Recombination Operators for Evolutionary Algorithms.

Daniela Zaharie
Faculty of Mathematics,
The West University of Timişoara, bv. V. Pârvan no. 4, 1900-Timişoara, ROMANIA,
email: zaharie@tim1.math.uvt.ro

Abstract

In this paper is analyzed the influence of some recombination operators on the convergence and exploitative power of a class of evolutionary algorithms without selection. Two variants of combining the variation operators are analyzed: sequential and alternate application of recombination and mutation. A hybrid type of recombination is proposed and its behaviour is illustrated by some experiments on classical test functions.

Keywords: evolutionary algorithms, fitness gain, relative progress rate.

1 Introduction

An evolutionary algorithm (EA) is a stochastic optimization method that uses a *population* of explorers in the solution space which evolves on the basis of three nature inspired mechanisms: *selection*, *mutation* and *recombination*. In the last two decades the EAs have been extensively used in solving optimization problems arising in different domains such as: *system design and modelling* (for optimal structural design, shape optimization and parameters identification), *production planning* (for solving allocation problems and job shop scheduling), *artificial intelligence* (for learning processes and neural network topology optimization), *physics and data analysis* (for determining low energy configurations in systems of many particles and parameters estimation in regression models). For a review of EAs applications see [1]

Some of the properties of EAs which recommend them as valuable alternatives to the classical optimization methods are: (a) they use minimal informations on the objective functions (do not use derivatives, so they can be called 0-order methods); (b) they can handle both constrained and unconstrained optimization problems, both on discrete, continuous and mixed search spaces; (c) they have the ability to avoid local minima of the objective function being appropriate for global optimization tasks; (d) are robust against disturbances in the determination of the objective function values; (e) are flexible, easy to be implemented and hybridized with domain dependent heuristics.

The aim of this paper is to analyze the influence of some types of recombination operators on a class of EAs used in solving a global minimization problem. Let us consider the problem of finding $x^* \in D$ with $f(x^*) = \min_{x\in D} f(x)$ for $f : D \subset I\!\!R^n \to I\!\!R$ a function defined on a bounded and convex set $D \subset I\!\!R^n$. The general structure of an EA is:

`Initialize` $\overline{x}(0)$, $t = 0$
`Repeat`
    `Apply selection` $(\overline{x}(t) \to \overline{x}'(t))$
    `Apply variation operators` $(\overline{x}'(t) \to \overline{x}''(t))$
    `Replace the old population with the new one` $(\overline{x}''(t) \to \overline{x}(t+1))$
    $t = t + 1$
`Until a stopping criterion is satisfied`

In the above algorithm $\overline{x}(t) \in D^m \overset{\text{not}}{=} \mathcal{P}$ denotes the population corresponding to generation t. The population size is m and $\mathcal{P}$ denotes the set of all m-sized populations with elements from D. The selection step has the role of preparing an intermediary population $x'(t)$ which will be transformed by the variation operators. Its purpose is the *exploitation* of the search space by emphasizing good solutions [4]. The purpose of variation operators (mutation and recombination) is to use the parent population formed after the selection to create an offsprings population. They assure both the exploitation and the *exploration* of the search space. The replacement step determines which members of the previous population are replaced by the elements generated through the variation operators. The stopping criterion is determined either by a maximal number of generations or by some properties of the population.

Since our interest is on the role and comparison of some variation operators (mainly recombination) we will consider an EA without an explicit selection step. The replacement step will be used only to assure the elitism of the algorithm (the best element is preserved in the next generation).

Some of the main problems in designing an EA is to choose the appropriate variation operators and to establish how to combine them. While the first problem has already been approached [2][7] little interest has been shown on the second problem.

Regarding the first problem we will consider that the mutation consists in an additive random perturbation of population elements. The interest will be in comparing the behaviour of some recombination operators: classical intermediate and discrete recombination and some particular variants which use the best element of the current generation. Regarding the second problem we will analyze two cases: one when the recombination and mutation are applied sequentially (the usual way of combining the two operators) and one when they are applied alternatively (in one generation, an element of the population is transformed either through recombination or through mutation but not through both of them). The idea of an alternate application of operators have been introduced in [5] for selection and mutation.

The paper is organized as follows. Section 2 presents the recombination types which will be studied. In Section 3 is analyzed the influence of the recombination on the convergence reliability while in Section 4 is studied the influence on the convergence velocity. Some numerical results for classical test functions are presented in Section 5.

2 The structure of the algorithm and recombination operators

We will analyze an EA without explicit selection. The transformations of the population $\overline{x}(t)$ during one generation are:

`For` $i = \overline{1, m-1}$
  $X_i(t+1) = M \circ R(\overline{x}(t))$
$x_m(t) = x_*(t)$

with $M \circ R$ the variation operator whose effect depends on how the mutation and recombination are combined. The variation operators are applied in a probabilistic manner, i.e. for each i the mutation is applied with probability p_m and the recombination is applied with probability p_r. In the following we will consider that the mutation consists in adding a zero mean random variable, ξ, to the current element, $M(x_i(t)) = x_i(t) + \xi$. Thus, for the sequential application of recombination and mutation one obtains:

$$X_i(t+1) = \begin{cases} R(\overline{x}(t)) + \xi & \text{with probability } p_m p_r \\ R(\overline{x}(t)) & \text{with probability } (1-p_m)p_r \\ x_i(t) + \xi & \text{with probability } p_m(1-p_r) \\ x_i(t) & \text{with probability } (1-p_m)(1-p_r) \end{cases} \tag{1}$$

while for the alternative application of operators one obtains:

$$X_i(t+1) = \begin{cases} R(\overline{x}(t)) & \text{with probability } p_r \\ x_i(t) + \xi & \text{with probability } p_m = 1 - p_r \end{cases} \tag{2}$$

It is obvious that X_i are random variables. In the following, the random variables will be denoted by capital letters.

The recombination operator acts in two steps: (i) it selects $2 \leq \mu \leq m$ elements (called parents) from the current population; (ii) it combines them to obtain a new offspring. The parents are usually randomly selected from the population i.e. if the current population is $\{x_1, x_2, \ldots, x_m\}$, the parents set is a random selection $\{x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_\mu}\}$ with $(\alpha_l)_{l=\overline{1,\mu}}$ random variables with values in $\{1, 2, \ldots, m\}$.

The random selection of parents can be carried out in one of the two variants: with or without replacement. From the viewpoint of the arity of the recombination operator, the most frequentely used are those with $\mu = 2$ (binary recombination) and with $\mu = m$ (global recombination). Depending on the way the information of the parents are mixed, different variants of recombination has been proposed. Two of the most used variants are:

(a) intermediate recombination:

$$y^i = \sum_{l=1}^{\mu} c_l x^i_{\alpha_l}, \quad i = \overline{1,n}, \quad c_l \in (0,1), \quad \sum_{l=1}^{\mu} c_l = 1 \tag{3}$$

(b) discrete recombination:

$$y^i = \begin{cases} x^i_{\alpha_1} & \text{with probability } p_1 \\ \vdots & \\ x^i_{\alpha_\mu} & \text{with probability } p_\mu \end{cases} \qquad i = \overline{1,n}, \quad p_l \in (0,1), \quad \sum_{l=1}^{\mu} p_l = 1 \tag{4}$$

The main theoretical results have been obtained for global intermediate and discrete recombination with distinct parents [2]. Few results have been obtained regarding the behaviour of recombination for $2 \leq \mu < m$. Thus we will analyze discrete and intermediate recombination with $\mu < m$ and a particular case of binary intermediate recombination which combines an arbitrary element with the best element of the current population denoted by $x_*(t)$ $(f(x_*(t)) = \min_{x \text{ in } \overline{x}(t)} f(x))$. This particular recombination is characterized by $y^i = cx^i + (1-c)x^i_*, i = \overline{1,n}, c \in (0,1)$. Since this recombination selects one of the parents by using the value of the objective function, it realizes an implicit selection which favorizes the elements with low values of the objective function.

The theoretical study of the recombination operators follows the steps: (i) analyzes if the recombination type influences the convergence reliability of the algorithm; (ii) analyzes the influence of the recombination on the exploitation power of the algorithm measured through the progress rate.

3 Influence of the recombination on the convergence properties

Due to the stochastic character of variation operators and of their combination, the EA generates a Markov chain on $\mathcal{P}$ characterized by a stochastic kernel, $\overline{K}(\overline{x}, A)$, which describes the transition probability from a population $\overline{x} = (x_1, \ldots, x_m)$ to a set of populations $A \subset \mathcal{P}$. Since the transformation of each element is independent from the transformations of the others, the population transition kernel $\overline{K}(\overline{x}, A)$ can be decomposed as:

$$\overline{K}(\overline{x}, A) = \prod_{i=1}^{m-1} K^i(\overline{x}, A_i) \cdot 1_{A_m}(x_*), \quad A = \prod_{i=1}^{m} A_i, A_i \subset D \tag{5}$$

Since the transformations are similar for all elements we can consider that the individual transition kernel are identical, $K^i = K$. Moreover, depending on the way how the variation operators are combined, the individual kernel, K, can be decomposed as follows. In the case of sequential application of recombination and mutation:

$$K(\overline{x}, A_i) = p_m p_r \int_D K_R(\overline{x}, dy) K_M(y, A_i) + p_r(1-p_m) K_R(\overline{x}, A_i) + p_m(1-p_r) K_M(x_i, A_i) + \tag{6}$$

$$+(1-p_m)(1-p_r) 1_{A_i}(x_i).$$

In the case of alternative application of mutation and recombination:

$$K(\overline{x}, A_i) = p_r K_R(\overline{x}, A_i) + (1-p_r) K_M(x_i, A_i). \tag{7}$$

Both in (6) and (7) K_R and K_M are the transition kernels corresponding to recombination and mutation operators, respectively.

To analyze the convergence behaviour of the algorithm we will follow the technique and some of the results from [6]. Let $f^* = \min_{x \in D} f(x) = f(x^*) > -\infty$ the global minimal value and $f_*(\overline{x}(t)) =$

$\min_{x \text{ in } \overline{x}(t)} = f(x_*(t))$ the best value of the objective function found in population $\overline{x}(t)$. We also consider the notion of ϵ-vicinity of the global minimum defined by $A_\epsilon = \{\overline{x} \in \mathcal{P} | f_*(\overline{x}) - f^* < \epsilon\}$ which is the set of all populations which have at least the best objective function value at a distance less than ϵ from the global minimum value. Rudolph stated that the convergence reliability of an EA means ability of reaching an ϵ-vicinity of the global minimum. More specifically, an EA is convergent in probability if for any $\epsilon > 0$ the associated Markov chain converges in probability to the ϵ-vicinity A_ϵ.

Theorem 3.1 *Let us consider an arbitrary value $\epsilon > 0$ and the corresponding ϵ-vicinity $A_\epsilon = A_\epsilon^1 \times A_\epsilon^2 \times \ldots \times A_\epsilon^m$, $A_\epsilon^i \subset D$. If the individual mutation kernel K_M satisfies the property: there exists $\delta > 0$ such that $K_M(x_i, A_\epsilon^i) \geq \delta$ for $i = \overline{1,m}$ and if $p_m > 0$ then the algorithm is convergent in probability both for the sequential and alternative application of the recombination and mutation operators.*

Proof. We will use Theorem 1 from [6]. We have to verify that:

(i) $\overline{K}(\overline{x}, A_\epsilon) = 1$ if $\overline{x} \in A_\epsilon$;

(ii) there exists $\delta' > 0$ such that $\overline{K}(\overline{x}, A_\epsilon) \geq \delta'$ for all $\overline{x} \in A_\epsilon^c = \mathcal{P} \backslash A_\epsilon$.

Using (5) it follows that if $\overline{x} \in A_\epsilon$ then $\overline{K}(\overline{x}, A_\epsilon) = 1$, i.e. the hypothesis (i) is satisfied.

Let $\overline{x} \in A_\epsilon^c$. From (6) and (7) is easy to obtain that

$$K(\overline{x}, A_\epsilon^i) \geq p_m(1 - p_r) K_M(x_i, A_\epsilon^i) + p_m p_r \int_D K_R(\overline{x}, dy) K_M(y, A_\epsilon^i) \geq p_m \delta$$

and

$$K(\overline{x}, A_\epsilon^i) \geq (1 - p_r) K_M(x_i, A_i) \geq (1 - p_r)\delta = p_m \delta,$$

respectively.

Since the order of the population elements is not essential, recomposing the transition kernel using (5) it follows that $\overline{K}(\overline{x}, A_\epsilon) \geq (p_m \delta)^{m-1} = \delta' > 0$. Both hypothesis of Theorem 1 in [6] are satisfied, hence the algorithm is convergent.

∎

The key property which assures the convergence is some kind of ergodicity of the mutation. If the mutation is based on the addition of a random variable whose support covers D then the hypothesis (ii) is satisfied disregarding the recombination operator or the way the two operators are combined.

4 Influence of the recombination on the exploitative power

The exploitative power of an EA expresses the ability of the algorithm to step into the direction of desired improvement (which can be the local gradient or the global optimum direction) [3]. Stepping into a "good" direction means progress of the algorithm. The progress can be quantified by expected rates of change from one generation to the next. The rate of change can be measured either in the parameter space (relative progress rate [7]) or in the fitness space (expected fitness gain [2]).

For a given element $x_i(t)$ of generation t, the relative progress rate is the random variable $\rho(t) = 1 - \|X_i(t+1) - x^*\| / \|x_i(t) - x^*\|$. The fitness gain is defined as $\delta(t) = f_*(\overline{X}(t+1)) - f_*(\overline{x}(t))$ i.e. the difference between the best value of the new generation and the best value of the previous one.

If we have to compare the effectiveness of two variation operators with respect to their exploitative power a quantitative power could be: $\delta_r(t) = E(f_*(\overline{X}'(t+1) - f_*(\overline{X}''(t+1)) | \overline{X}(t) = \overline{x}(t))$ where $\overline{X}'(t+1)$ is the population obtained by applying the first type of variation operators and $\overline{X}''(t+1)$ is that obtained by applying the second type of variation operators. In both cases the previous generation is the same: $\overline{x}(t)$. If $\delta_r(t) \leq 0$ we will say that the first variation operator has a better local exploitative power than the second one.

To compute the fitness gain, the objective function must be known. Most of the analytical results for EAs have been obtained for the sphere model, $f(x) = \sum_{i=1}^n (x_i - x_i^*)^2$. This is the case we will analyze. Since the interest is on comparing the recombination operators we shall consider that the mutation operator is the same.

Let us consider $\overline{x}(t)$ be the current population and $(x_l)_{l=\overline{1,\mu}}$ a set of parents obtained through selection with replacement. For a given set of parents, the intermediate recombination (3) produces only one offspring while the discrete recombination (4) can produces μ^n offsprings.

Let y be the offspring obtained by intermediate recombination with $c_l = 1/\mu$, $l = \overline{1,\mu}$. Then $y^j = \dfrac{1}{\mu}\displaystyle\sum_{l=1}^{\mu} x_l^j$ and

$$E(f(y)) = f(y) = \sum_{j=1}^{n}\left(\frac{1}{\mu}\sum_{l=1}^{\mu} x_l^j - x^{*j}\right)^2 = \frac{1}{\mu^2}\sum_{j=1}^{n}((x_1^j - x^{*j}) + \ldots + (x_\mu^j - x^{*j}))^2.$$

Let Z be the random variable which represents the offspring generated by the discrete recombination. If $p_l = 1/\mu$, $l = \overline{1,\mu}$, Z can take μ^n values each one with the same probability $1/\mu^n$. Thus

$$E(f(Z)) = \frac{1}{\mu^n}\sum_{k=1}^{\mu^n}\sum_{j=1}^{n}(z_{(k)}^j - x^{*j})^2$$

with $z_{(k)}^j$ the j-th component of the k-th offspring. For each position $j \in \{1,\ldots,n\}$ there exist μ^{n-1} offsprings having the same value. Thus

$$E(f(Z)) = \frac{\mu^{n-1}}{\mu^n}\sum_{j=1}^{n}((x_1^j - x^{*j})^2 + \ldots + (x_\mu^j - x^{*j})^2).$$

The relative fitness gain is:

$$E(f(Z)) - f(y) = \frac{1}{\mu^2}\sum_{j=1}^{n}\left[(\mu-1)\sum_{l=1}^{\mu}(x_l^j - x^{*j})^2 - \sum_{l\neq k}(x_l^j - x^{*j})(x_k^j - x^{*j})\right] \geq 0.$$

Hence we obtained that for each given set of parents $f(y) \leq E(f(Z))$. Averaging over the all sets of parents one obtains $E(f_*(X'(t+1)) - f_*(X''(t+1))) \leq 0$ for $X'(t+1)$ obtained by intermediate recombination and $X''(t+1)$ obtained by discrete recombination. This result means that for the sphere model case the intermediate recombination ensures an improvement of the objective function greater than the discrete recombination. This suggests that in the case of a minimization problem, if the current population (or at least the set of parents) belongs to a region where the objective function is almost convex it is appropriate to use intermediate recombination while in regions where the objective function is concave it is better to use discrete recombination. These remarks suggest to adaptively choose the recombination type depending on the local properties of the objective functions.

It is easy to remark that for the sphere model the inclusion of the best element of the population in the parents set increases the improvement in the objective function value both in the case of intermediate and discrete recombination.

For a deeper analysis of the influence of recombination on the progress of the algorithm we shall consider the particular case of intermediate recombination with the best element x_*. We will estimate the relative progress rate (in the sense of Schwefel) for the case of a mutation operator based on a normally distributed additive perturbation and the specified recombination operator.

Let $\overline{x}(t)$ be the current population and $x(t)$ a fixed element different from $x_*(t)$. The new element is a random variable $X(t+1)$. We consider three cases: only mutation (X'), recombination and mutation applied alternatively (X'') and recombination and mutation applied sequentially (X''').

These three random variables are defined as follows:

$$X' = \begin{cases} x + \xi & \text{with probability } p_m \\ x & \text{with probability } 1 - p_m \end{cases} \qquad X'' = \begin{cases} cx + (1-c)x_* & \text{with probability } p_r \\ x + \xi & \text{with probability } 1 - p_r \end{cases}$$

$$X''' = \begin{cases} cx + (1-c)x_* + \xi & \text{with probability } p_r p_m \\ cx + (1-c)x_* & \text{with probability } p_r(1-p_m) \\ x + \xi & \text{with probability } p_m(1-p_r) \\ x & \text{with probability } (1-p_m)(1-p_r) \end{cases}$$

with $\xi = (\xi^1, \ldots, \xi^n)$, $\xi^i \sim N(0, \sigma^2)$. A relation between expected relative progress rates corresponding to these three variants is stated in the following proposition.

Proposition 4.1 *If n is great, $p_m = 1 - p_r$ and $\|x_* - x^*\| < \|x - x^*\|$ then $E(\rho') < E(\rho''') < E(\rho'')$.*

Before starting with the proof of Proposition we ennounce a Lemma:

Lemma 4.1 *Let $Z_1, \ldots, Z_k$ be bounded random variables which are independent with respect to any discrete random variable V having the distribution $P(V = v_j) = p_j$, $j = \overline{1,k}$, $p_j \in [0,1]$, $\sum_{j=1}^k p_j = 1$. Then the random variable*

$$Z = \begin{cases} Z_1 & \textit{with probability } p_1 \\ \vdots & \\ Z_k & \textit{with probability } p_k \end{cases}$$

has the mean $M(Z) = \sum_{j=1}^k p_j M(Z_j)$.

Proof of lemma. Let $Z_j^+ = \sup(Z_j, 0)$ and $Z_j^- = -\inf(Z_j, 0)$ be the positive part of Z_j and the negative part of Z_j, respectively. Then $Z_j = Z_j^+ - Z_j^-$. Moreover

$$Z^+ = \sup(Z,0) = \begin{cases} Z_1^+ & \text{with probability } p_1 \\ \vdots & \\ Z_k^+ & \text{with probability } p_k \end{cases} \qquad Z^- = -\inf(Z,0) = \begin{cases} Z_1^- & \text{with probability } p_1 \\ \vdots & \\ Z_k^- & \text{with probability } p_k \end{cases}$$

and $Z = Z^+ - Z^-$. Thus $M(Z) = M(Z^+) - M(Z^-)$ which means that it suffices to prove the lemma for positive random variables. Such a results has been proved in [8], thus the present lemma is proved. ■

Proof of proposition. First we will prove that $\|x_* - x^*\| < \|x - x^*\|$ implies $\|cx + (1-c)x_* - x^*\| < \|x - x^*\|$. Indeed $\|c(x - x^*) + (1-c)(x_* - x^*)\| \leq c\|x - x^*\| + (1-c)\|x_* - x^*\| < \|x - x^*\|$. In the following we use the approximation technique of Schwefel [7]. Let us denote $r = \|x - x^*\|$, $\overline{\sigma} = \sigma n / r$ (normalized standard deviation) and $\gamma = \|cx + (1-c)x_* - x^*\| / \|x - x^*\| < 1$. Using the result from [7] obtained for the case of mutation only it follows that ρ' can be approximated by a random variable with the structure:

$$\rho' \simeq \begin{cases} Z' \sim N\left(-\frac{\overline{\sigma}^2}{2n}, \frac{\overline{\sigma}^2}{n^2}\right) & \text{with probability } p_m \\ 0 & \text{with probability } 1 - p_m \end{cases}$$

Using Lemma it follows that $E(\rho') \simeq p_m\left(-\frac{\overline{\sigma}^2}{2n}\right)$. Since the recombination result is completely determined by x and x_* it immediately follows that $E(\rho'') \simeq (1 - p_r)\left(-\frac{\overline{\sigma}^2}{2n}\right) + p_r(1 - \gamma)$.

To compute $E(\rho''')$ it is necessary to estimate the distribution of $r_1(t+1) = \|cx + (1-c)x_* + \xi - x^*\|$. Using the properties of normal distribution when n is great we obtain that $r_1(t+1)$ can be approximated by a random variable having the expression $\sqrt{r'^2 + n\sigma^2 + \sqrt{2n\sigma^4 + 4\sigma^2 r'^2} Z}$ with $r' = \|cx + (1-c)x_* - x^*\|$ and $Z \sim N(0,1)$.

We estimate the progress rate component corresponding to the case when both recombination and mutation are applied:

$$\rho_1''' = 1 - \frac{\|cx + (1-c)x_* + \xi - x^*\|}{\|x - x^*\|} \simeq 1 - \sqrt{\left(\frac{r'}{r}\right)^2 + \frac{n\sigma^2}{r^2} + \sqrt{\frac{2n\sigma^4}{r^4} + \frac{4\sigma^2 r'^2}{r^4}} Z} =$$

$$= 1 - \sqrt{\gamma^2 + \frac{\overline{\sigma}^2}{n} + \sqrt{\frac{2}{n^3}\overline{\sigma}^4 + \frac{4}{n^2}\overline{\sigma}^2\gamma^2} Z} = 1 - \sqrt{\gamma^2 + \frac{\overline{\sigma}^2}{n} + \frac{\overline{\sigma}^2}{n}\sqrt{\frac{4\gamma^2}{\overline{\sigma}^2} + \frac{2}{n}} Z}.$$

Since n is great, the term $2/n$ can be neglected. Thus ρ''' can be approximated by the random variable:

$$1 - \sqrt{\gamma^2 + \frac{\overline{\sigma}^2}{n} + \frac{2\gamma}{n}\overline{\sigma} Z} = 1 - \sqrt{1 + (\gamma^2 - 1 + \frac{\overline{\sigma}^2}{n} + \frac{2\gamma}{n}\overline{\sigma} Z)}.$$

Using the first order Taylor expansion of $\sqrt{1+x}$ in 0 it follows that

$$\rho_1''' \simeq 1 - \left(1 + \frac{\gamma^2 - 1}{2} + \frac{\overline{\sigma}^2}{2n} + \frac{\gamma}{n}\overline{\sigma} Z\right) = \frac{1 - \gamma^2}{2} - \frac{\overline{\sigma}^2}{2n} - \frac{\gamma}{n}\overline{\sigma} Z.$$

Thus ρ_1''' has a normal distribution with the mean $\dfrac{1-\gamma^2}{2} - \dfrac{\overline{\sigma}^2}{2n}$ and variance $\dfrac{\gamma^2}{n^2}\overline{\sigma}^2$.

Using Lemma it follows that

$$E(\rho''') = p_r p_m \left(-\frac{\overline{\sigma}^2}{2n} + \frac{1-\gamma^2}{2}\right) + p_r(1-p_m)(1-\gamma) + p_m(1-p_r)\left(-\frac{\overline{\sigma}^2}{2n}\right) =$$

$$= p_m\left(-\frac{\overline{\sigma}^2}{2n}\right) + p_r p_m \frac{1-\gamma^2}{2} + p_r(1-p_m)(1-\gamma).$$

If $p_m = p_r$ then we have:

$$E(\rho') \simeq p_m\left(-\frac{\overline{\sigma}^2}{2n}\right) \qquad E(\rho'') \simeq p_m\left(-\frac{\overline{\sigma}^2}{2n}\right) + (1-p_m)(1-\gamma)$$

$$E(\rho''') \simeq p_m\left(-\frac{\overline{\sigma}^2}{2n}\right) + p_m(1-p_m)\frac{1-\gamma^2}{2} + (1-p_m)^2(1-\gamma).$$

Since $\gamma < 1$ it is obvious that $E(\rho') < E(\rho'')$ and $E(\rho') < E(\rho''')$. On the other hand, $E(\rho''') - E(\rho'') = p_m(1-p_m)(1-\gamma)\left(\frac{1+\gamma}{2} - 1\right) < 0$.

■

The hypothesis $\|x_* - x^*\| < \|x - x^*\|$ for $x \neq x_*$ (meaning that the best element of the population is closer to the global minimum than all other elements of the population) is satisfied for all x if $f(x) = \|x - x^*\|^2$ or $f(x) = g(\|x - x^*\|^2)$ with g a strictly increasing function. Moreover, the hypothesis could be satisfied for some x even for objective functions which are deterministic or stochastic perturbations of the above models. For example if $f(x) = g(\|x - x^*\|^2) + p(x)$ the hypothesis is satisfied at least for those x of population which satisfy $p(x_*) > p(x)$.

The above Proposition states that under some plausible hypothesis the algorithm with the proposed recombination operator has a relative progress rate greater that of the algorithm which uses only mutation. Moreover, if the mutation and recombination probabilities verify $p_m + p_r = 1$ then the alternate application of operators assure a relative progress rate higher than that corresponding to the sequential application of the operators.

5 Experimental results

The theoretical results presented above suggests that the recombination operators can not be considered as universally applicable. For one type of objective functions one recombination operator can be better than another while for another type of objective functions the situation could be reversed. This generates a natural idea: to let the algorithm to choose the appropriate recombination operator based on the local improvement which it assures.

The aim of the experimental analysis was to compare the behaviour of five recombination operators: binary intermediate recombination (with an arbitrary (*Intermediate*) and with the best element (**Intermediate*) respectively), binary discrete recombination (with an arbitrary (*Discrete*)and with the best element (**Discrete*) respectively) and a hybrid variant which selects one of the four previous ones based on the current behaviour (is selected the operator which ensures the greatest improvement of the objective function).

The experiments have been carried out using an algorithm with alternate application of mutation and recombination with $p_r = 0.5$ (between the two variants of operators combination have not been remarked significant differences through numerical experiments). The mutation operator consists in adding to each component a normally distributed random variable with zero mean and a dispersion $\sigma^2(t)$, variable in time. The time dependence which we used was $\sigma(t) = \sigma_0/\ln(t+1)$. The stopping criterion is: "either 10000 generations have been produced or $|f_* - f^*| < 10^{-3}$".

The test functions which we used were some classical ones having many local minima:
(i) (Rastrigin) $R(x_1,\ldots,x_n) = 0.2\sum_{i=1}^n x_i^2 - 2(\cos(10\sum_{i=1}^n x_i) - 1)$, $x_i \in [-5.12, 5.12]$, $x^* = (0,\ldots,0)$, $R^* = 0$.
(ii) (Griewank) $G(x_1,\ldots,x_n) = \frac{1}{400n}\sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos(x_i/\sqrt{i}) + 1$, $x_i \in [-400, 400]$, $x^* = (0,\ldots,0)$, $G^* = 0$.

(iii) (Ackley) $A(x_1,\ldots,x_n) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + e$, $x_i \in [-32,32]$, $x^* = (0,\ldots,0)$, $A^* = 0$.

In Table 1 are presented results obtained in 10 independent runs: columns R_*, A_* and G_* contain the best value obtained in the last generation, while the columns $E(R_*)$, $E(A_*)$ and $E(G_*)$ contain the average over the 10 runs of the best value in the last generation. The results preceded by a * are in the ϵ-vicinity of the global minimum (for $\epsilon = 10^{-3}$). To assure a fair comparison, for the first four recombination types, the population size was $m = 30$ while for the hybrid variant we used $m = 9$.

The experimental results suggests that the hybrid variant has a good behaviour for all test functions because it uses the exploitative power of each of the other implemented recombination operators.

Recombination	R_*	$E(R_*)$	A_*	$E(A_*)$	G_*	$E(G_*)$
Intermediate	2.9707	10.8636	0.0327	0.0733	0.6091	0.7996
*Intermediate	7.9597	18.1082	*0.0091	0.4709	0.0842	0.2109
Discrete	0.0161	0.9930	0.0207	0.0466	0.0865	0.3053
*Discrete	*0.0025	*0.0072	*0.0068	*0.0085	0.0223	0.0504
Hybrid	*0.0037	*0.0077	*0.0069	*0.0092	*0.0071	*0.0155

Table 1. Results for the test functions R, A and G

Acknowledgement. This work is supported by Ministry of Education of Romania under Grant 3525/2000.

References

[1] Bäck, T., Hoffmeister,F., Schwefel, H.P. (1992) Applications of Evolutionary Algorithms, *Technical Report SYS-2/92*, University of Dortmund.

[2] Beyer, H.G. (1995) Towards a Theory of Evolution Strategies: on the benefits of sex $(\mu/\mu,\lambda)$ theory, *Evolutionary Computation*, **3** (1), 81-111.

[3] Beyer, H.G. (1999) On the Dynamics of EAs without Selection, in *Foundations of Genetic Algorithms*, W. Banzaf and C. Reeves, eds., Morgan Kaufmann, 5–26.

[4] Beyer, H.G. and Deb, K. (1999) On the Analysis of Self Adaptive Evolutionary Algorithms, *Technical Report CI-69/99*, University of Dortmund.

[5] François, O. (1998) An Evolutionary Strategy for Global Minimization and its Markov Chain Analysis, *IEEE Trans. on Evolutionary Computation*, **2**, 77-90.

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

[7] Schwefel, H.P. and Bäck, T. (1995) Evolution Strategies II: Theoretical Aspects, in *Genetic Algorithms in Engineering and Computer Science*, J. Périaux et al., eds., J. Wiley & Sons, 127 – 140.

[8] Zaharie, D. (2000) On an Evolutionary Algorithm for Global Optimization, *Analele Univ. Timişoara*, seria Matematică-Informatică, **XXXVIII** (2), 203–216.