1-\lambda)^2\overline{x^2})+2pF^2. \tag{4}$$

After some computations, the second term becomes:

$$\frac{1}{m^2}\sum_{k\neq l} E(Z_k)E(Z_l) =$$

$$= \frac{1}{m^2}\sum_{k\neq l}((1-p)x_l + p\lambda x_* + p(1-\lambda)\overline{x})((1-p)x_k + p\lambda x_* + p(1-\lambda)\overline{x}) =$$

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

$$+\left((1-p)^2 + \frac{2p(1-p)(1-\lambda)(m-1)}{m} + \frac{p^2(1-\lambda)^2(m-1)}{m}\right)\overline{x}^2 +$$

$$+\frac{2(m-1)}{m}\lambda p(1-p+p(1-\lambda))x_*\overline{x}. \tag{5}$$

Computing the difference between (4) and (5) one obtains (2) and the theorem is proved.

Particular cases.
Case 1: $\lambda = 0$. In this case the best element of the population is not taken into consideration and (2) becomes

$$E(Var(Z)) = \left(2pF^2 + 1 - \frac{2p}{m} + \frac{p^2}{m}\right)Var(x). \tag{6}$$

If $Q = 1$ this result is similar with that presented in (Zaharie, 2001). We remark that from the viewpoint of variance evolution using $Q > 1$ is equivalent with increasing the value of F. However the variance evolution is not the unique factor which influences the DE convergence properties, thus using $Q > 1$ could be beneficial in some situations. Such an analysis is out of the scope of this paper.
Case 2: $p = 1$. In this case all the population elements are replaced (in the variation step) with the new generated elements and (2) becomes:

$$E(Var(Z)) = \left(2F^2 + \frac{m-1}{m}(1-\lambda)^2\right)Var(x). \tag{7}$$

In both particular cases presented above a linear dependence exists between $E(Var(Z))$ and $Var(x)$. In the general case this would be true

only if we could accept the approximation $(\overline{x} - x_*)^2 \simeq K \cdot Var(x)$. To simplify the analysis we shall accept this approximation and we will introduce the notation:

$$c = 2pF^2 + \frac{(1-p)^2}{m} + \frac{m-1}{m}(p(1-\lambda)^2 + (1-p)) + K\frac{m-1}{m}p\lambda^2(1-p) \quad (8)$$

such that $E(Var(Z)) \simeq cVar(x)$. If $c < 1$ then the variation operators induce a decrease of the population variance. This could induce after selection a too fast decrease of the global population variance ($E(Var(W))$). Thus the parameter space region characterized by $c < 1$ can be considered as included in the premature convergence region. The surface described by $c = 1$ can be considered as a border between the premature and good convergence regions. On the other hand if c is too high then it induces a too slow decrease (or even an increase) of the global population variance which could imply slow convergence. To estimate the upper bound of c (the value which we suppose that determines the separation between the good and slow convergence regions) we analyzed some reported experimental results (see Table 1) and also we made our set of experiments (see Table 2). Our testbed (see Table 3) is a subset of that used in (Yao, 1999). The parameters used in tests whose results are presented in Table 2 correspond to the following values of c: $c = 0.968$ (for $F = 0.1$), $c = 1.541$ (for $F = 1.2$) and $c = 2.56$ (for $F = 2$). We have deliberately chosen a small population ($m = 10$) to emphasize that for some objective functions (e.g. Rastrigin) if the population is not large enough then the three regions can not be easily delimitated. The numerical results suggest that an empirical upper bound for c could be near 2 (depending on the objective function) but a rigorous one seems to be difficult to obtain. Therefore in the following we will search for parameters values near the surface corresponding to the lower bound ($c = 1$) even if this strategy does not induce an optimal behavior of the algorithm but only a satisfactory one.

To gather more information on the role of parameters in the convergence process we need to know how they influence the global population variance evolution (when selection is also considered). We tried to obtain such information in an empirical way analyzing the relationship between $\langle Var(W) \rangle$ and $Var(x)$. Using values of the expected population variance collected during the first 100 generations we obtained results as those illustrated in Figure 2.

Parameters			c	Reported
m	F	p		behavior
10	0.5	0.3	1.099	GC (Storn, 1995)
6	0.95	0.5	1.777	GC (Storn, 1995)
10	0.8	0.3	1.333	GC (Storn, 1995)
10	0.75	0.5	1.487	GC (Storn, 1995)
30	1	0.3	1.583	GC (Storn, 1995)
30	0.8	1	2.246	GC (Storn, 1995)
100	0.65	1	1.835	GC (Storn, 1995)
20	0.1	0.9	0.968	PC (Liu, 2002)
20	0.9	0.9	2.408	GC (Liu, 2002)
20	1.5	0.9	5	SC (Liu, 2002)

Table 1. Values of c for some results presented in the literature in the case $\lambda = 0$, $Q = 1$ (GC-good convergence, PC- premature convergence, SC-slow convergence)

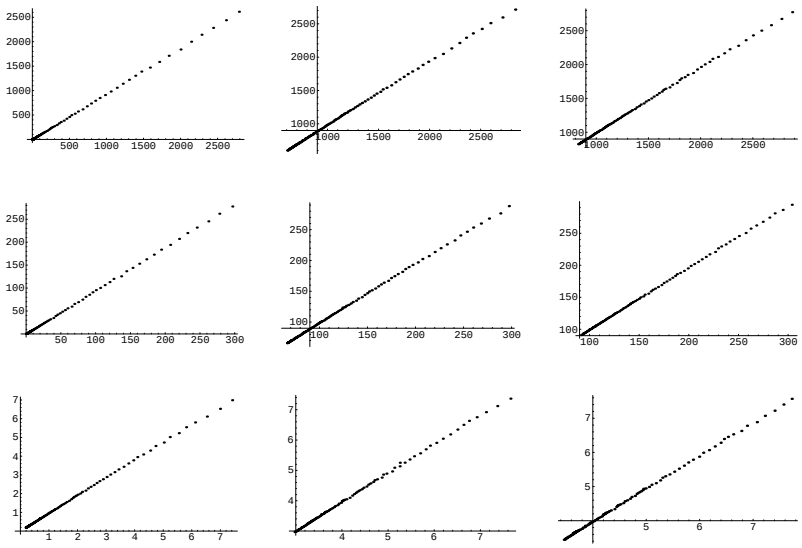
F	Sphere	Ackley	Rastrigin
0.1	PC	PC	PC (quick)
1.2	GC	GC	PC (slow)
2	SC/PC	SC/PC	PC (very slow)

Table 2. Behavior of DE for $\lambda = 0$, $p = 0.2$, $m = 10$, $Q = 1$ (GC-good convergence, PC- premature convergence, SC-slow convergence)

Name	Expression	Domain
Sphere	$f_S(x) = \sum_{i=1}^{n} x_i^2$	$[-100, 100]^n$
Rosenbrock	$f_{Ro}(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	$[-30, 30]^n$
Rastrigin	$f_R(x) = \sum_{i=1}^{n} [x_i^2 - 10\cos(2\pi x_i) + 10]$	$[-5.12, 5.12]^n$
Ackley	$f_A(x) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) -$ $\exp(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)) + 20 + e$	$[-32, 32]^n$
Griewangk	$f_G(x) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n}\cos(x_i/\sqrt{i}) + 1$	$[-600, 600]^n$

Table 3. Test functions

From these illustrations we can infer that there exists a linear dependence between $\langle Var(W) \rangle$ and $Var(x)$. Since $Var(x) = 0$ obviously implies



(a) (b) (c)

Fig. 2. Dependence between the expected population variances after and before transformation (perturbation and selection). Test functions: sphere (top), Ackley (middle), Rastrigin (bottom). Parameters: $p = 0.2$, $m = 50$ (a) $F = 0.1$ (b) $F = 1.2$ (c) $F = 2$

$Var(W) = 0$ it follows that:

$$\langle Var(W) \rangle = aVar(x) \tag{9}$$

Some estimations of the coefficient a obtained by linear regression from the data collected during 100 generations are presented in Table 4. All these results suggest a decrease of the global population variance.

F	Sphere	Ackley	Rastrigin
0.1	0.93248	0.94078	0.95105
1.2	0.97596	0.97836	0.98628
2	0.98149	0.98389	0.98908

Table 4. Empirical values of the linear dependence coefficient (a) between $\langle Var(W) \rangle$ and $Var(x)$

4. The adaptation scheme The starting idea of the adaptive parameter control which we propose is to combine theoretical results (i.e. relation (2)) with empirical knowledge and expertise collected through numerical experiments. Since the parameters influence in an interacting manner the convergence properties of DE it is difficult to automatically adapt all the parameters. Thus our first approach is to adapt only one parameter while the other ones are fixed. We consider m, λ and p fixed while F will be the variable parameter. Also we will use $Q = 1$. The basic idea of adapting F is to choose it such that it "corrects" the population variance in order to avoid the undesirable regions.

At each iteration we will compute for each component a new value of the parameter F_j as follows:

$$F_j(g+1) = \varphi\left(\frac{Var(x_j(g))}{Var(x_j(g+1))}, m, p, \lambda, K\right) \tag{10}$$

where $\varphi(c, m, p, \lambda, K)$ is the positive solution of the equation in F:

$$2pF^2 + \frac{(1-p)^2}{m} + \frac{m-1}{m}(p(1-\lambda)^2 + (1-p)) + K\frac{m-1}{m}p\lambda^2(1-p) = c \tag{11}$$

Since the factor $Var(x_j(g))/Var(x_j(g+1))$ can be viewed as an approximation of $1/a$, the values $F_j(g+1)$ chosen as in (10) will have the following effect: (i) if the population variance tends to decrease ($a < 1$) then F_j is modified such that it induce an increase of the variance through mutation and crossover; (ii) in the case when the population variance tends to increase ($a > 1$) then F_j is modified such that it induce a decrease of the variance.

Experimental results for the test functions listed in Table 3 are presented in Table 5. The values of the fixed parameters are: $m = 50$, $K = 1$. The stopping condition which we used is: "$f_* < 10^{-3}$ or $\langle Var(x)\rangle < 10^{-12}$ or the generation number is greater than a maximal value" (f_* is the best value of the objective function found into the population). The algorithm is considered successful if the first part of the condition is satisfied and it prematurely converges if only the second part is true. If the third part of the condition is true then it is considered that a slow convergence situation occurs. The maximal number of generation was 5000 for f_S, f_R, f_A and f_G while for f_{Ro} it was set to 10000.

The columns in Table 5 have the following significance: λ - the parameter used in swarm-type term; S - number of successes in 50 runs; $\langle g_S\rangle$ -

averaged number of generations until the optimum is approached; f_* - the best function value found during all 50 runs; $\langle f \rangle$ - the averaged function value; PC - number of cases of premature convergence; $\langle g_P \rangle$ - averaged number of generations until premature convergence occurred and the optimum has not been attained; $\langle Var(x) \rangle$ - averaged (over all the components and all 50 runs) value of the population variance in the last generation; $\langle F \rangle$ - averaged value of the parameter F. All averaged values are computed based on 50 independent runs of the algorithm.

Fct.	λ	S	$\langle g_S \rangle$	f_*	$\langle f \rangle$	PC	$\langle g_P \rangle$	$\langle Var(x) \rangle$	$\langle F \rangle$
f_S	0	50	283	0.0007	0.0009	–	–	0.00004	0.2583
	1	50	211	0.0006	0.0008	–	–	0.00003	0.5677
f_{Ro}	0	–	–	9.6188	65.76	9	2511	0	0.1994
	1	38	4297	0.0009	0.8426	7	5117	0.00707	0.5234
f_R	0	1	584	0.0009	2.905	49	844	0	0.255
	1	–	–	11.93	26.38	50	659	0	0.5772
f_A	0	50	346	0.0007	0.0009	–	–	0	0.251
	1	49	267	0.0007	0.024	1	395	0	0.5747
f_G	0	49	300	0.0006	0.001	1	634	0.00146	0.2569
	1	28	220	0.0007	0.0055	22	465	0.0006	0.5882

Table 5. Results of DE with adaptive F_j ($p = 0.5$)

The adaptive (with respect to F) variant of the DE algorithm is presented in Figure 3. To simplify the relations we fixed $\lambda = 0$ and $Q = 1$ (the adaptive variant of DE1 scheme). The computed values of F_j are accepted only if they are between the limits: $F_{inf} = \sqrt{1/m}$ and $F_{sup} = 2$. To obtain the adaptive algorithm for $\lambda = 1$ (and using $K = 1$) we have to make the following modifications: $t_j = (1 - p_j)^2/m + (m - 1)(1 - p_j^2)/m$ the other relations remaining identical.

The adaptation process can be extended by alternatively applying adaptation on F_j and on p_j (for instance, at odd generation parameters F_j are adjusted, while at even generations are adjusted the parameters p_j). The adaptation rule for p_j is obtained by solving equation (11) with respect to p.

5. Conclusions The main particularities of the adaptive DE algorithm are:

- The parameters F and p becomes individual, each component of the population elements having its own parameters values.

`Initialize` $P(0) = \{x_1(0), \ldots, x_m(0)\}$, $g = 0$
`Compute` $F_j = \sqrt{1/m}$, $c_j = 1$, $Var(x_j(g))$, $j = \overline{1,n}$
`Repeat`
Mutation step:
`For` $l = \overline{1,m}$ and $j = \overline{1,n}$
$y_l^j = \lambda x_*^j(g) + (1-\lambda) x_{\alpha^l}^j(g) + F_j \cdot (x_{\beta^l}^j(g) - x_{\gamma^l}^j(g))$
Crossover step:
`For` $l = \overline{1,m}$ and $j = \overline{1,n}$

$$z_l^j = \begin{cases} y_l^j & \text{with probability } p \\ x_l^j(g) & \text{with probability } 1-p \end{cases}$$

Selection step:
`For` $l = \overline{1,m}$
`If` $f(z_l) < f(x_l(g))$ `then` $x_l(g+1) = z_l$ `else` $x_l(g+1) = x_l(g)$
Compute the population variance:
`Compute` $Var(x_j(g+1))$ and
$c_j = Var(x_j(g))/Var(x_j(g+1)), \quad j = \overline{1,n}$
Compute values for individual F parameters:
`For` $j = \overline{1,n}$
$t_j = (1-p_j)^2/m + (m-1)/m$
`If` $c_j < t_j$ `then` $F_j = F_{inf}$ `else` $F_j = \sqrt{(c_j - t_j)/(2p)}$
`If` $F_j < F_{inf}$ `then` $F_j = F_{inf}$
`If` $F_j > F_{sup}$ `then` $F_j = F_{sup}$
$g = g + 1$ (increment the generation counter)
`Until a stopping criterion is satisfied`.

Fig. 3. The adaptive variant of the DE1 scheme

- The adaptation relations are based on the theoretical results concerning the population variance evolution. Relations like (2), (6) or (7) allows us to see how the parameter interact and how contribute each of them to the population variance evolution. At each generation the parameters are adjusted such that they are in the good convergence region, near the border which separates this region from the premature convergence region. Due to this option the adaptation rules does not assure the best convergence speed its

main aim being to avoid premature convergence.

- Since the convergence properties of DE are not completely determined by the population variance evolution introducing into the adaptation process other information regarding the algorithm behavior could be beneficial.
- Compared with the results reported for fast evolutionary programming (Yao, 1999), the results presented in Table 5 suggest that the adaptive DE is competitive.

REFERENCES

H.A. Abbass; The Self-Adaptive Pareto Differential Evolution Algorithm, *Proceedings of 2002 Congress on Evolutionary Computation*, pp. 831-836, 2002.

A.E. Eiben, R. Hinterding; Parameter Control in Evolutionary Algorithms; *IEEE Trans. on Evolutionary Computation*, vol. 3, no. 2, pp. 124-141, 1999.

R. Gämperle, S.D. Müller, P. Koumoutsakos; A Parameter Study for Differential Evolution, in A. Grmela, N.E. Mastorakis (eds.) *Advances in Intelligent Systems, Fuzzy Systems, Evolutionary Computation*, WSEAS Press, pp. 293-298, 2002.

J. Liu, J. Lampinen; Adaptive Parameter Control of Differential Evolution, in R. Matoušek, P. Ošmera (eds.), *Proc. of Mendel 2002, 8th International Conference on Soft Computing*, pp. 19-26, 2002.

T. Šmuc; Improving Convergence Properties of the Differential Evolution Algorithm, in R. Matoušek, P. Ošmera (eds.), *Proc. of Mendel 2002, 8th International Conference on Soft Computing*, pp. 80-86, 2002.

R. Storn, K. Price; Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, *Technical Report TR-95-012*, ICSI, 1995.

C. Wei, Z. He, Y. Zhang, W. Pei; Swarm Directions Emedded in Fast Evolutionary Programming, Proceedings of 2002 Congress on Evolutionary Computation, pp. 1278-1283, 2002.

X. Yao, Y. Liu, G. Lin; Evolutionary Programming Made Faster,*IEEE Trans. on Evolutionary Computation*, vol. 3, no. 2, pp. 82-102, 1999.

D. Zaharie; On the Explorative Power of Differential Evolution Algorithms, *Analele Univ. din Timişoara*, vol. XXXIX, special issue on Computer Science, pp 249-260, 2001.

D. Zaharie; Critical Values for the Control Parameters of Differential Evolution Algorithms, in R. Matoušek, P. Ošmera (eds.), *Proc. of Mendel 2002, 8th International Conference on Soft Computing*, pp. 62-67, 2002.

University of the West,
Faculty of Mathematics
Department of Computer Science
B-dul V. Pârvan, 4
Timişoara, 1900, ROMANIA
email:dzaharie@info.uvt.ro