Modified GA with the possibility of selecting a selection operator according to a set criterion

Razija Turcinhodzic, Zikrija Avdagic and Samir Omanovic

***Abstract*—**Genetic algorithms are used to solve complex problems in various areas. Research related to genetic algorithms mainly focuses on its three operators: selection, crossover, and mutation. The need to improve the algorithm has led to the creation of different operators out of the three mentioned, many of which are adapted to specific problems. This paper deals with the most commonly used selection operators, and their influence on the efficiency and robustness of the genetic algorithm. The idea behind this paper is to combine selection operators inside the genetic algorithm during its execution to decrease the risk of selecting the inappropriate selection operator for the considered test function. Operators are combined so that preference in the current generation is given to the operator which produces the most suitable population according to the set criteria after crossover and mutation. The criteria used in this paper are the best average overall fitness of the population and the best individual fitness. This research has shown that the change in selection operators within genetic algorithm has positive effects on its functionality.

***Keywords*—**Fitness, genetic algorithm, mathematical model, population estimation, selection operators.

I. Introduction

THE study of genetic algorithm (GA) is aimed at improving the algorithm efficiency, and mainly focuses on three operators: selection, crossover, and mutation. Each of these operators is important for the operation of the algorithm. This paper deals with the selection operator. Selection operator is of great importance because of its impact on GA in general [1] - [4]. The selection determines the evolutionary direction of GA and directly reflects “the survival of the fittest” theory of biological evolution. A higher selection pressure enables a higher number of copies of the best chromosome which is involved in the creation of a new population. A lower selection pressure is desirable at the beginning of the GA to provide a uniform search of the domain. A high selection pressure is recommended at the end of the algorithm in order to exploit the parts of the domain identified during the search as potential parts (sub domains) in which the solution to the problem could be found. At the beginning of the algorithm, a wide diversity of genetic materials is preferred. The initial population is thus chosen randomly to cover the domain as uniformly as possible (if it is assumed that no additional information about the considered problem is known). Through this, it can be directly seen that the selection operator has a significant impact on the GA efficiency, which is verified in some papers which deal with selection operators [5] - [7]. The most commonly used selections are roulette wheel, tournament and ranking selection. The importance of elitism is particularly pointed out as well as the possibility of combining two different selections when selecting individuals which will take part in creating a new generation offspring [8].

The development of different types of selection has opened room for the discussion about their suitability for solving a specific problem. There are claims which champion and question the superiority of a certain selection within the GA. It is not possible to assess which selection operator is more suitable for a particular type of problem. This paper focuses on the possibility of improving the GA by using multiple selection operators (modified GA).

Certain functions from the test functions group [9] were utilised to test the modified GA. The problem of searching for the global optimum (minimum or maximum) of the functions with one or more extremes (local or global) is discussed as well. The GA can fall into the local extremes; therefore the probability of finding the solution in the domain of optimum (ε area) is taken as one of the algorithm efficiency criteria.

Available literature did not provide the GA which combines multiple selection operators within it. In a related paper, adjustment of tournament selection has been performed by using fuzzy systems [10]. Furthermore, combinations of multiple GAs have also been used to exclude the possibility of selecting an inappropriate algorithm for the given problem [11]. Modified GA will be derived from the classic GA by making changes in a part of the selection (classifier). Classic GA operates with a single, parameter fixed selection operator, so its comparison with the modified algorithm shows how selection affects efficiency of GA. Crossover and mutation parameters remain unchanged. A large number of GAs and selection operators makes it impossible to test of all of their

Razija Turcinhodzic is with the Faculty of Electrical Engineering in Sarajevo - Department for Computer Science and Informatics, Zmaja od Bosne bb, Kampus Univerziteta, 71000 Sarajevo, Bosnia and Herzegovina (e-mail: razija.turcinhodzic@etf.unsa.ba).

Zikrija Avdagic is with the Faculty of Electrical Engineering in Sarajevo - Department for Computer Science and Informatics, Zmaja od Bosne bb, Kampus Univerziteta, 71000 Sarajevo, Bosnia and Herzegovina (e-mail: zikrija.avdagic@etf.unsa.ba).

Samir Omanovic is with the Faculty of Electrical Engineering in Sarajevo - Department for Computer Science and Informatics, Zmaja od Bosne bb, Kampus Univerziteta, 71000 Sarajevo, Bosnia and Herzegovina (e-mail: samir.omanovic@etf.unsa.ba).

combinations. Therefore, one of the existing GAs and three most commonly used selection operators have been selected.

II. The Modified Genetic Algorithm

The GA's performance is affected by encoding methods, operators, GA's initial parameter settings and success criteria (fitness) [12] - [14]. This research uses a classic GA with binary coding, elitism, fixed crossover and mutation operator values, which is the reference algorithm for comparing the results obtained with the modified GA. All experiment groups have the same initial parameters. The only difference is the number of selection operators.

In order to create the initial population, individuals are chosen in a manner defined within a given classic GA. Generally, the initial population should be chosen randomly, and the entire domain, within which a solution is sought after, should be covered. The initial population size remains unchanged throughout the algorithm, and is given as one of the input parameters. The original algorithm also works with a single selection operator. The termination criterion for this algorithm is reaching the number of generations set beforehand.

The existing papers dealing with selection operators have a different approach than those specified herein [4], [11], [15], [16]. Selection dictates the manner of selection of parent individuals. Different selection methods raise the question which of them should be used. Modified GA was developed to examine the difference between individual selection operators and their combination for the selected test functions.

Classic GA is modified so that it uses multiple selections in parallel: ranking selection function based on a normalised geometric distribution (norm. geom.), roulette and tournament selection. Classic GA is extended to include a new component and we will call it classifier S. The function of classifier S is to select, according to a set criterion, one (the most favourable) population out of multiple populations offered, each created with a different selection operator. This eliminates the need to select one type of selection, and offers the possibility for adapting the algorithm to the given problem. Modified GA is given in Fig.1.

In one of the scenarios, classifier is defined to compare the fitness value of the best individual for given populations, as well as the average overall fitness value of the given populations, and select the population with the best individual and the best overall fitness. If it happens that one population has an individual with the best fitness, while the other population has the best overall average fitness, the following criterion is used: if the difference in fitness values of the best individuals is small enough (10^{-2}) to be disregarded, the population with the best overall average fitness is selected. If that is not the case, the next criterion is used: if the difference in the populations' overall average fitness values is so insignificant that they can be disregarded, the population with the best individual is selected. If none of these conditions are met, the absolute difference between the best individuals within populations, and the absolute difference between average overall fitness values of given populations are analysed separately. If the difference is greater in fitness values of the individuals, the population with the best individual becomes the one to continue the GA. If the difference is larger in average overall fitness values of the populations, then the leading population becomes the one to continue the GA.

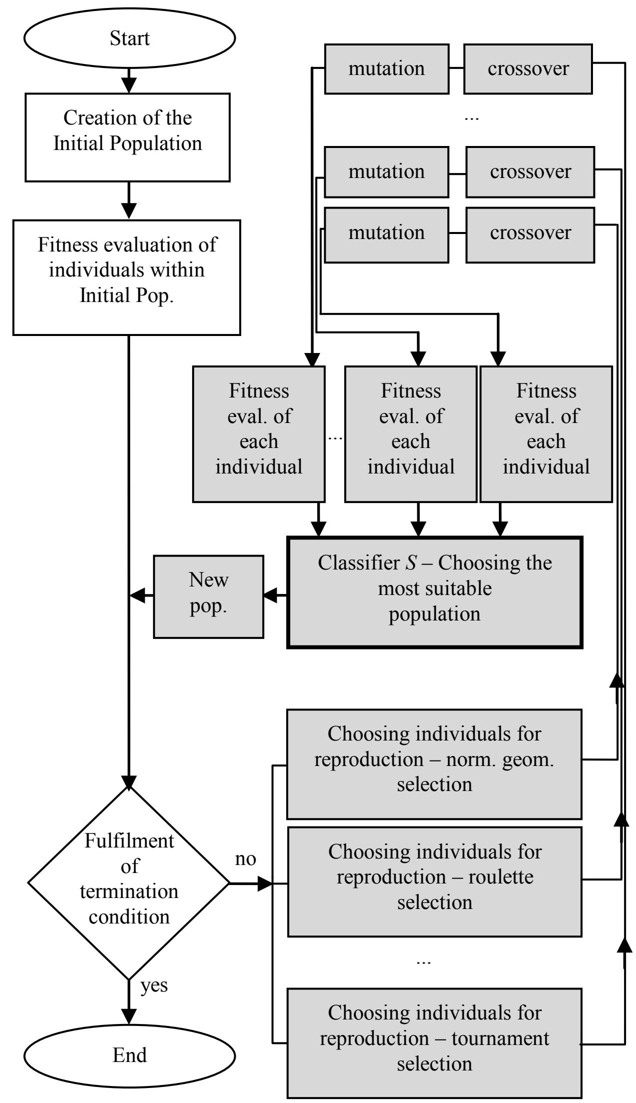

Figure 1. Modified Genetic Algorithm

The development of the modified GA (programme), as well as the simulations, was performed in MATLAB. A program called GAOTv5 was used to represent the performance of a simple (classic) GA (created in MATLAB by C.R. Houck, J.A. Joines, M.G. Kay). This program solves the problem of searching for the global optimum of the given function with one or more variables. Following programme modifications, simulation was performed using classic and modified algorithm on the same group of test functions, and the result were analysed.

III. Mathematical Model of the Modified GA

Block-scheme of classic GA can be represented as in Fig. 2.

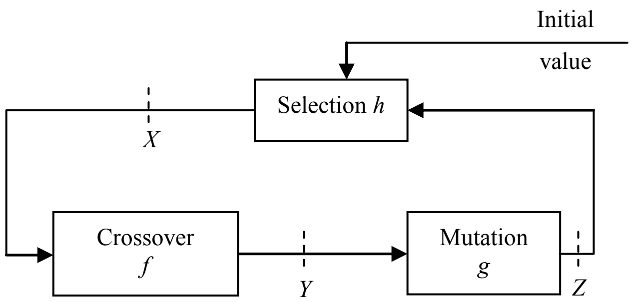

Figure 2. Classic genetic algorithm

Let $X = (x_1, x_2, \dots x_m)$, $Y = (y_1, y_2, \dots y_m)$, and $Z = (z_1, z_2, \dots z_m)$, be sets of m random variables. Size m is the number of individuals in the population.

Classic GA can be presented using the following mathematical model:

$$X \xrightarrow{f} Y \xrightarrow{g} Z \xrightarrow{h} X$$

whereby

- the crossover operator defines the mapping function f,
- the mutation operator defines the mapping function g,
- the selection operator defines the mapping function h.

A simplified block-scheme of classifier S is illustrated in Fig. 3, whereby size k stands for the number of populations classifier operates with. This number depends on the number of selections used for modified GA. Let the classifier output be the set α defined by the relation:

$$\alpha = \sup_\omega (Z_1, Z_2, \dots Z_k),$$

where $\alpha = (\alpha_1, \alpha_2, \dots \alpha_m)$ is a set of m random variables. Classifier S selects the best Z_j from the set $(Z_1, Z_2, \dots Z_k)$ using the given criterion ω. According to this $\alpha = Z_j$.

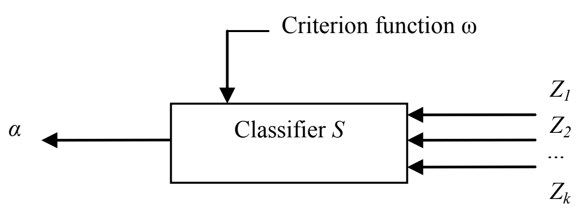

Figure 3. Classifier S

The aim is to determine the function of criterion ω that will give the highest probability of finding the optimum test function in the given area (ε area) of the global test function extremum.

The following values can be considered for the criterion function ω:

1. the fitness value of the best individual for the given population,
2. the best average overall fitness value of the population,
3. min/max dispersion value of individuals comparing to the average value of individuals,
4. min/max dispersion value of individuals in relation to the best individual in the population,
5. combinations of criteria.

A. ε Area

Let $x = (x_1, x_2, \dots x_n)$ be the n-dimensional function which is optimum in the point $x^0 = (x_1^0, x_2^0, \dots x_n^0)$.

Let $D_i, i = 1,2, \dots, n$ be the domain of the i^{th} component of the function x.

Let ε area around the optimal point x^0 be defined with points that deviate by ±5% of the domain size $D_i, i = 1,2, \dots, n$.

The ε area is then defined with the relation $\varepsilon_i = x_i^0 \pm \mathrm{D_i}/20, i = 1,2, \dots, n$. In this paper, any solution of the n-dimensional function x, belonging to the ε area of optimum x^0 defined in such a way, is satisfactory.

IV. Two–Dimensional Functions

A. Description of the Experiment

This experiment will include three separate selection operators (norm. geom. (N), roulette (R) and tournament (T)) and their combinations. It makes the total of 7 combinations (Table I).

Table I: Selections and their combinations

Selection operator		
1.	norm. geom.	**N**
2.	roulette	**R**
3.	tournament	**T**
4.	norm. geom. and roulette	**NR**
5.	norm. geom. and tournament	**NT**
6.	roulette and tournament	**RT**
7.	norm. geom., roul. and tourn.	**NRT**

The aim of this group of experiments is to analyse if any of those 7 combinations offer the best results in the process of searching for the given function's global optimum. The best operator is the one which provides the best chance of finding the optimum, and the one which finds the optimum with the smallest margin of error.

Mutation and crossover are invariant in all the experiments. Depending on the test function, s simpleXover program function with the crossover probability of 0.4 or 0.6 is used for crossover, and a binary mutation with the mutation probability of 0.005 or 0.05, is used for mutation. Each selection receives additional parameters if needed. These parameters are invariant for all the experiments. Norm. geom. selection has 0.08 probability of selecting the best individual, the roulette selection has no other additional parameters, and the tournament selection operates with the tournament size 4.

When combining selection operators, the selection of the most suitable selection was performed based on the biggest average overall fitness value, and the best individual criteria. Three types of selection are applied separately on the current population, followed by crossover and mutation, while determining which one of the three resulting populations yields the best results based on the set criterion. The best population is selected in continuing the process.

All experiments span over 25 generations. The number of individuals within populations varies. The number of individuals is 5, 10, 15, 20, 30, 40 and 45, and it remains the same for all generations. The programme draws two graphs. One graph represents the appearance of the function as well as the individuals which constitute the initial population, and individuals from the final population. The other graph consists of two curves which represent the population's average fitness value and the best individual value in every generation. The y-axis in second graph represents fitness value, while x-axis represents the number of the generation.

Each experiment which fixates the selection and the number of individuals was successively performed 30 times. 7 different population sizes with 7 abovementioned selection operator combinations are used. The overall number of experiments is 7·7·30 = 1470. The quality of success in finding the extremum with a certain operator is determined by:

- the probability of falling into a pre- defined ε area (local extremum avoided).
- the relative error which defines the deviation of the discovered extremum in relation to the actual extremum.

Three test functions will be used: one is the function with one independent variable, and the other two are functions with two independent variables.

B. Test functions and results

The following three test functions were selected: one-dimensional function with 7 local extremums (F2), and two two-dimensional functions, Goldstein Price (F3) and Six-Hump Camel Back (F4).

The analytical form of the F2 function is $f(x)=x+10\sin(5x)+7\cos(4x)$. The function maximum is established within the interval [a b] (a = 0 and b = 9), and its value is 25. Fig.4 represents the results of one of the experiments for which the size of the population is 20, and for norm.geom. selection operator.

Goldstein Price function (F3) with two independent variables is shown on Fig. 5. The absolute minimum of the given function is at the point (0,-1), and the value of the function is 3. The highest value of the function is 1 014 600. Since the difference between the minimum and the maximum value of the function is large for such a small domain of independent variables, it can cause even double digit deviations from the optimum even when the final individual is within its domain.

Table II shows the results of the experiment for the F3 function (results for the F2 function are not presented).

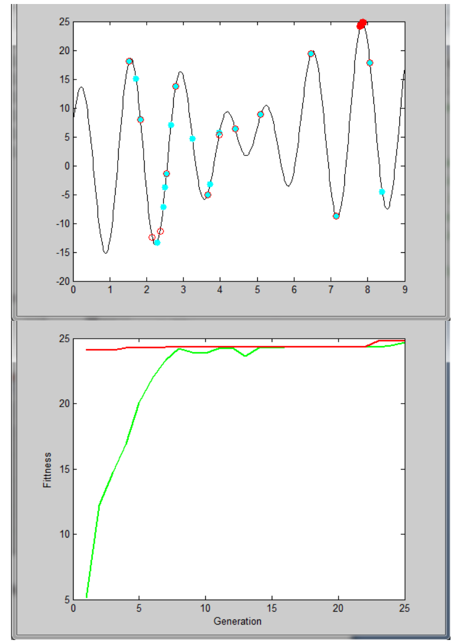

Figure 4: Norm. geom. selection with 20 individuals for the F2 function

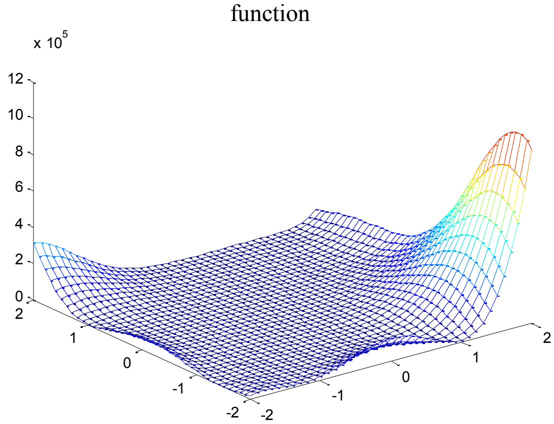

Figure 5: Goldstein-Price Function

Table II. Statistics for Goldstein Price Function (F3)

	Selections							
	N		*R*		*T*		*NRT*	
pop. size	*no*	*prob*	*no*	*prob*	*no*	*prob*	*no*	*prob*
5	23	0.77	21	0.7	17	0.57	19	0.63
10	26	0.87	16	0.53	21	0.7	28	0.93
15	28	0.93	17	0.57	24	0.8	29	0.97
20	29	0.97	11	0.37	29	0.97	30	1
30	29	0.97	16	0.53	30	1	30	1
40	30	1	24	0.8	29	0.97	30	1
45	30	1	19	0.63	30	1	30	1

	Selections					
	NR		*NT*		*RT*	
pop. size	*no*	*prob*	*no*	*prob*	*no*	*prob*
5	19	0.63	14	0.47	21	0.7
10	27	0.9	23	0.77	25	0.83
15	30	1	28	0.93	28	0.93
20	30	1	29	0.97	29	0.97
30	30	1	30	1	30	1
40	30	1	30	1	30	1
45	30	1	30	1	30	1

Column 1 represents the size of the population. Two columns are given for each of the 7 operators (N, R, T, NRT, NR, NT, RT), where the first (no) represents the overall number of the solutions found in the domain of the global extremum in 30 experiments, and the second column (prob) represents the probability of being in the domain of the global extremum (ε area).

Six-hump Camel Back function (F4) is presented in Fig. 6. The function has two global and two local minima. Global extremums are at points (0.089842, -0.712656) and (-0.089842, 0.712656). The function value at these points is -1.031628453.

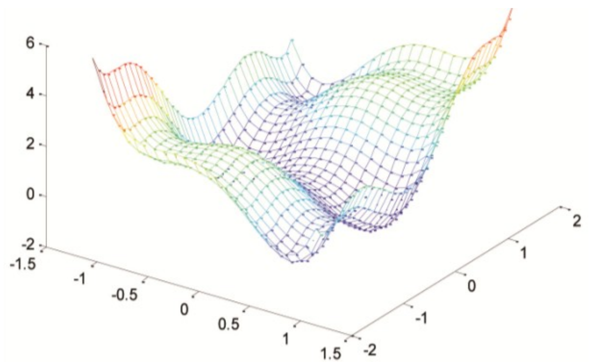

Figure 6. Six-Hump Camel Back Function (F4)

Table III shows the statistical results for the F4 function.

Table III: Statistics for Six-Hump Camel Back function (F4)

	Selections							
	N		*R*		*T*		*NRT*	
pop. size	*no*	*prob*	*no*	*prob*	*no*	*prob*	*no*	*prob*
5	28	0,93	18	0,6	27	0,9	29	0,97
10	29	0,97	24	0,8	30	1	30	1
15	30	1	17	0,57	28	0,93	30	1
20	30	1	23	0,77	30	1	30	1
30	30	1	25	0,83	30	1	30	1
40	30	1	29	0,97	30	1	30	1
45	30	1	29	0,97	30	1	30	1

	Selections					
	NR		*NT*		*RT*	
pop.size	*no*	*prob*	*no*	*prob*	*no*	*prob*
5	28	0,93	30	1	27	0,9
10	30	1	30	1	30	1
15	30	1	30	1	30	1
20	30	1	30	1	30	1
30	30	1	30	1	30	1
40	30	1	30	1	30	1
45	30	1	30	1	30	1

C. *Analysis of the Results*

Based on previously mentioned experiments for the F2 function, it can be observed that the norm. geom selection in some cases yields better results than multiple selection combinations. In the F3 function it is obvious that the roulette selection is unfavourable and that any of the multiple selection combinations provides better results for all sizes of the population, with the exception of the population size 5.

The F3 function, unlike the F2 function, shows a large span between the minimum and the maximum, which causes large deviations from the global extremum even in the cases when it is found in its ε area. A large span of the function's value at the small interval from which independent variables x_1 and x_2 are selected causes large deviations. The table which presents the deviation of the best individual fitness from the extremum fitness (not shown here) shows that in the roulette selection the best individual rarely approaches the optimum with the fitness value under 10 (optimum is 3). Furthermore, when combining the roulette selection with other selections, in the case of 10 and more individuals, it gives better results. This demonstrates that a wrong choice of selection can significantly affect the result of the GA. In the F2 function, all the deviations are within the range of 1% to 4%. In the F4 function, if the individual is found within its domain, deviations from the extremum are lesser when compared to the F3 function (table not shown). The reason for this is the nature of the function.

Table IV offers the overview of the best operators for all three test functions, and all population sizes. For each function, the Table offers two best operators for all 7 population sizes. Quality assessment of the operator is performed based on the probability of finding the solution approaching the global optimum (ε area).

The experiment with 45 individuals within the population has not been done for the F2 function, as well as the experiments which combine two selection operators for the population of 40. Seeing that there are cases where certain operators and their combinations offer almost equal probabilities of finding the solution approaching the optimum, multiple operators are listed. If two or more operators offer the same probability, they are bold. The mark 2S means that all three combinations of the two operators are implied. The mark *-R means that all the operators and their combinations, with the exception of the roulette selection, are implied.

Table IV. The best selections for individual test functions and population sizes

	Selection combinations yielding best results								
pop. size	*F2*			*F3*			*F4*		
5	NR	RT		NRT	N		NRT	**N**	**NR**
10	NR	NRT		N	NRT		**T**	**NRT**	**2S**
15	NRT	NT		N	RT		**N**	**NRT**	**2S**
20	NT	N		**NR**	**NT**	**RT**	***-R**		
30	NR	N		**T**	**NRT**		***-R**		
40	R	T		**N**	**NRT**	**2S**	***-R**		
45	-	-	-	***-R**			***-R**		

After completing the experiments, it is obvious that it is impossible to give preference to one selection operator or its combination over the other(s). The nature of the function and the size of the population affect the probability of finding the solution approaching the optimum, and deviation as well. The optimal size of the population depends on the domain of the independent variables and the nature of the function. This includes: the number of global and local optimums, the probability that the individual falls within the domain of the global optimum, the function's co-domain span, and the number of independent variables. When a function has a big gradient influx in the optimum domain, the final individual's fitness deviates from the global extremum fitness more than in a function with a smaller gradient influx in the optimum domain. The F3 function has the largest gradient influx in the optimum domain of all the tested functions.

Table V gives the overview of the worst operators and their combinations for the three given functions. The remarks are the same as for Table IV.

Table V: The worst selections for individual test functions and individual population sizes

pop. size	Selection combinations yielding worst results								
	F2			F3			F4		
5	T	**R**	**NT**	R	NR		R		
10	R			R	**T**	**NR**	R		
15	N			R	T		R		
20	**R**	**T**	**NRT**	R			R		
30	NRT			R	NR		R		
40	NRT			R	T		R		
45	-	-	-	R			R		

In order to check which of the selections and selection combinations has the highest probability of finding the solution, the average value of each selection and its combinations is calculated. The last row in Table VI ,AVR, represents the average probability value of finding the solution approaching the optimum for individual test functions.

Table VI: Average value of the probability (AVR) to find the solution approaching the optimum of individual test functions

fun.	Selections						
	N	R	T	NRT	NR	NT	RT
F2	0.578	0.550	0.556	0.594	0.613	0.553	0.533
F3	0.871	0.142	0.819	0.857	0.800	0.823	0.857
F4	0.986	0.714	0.976	0.995	0.990	1.000	0.986
AVR	0.812	0.469	0.784	0.815	0.801	0.792	0.792

The table shows that the NRT combination offers the highest average probability of finding the solution. Next in line is the N selection, followed by three selections which combine two selection operators.

Even though one of these selections or their combinations cannot be given preference over the other(s), it can be noted that combining the selections improves the probability of finding the solution when, among the chosen combined selections, there is one which offers much worse results than the average of other combinations. In functions F3 and F4 it is found that the roulette selection is the least favourable for these types of function. The probability of finding the solution approaching the optimum using the roulette selection is lower compared to these probabilities for other operators and their combinations. This means that combining the roulette selection with other selections gives better results than using only the roulette selection.

Figure 7 summarizes the results of 20 experiments. The chart on the left shows the probability of individual selections giving the best results out of 20 experiments, while the chart on the right shows the probability of individual selections giving the worst results out of 20 experiments. The assessment of the selections' quality is done based on the probability of finding the solution approaching the optimum.

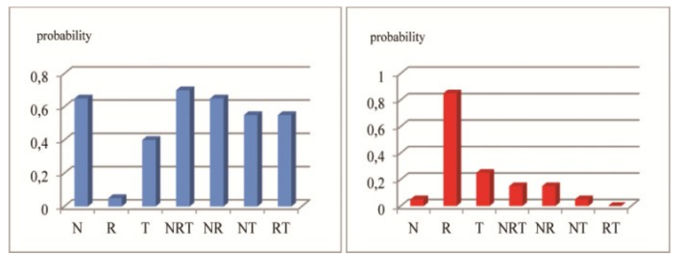

Figure 7: Selections with the best results (left) and selections with the worst results (right)

This segment presents results related to only one tested classifier criterion. This criterion gave the best results and was chosen as such to be used in subsequent tests with the n-dimensional functions. The results with dispersion did not prove to be efficient and were excluded from further analysis. Dispersion gives good results if it includes n copies of the best individual, but this experiment combines individuals from different populations and does not belong to this group of experiments.

V. N-DIMENSIONAL FUNCTIONS

Experiments in the following segment will be performed as in previous research but with the n-dimensional functions.

A. Description of the Experiment

Scientists have used various function groups in order to study the performance of GA, and a lot of research involves the definition of control parameters and their potential adjustment. There are still no general conclusions when it comes to the optimum parametrisation of operators. The paper [17] offers the most commonly used sets of test functions in the analysis of various GA models, and proposed sets of parameters.

Three functions have been selected from this set of functions. The first one is known as De Jong function F1. It is a smooth, unimodal, strongly convex and symmetric function. The other two functions are Schwefel and Rastrigin functions and are typical examples of non-linear multimodal functions. Rastrigin's function is a difficult issue for GA due to the large area that needs to be searched (domain), and a large number of local minima. Schwefel's function is easier than Rastrigin's function. It is characterised by a second-best minimum which is far from the global optimum. To make it easier to follow, functions will be abbreviated to: F1 (De Jong function), F6 (Schwefel's function), and F7 (Rastrigin's function).

Experiments will be performed on these three functions with 10 and 20 independent variables. Since crossover and mutation are not subject to the analysis in these experiments, their probabilities are fixed at 0.7 for crossover and 0.01 for mutation. Population size plays an important role in GA. Optimal population size depends on the domain complexity. In order to use the same population size for all experiments, a population size of 80 or 240 is chosen. This size satisfies 10 independent variables for Rastrigin's function, but the probability to find a solution near the optimum is low. For 20 independent variables the selected population size is insufficient to find a solution near the optimum. Increasing the population size will not be performed in this experiment, because necessary information is obtained from 10 independent variables, if we consider the difficulty of Rastrigin's function.

Finding a suitable set of test problems is not an easy task, since a particular combination of properties represented by any given set of test functions does not allow making general statements about GA performances. It is only possible to draw limited general conclusions on the effectiveness of GA. It is very likely that there are test functions which will give different results.

For the computational limited GA, a too large or too small population reduces the quality of the final solution. The paper [17] suggests that the population size should be 200-250, and that the crossover probability percentage should be 0.7-0.75. According to these recommendations, the population size is set to 240 and the crossover probability to 0.7. The GA stops after 500 generations. Each experiment is performed 20 times.

Classifier will have one or 3 selections (N, R, T, 3N, 3R, 3T, and NRT combinations). The results of the GA success will be determined by the probability of finding the solution in the given extremum area (ε area). There will be two different criterion functions:

- the average overall fitness value of the given population (C1)
- the average overall fitness value of the given population and the fitness value of the best individual (C2).

Classic elitism will be used in both cases. Dispersion criterion is left out in these experiments since the tests on two-dimensional functions using this criterion did not provide favourable results.

B. Results and analysis of the Results

A series of experiments is performed on the F1, F6 and F7 test functions with 10 and 20 independent variables. Classifier uses the criterion of the highest average overall fitness value of the given population (**C1**). Table VII shows the statistical results of all experiments. Column 1 and 2 represent the selection operators and the population size respectively. The table contains three columns for each test function (10 and 20 independent variables). The first column (gen) is the average number of generations needed to get in the ε area, the second column (no) represents the number of experiments for which the solution was found in ε area, out of 20 experiments, and the third column (prob) is the probability of finding the solution. The results for Rastrigin's function with 20 independent variables are left out, because the solution is not found in this case.

Experiments use N, R, T selections, and their combinations: 3N, 3R, 3T i NRT. The population size in experiments is 80 or 240. For populations with 240 individuals only 4 combinations are used: N, R, T and NRT. Two different combinations are performed for N, R and T in case of 80 individuals in the population. The first one works with classic GA, and the second one works with the same algorithm as the NRT combination, but uses the same selection operators: 3N, 3R and 3T.

Table VII: Statistic values for criterion C1 and functions F1, F6 and F7

Sel.	Pop. size	Function								
		De Joung 1 (F1)						**Rastrigin (F7)**		
		10			20			10		
		gen	no	prob	gen	no	prob	gen	no	prob
NRT	240	10,5	20	1	23	20	1	29,5	2	0,1
N	240	9,4	20	1	20	20	1	25,1	2	0,1
R	240	0	0	0	0	0	0	0	0	0
T	240	14,8	20	1	35,5	20	1	60,8	6	0,3
NRT	80	18,4	20	1	39,3	20	1	0	0	0
N	80	16,4	20	1	37,2	20	1	36	2	0,1
R	80	0	0	0	0	0	0	0	0	0
T	80	21,2	20	1	45,2	20	1	22,2	2	0,1
3N	80	16	20	1	35,4	20	1	0	0	0
3R	80	0	0	0	0	0	0	0	0	0
3T	80	19,3	20	1	41,9	20	1	14,1	1	0,05

Sel.	Pop. size	**Schwefel (F6)**						**AVR**
		10			20			
		gen	no	prob	gen	no	prob	
NRT	240	86,1	19	0,95	263	16	0,8	0,77
N	240	129	20	1	201	16	0,8	0,78
R	240	46,6	10	0,5	14	1	0,05	0,11
T	240	41,7	20	1	166	16	0,8	0,82
NRT	80	178	13	0,65	54,8	3	0,15	0,56
N	80	144	14	0,7	35	2	0,1	0,58
R	80	105	11	0,55	0	0	0	0,11
T	80	201	19	0,95	134	8	0,4	0,69
3N	80	153	14	0,7	216	16	0,8	0,7
3R	80	93,2	7	0,35	72,2	10	0,5	0,17
3T	80	182	16	0,8	150	17	0,85	0,74

Population size **80** and classic selections **N**, **R** and **T** give the following results: N and T find always the solution near the optimum for the F1 function with 10 variables, and R cannot find the area of optimum at all. The same is the case with 20 variables. T gains the best results for the F6 function, with the probability of 0.95 and 0.4, in case of 10 and 20 variables respectively. Selection R does not find the optimum solution. For the F7 function with 10 variables the best results are obtained with N and T with the probability of 0.1, and R cannot find the area of optimum. In case of **240** individuals in the population, N and T again find the solution for F1 and 10 variables with the probability of 1, and R cannot find the optimum solution at all. For function F6 with 10 variables, N and T have the probability of 1, and R of 0.5. For 20 variables N and T have the probability of 0.8, and R of 0.05. T (0.3) obtains the best results for the F7 function with 10 variables. R cannot find the optimum solution.

For population size **80** and **combination of selections**, each combination in case of the F1 function (10 and 20 variables) has the probability of 1, excluding 3R, which cannot find the optimum solution. 3T (0.8) has the highest probability for the F6 function with 10 variables, while NRT has the probability of 0.65. For 20 variables 3T has the probability of 0.85, and NRT of 0.15. In the F7 function with 10 variables only 3T can find the solution with the probability of 0.05.

In this group of experiments, classifier uses the average overall fitness value of the given population and the fitness of the best individual as the criteria (**C2**). The selection of the most suitable population according to these two criteria was described in detail earlier. Table VIII shows the statistical results of all experiments, each of which is performed 20 times.

Table VIII: Statistic values for criterion C2 and functions F1, F6 and F7

Sel.	Vel. pop.	Function								
		De Joung 1 (F1)						Rastrigin (F7)		
		10			20			10		
		gen	no	prob	gen	no	prob	gen	no	prob
NRT	240	10,9	20	1	21,8	20	1	19,3	2	0,1
N	240	9,4	20	1	20,7	20	1	31,5	3	0,15
R	240	0	0	0	0	0	0	0	0	0
T	240	14,5	20	1	35,5	20	1	137	13	0,65
NRT	80	15,5	20	1	36,3	20	1	13,8	1	0,05
N	80	15,9	20	1	37,9	20	1	0	0	0
R	80	0	0	0	0	0	0	0	0	0
T	80	20,1	20	1	46,6	20	1	59,2	3	0,15
3N	80	14,5	20	1	31,8	20	1	49,1	3	0,15
3R	80	0	0	0	0	0	0	0	0	0
3T	80	17,3	20	1	38,1	20	1	0	0	0

Sel.	Vel. pop.	Schwefel (F6)						AVR
		10			20			
		gen	no	prob	gen	no	prob	
NRT	240	87,9	20	1	190,85	16	0,8	0,78
N	240	92,5	20	1	187,95	15	0,75	0,78
R	240	58	10	0,5	0	0	0	0,167
T	240	36,2	20	1	191,7	17	0,85	0,9
NRT	80	193,55	20	1	87,7	6	0,3	0,67
N	80	199,8	17	0,85	12,8	1	0,05	0,58
R	80	91,35	8	0,4	0	0	0	0,13
T	80	208,15	17	0,85	40,1	2	0,1	0,62
3N	80	219,55	20	1	72,75	5	0,25	0,68
3R	80	68,65	11	0,55	10,85	1	0,05	0,2
3T	80	114,7	20	1	152,1	8	0,4	0,68

Population size **80** and classic selections **N**, **R** and **T** give the following results: N and T find always the solution near the optimum and R cannot find the optimum solution at all for the F1 function (with 10 and 20 variables). N and T gain equal probability (0.85) for the F6 function in case of 10 variables, while selection R is weaker more than twice (0.4). For F6 and 20 variables T (0.1) gives the best results, and R cannot reach the area of optimum. In the F7 function with 10 variables only T can find the solution with the probability of 0.15. In case of **240** individuals in the population, N and T again find the solution with the probability of 1, and R cannot find the area of optimum at all. For the F6 function with 10 variables, N and T have the probability of 1, and R is twice weaker (0.5). For 20 variables T has the best results (0.85), and R the worst (0). T (0.65) obtains the best results for the F7 function with 10 variables. R cannot find the optimum solution.

Results for population size **80** and **combination of selections** are: for the F1 function, for 10 and 20 variables, the probability for all combinations is 1, besides 3R, which cannot find the optimum solution. The same is true for the F6 function with 10 variables, with difference that the probability of R increases from 0 to 0.55. In case of the F6 function with 20 variables the best combination is 3T (0.4), followed by NRT (0.3), and the worst combination is 3R (0.05). In the F7 function with 10 variables only 3N and NRT can find the solution with the probability of 0.15 and 0.1 respectively. Two additional combinations cannot find the area of optimum.

Fig. 8 shows the average probability of finding the solution in ε area for single selections and their combinations for both criteria (C1 and C2).

In case of **C1** and **individual selections**, T gives the best results for all three functions regardless of the population size. R has the worst results. For all three functions **combination** 3R provides the worst results, except in one case, while 3T combination provides the best results in all cases. NRT can partly neutralise the worst selection (in this case R), apart from the F6 function with 20 variables. The combination of different selections cannot achieve such good results as the combination of the same selections, if the most appropriate selection for the considered function is chosen. It should be noted that the dominant selection in experiments with one and two variables is selection N. The worst selection in this case is also selection R. It can be concluded that it is not possible to predict which selection will give the best results for particular functions. It is not even possible to make this estimation for the same function, because the number of variables can play an important role as well.

In case of **C2** and **individual selections**, T gives the best results for all three functions regardless of the population size, while N can achieve equally good values in some cases. R has the worst results. For all three functions **combination** 3R gives the worst results. There is no best combination for C2. NRT combination mainly neutralises the worst selection R. NRT does not achieve such good results as the combination of the same selections, if the most appropriate selection for considered problem is chosen. C2 significantly increases the probability of finding the solution in the area of optimum for the NRT combination, but it can have a negative influence on other combinations (the F6 function with 20 variables). In this experiment selection N is again pointed out as a good selection, like it was in experiments performed on one- and two-dimensional functions. These experiments show that the selection choice has a significant influence on GA performances, thus there is a risk of choosing the inadequate selection for the considered function.

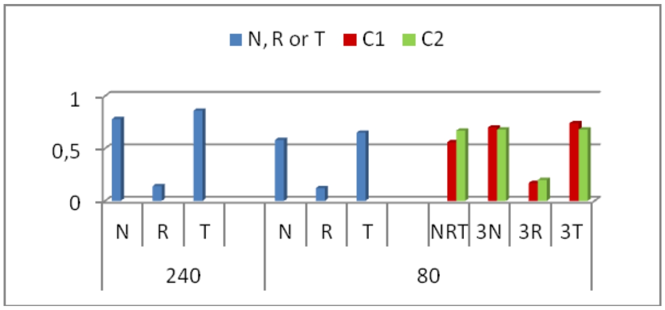

Figure 8: Results for individual selections, C1 and C2

Combining selections can improve the results of the algorithm. In some cases, it is better to combine the same selections, but the problem is to estimate which selection is adequate for which type of problem. If this is not possible, as is usually the case, it is better to combine different selections, rather than risk to chose the worst possible selection for the considered problem. Classifier criterion used to determine specified selections, i.e. the selection which creates the most appropriate population, has a significant role in the GA efficiency.

VI. Conclusion

The original idea behind this paper is the introduction of a new module—classifier—into the classic GA.

The function of the classifier is to choose the most suitable population among multiple populations, from which each was generated using a different selection operator, using set criteria ω in order to increase the GA efficiency. This choice is made in each cycle when forming a new population. N, R and T selection operators and their combinations are used in this paper. The criterion ω uses the best individual fitness, the best average population fitness and their combinations. The following test functions were used: one-dimensional, two-dimensional, 10-dimensional and 20-dimensional function. The efficiency of GA is measured by the probability of finding the functions optimum in the ε area of the optimum point given beforehand. The population size was different depending on the complexity of the test function.

The aim of this research was to try to determine the selection operator combination and criterion function ω that will give the best results to determine the optimum of all considered test functions and other potential test functions. Modified GA is in fact an extension of the classic GA with the classifier. This paper presents a mathematical model of the modified GA that also describes its functionality.

Experiments were carried out using simulation. Each experiment was repeated 20 or 30 times. Results were analysed statistically. It was not possible to determine which selection operator or its combination is the best for all considered test functions under the given constraints. But, in general, classifier with the NRT combination can conditionally be considered the most favourable for two reasons:

1. NRT combination gives the best results or was in top three selection operators in most experiments;
2. using classifier with different selection operators can significantly decrease the risk of selecting an inadequate selection operator for the considered test function.

For further research, it is necessary to: performs a similar analysis for other selection operators and their combinations; explore the influence of other criterion functions ω used and mentioned in this paper in relation to this problem; consider this research particularly from the aspect of clusters, fuzzy-logic and neural networks.

References

[1] N. M. Razali, J. Geraghty, “Genetic Algorithm Performance with Different Selection Strategies in solving TS“, *Proceedings of the World Congress of Engineering*, Vol II, London UK, 2011.

[2] A. P. Conrad, R. S. Roos, G. M. Kapfhammer, V. Petridis, “Empirically Studying the Role of Selection Operators During Search-Based Test Suite Prioritization“, *ACM SIGEVO Genetic and Evolutionary Computation Conference*, Portland – Oregon, 2010.

[3] A. E. Eiben, M. C. Schut, A. R. de Wild, “Boosting Genetic Algorithms with Self-Adaptive Selection“, *IEEE Congress on Evolutionary Computation*, Canada, pp. 477 – 482, 2006.

[4] M. Affenzeller, S. Wagner, “Offspring Selection: A New Self-Adaptive Selection Scheme for Genetic Algorithms“, *Adaptive and Natural Computing Algorithms: Proceedings of the International Conference in Coimbra*, Portugal, 2005.

[5] Huang Bao-Juan, Zhuang Jian, Yu De-Hong, “A Novel an Accelerated Genetic Algorithm“, *WSEAS Transactions on Systems and Control*, vol.3, no. 4, 2008.

[6] Jianhai Yu, Zhigango Mao, “A design method in CMOS operational amplifier optimization based on adaptive genetic algorithm“, *WSEAS Transactions on Circuits and Systems*, vo. 8, no. 7, pp. 548-558, 2009.

[7] M. F. Mohammad Kamal, T. K. Abdul Rahman, I. Musirin, “Application of improved genetic algorithms for loss minimisation in power system“, *Power and Energy Conference,* 2004.

[8] A. Neubauer, “The Intrinsic System Model of the Simple Genetic Algorithm with α-Selection, Uniform Crossover and Bitwisw Mutation“,Proceedings of the 14th WSEAS international conference on Systems: part of the 14th WSEAS CSCC multiconference, vol II, pp. 538-543, 2010.

[9] Z. Michalewicz, Genetic Algorithms + Data Structure = Evolution Programs. Springer – Verlag Berlin Heidkeberg, 1999.

[10] Soung-Min Im, Ju-Jang Lee, “Adaptive Crossover, Mutation and Selection Using Fuzzy system for Genetic Algorithms“, *Artificial Life and Robotics*, vol. 13, no. 1, pp. 129-133, 2008.

[11] F. Peng, K. Tang, G. Chen, X. Yao, “Population-Based Algorithm Portfolios for Numerical Optimization“, *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 5, pp. 782 – 800, 2010.

[12] A. E. Eiben, R. Hinterding and Z. Michalewicz, Parameter Control in Evolutionary Algorithms, *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 2, pp. 124 - 141, 1999.

[13] K. Weicker, Basic principles for understanding evolutionary algorithms, Fundamenta Informaticae XXI, 2003.

[14] M. C. Popescu, L. Popescu, N. Mastorakis, “Applications of genetic algorithms“, *WSEAS Transactions on Information Science and Applications*, vol. 6, no. 11, 2009.

[15] L. Li, Y. Ren, C. Yang, “Path Planning Based on Dynamic Sub-population Pseudo-Parallel Genetic Algorithm“, Journal of Environmental Technology and Engineering, vol.1, no. 1, 2009.

[16] J. A. Vrugt, B. A. Robinson and J. M. Hyman, “Self-Adaptive Multi Method Search for Global Optimization in Real-Parameter Spaces“, *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 243 – 259, 2009.

[17] J. G. Digalakis and K.G. Margaritis, “An experimental study of benchmarking functions for genetic algorithms“, *Intern. J. Computer Math*, vol. 79, no. 4, pp 403-416, 2002.