

C. Hervás-Martínez · D. Ortiz-Boyer

Analyzing the statistical features of CIXL2 crossover offspring

Published online: 7 June 2004
© Springer-Verlag 2004

Abstract We cannot deny the effort that the scientific community is devoting to the explanation of the features of the crossover operator of real-coded genetic algorithms and its effect over the evolutive process. This paper is another step in that direction, we analyze the behavior of the Confidence Interval Based Crossover using L_2 Norm (CIXL2). This crossover is based on the learning of the statistical features of localization and dispersion of the best individuals of the population. The crossover obtains, by means of a L_2 norm, the estimators of the parameters of localization and dispersion of the distributions of the fittest individuals. From this estimation three virtual parents are created using the localization parameter and the lower and upper bounds of the bilateral confidence intervals of the gene values of the best individuals of the population. This paper studies the statistical features of the offspring generated by this crossover and corroborates this study showing the behavior of the crossover in a set of test functions.

Keywords Real-coded genetic algorithms · Statistical features learning · Confidence intervals · L_2 -norm

1 Introduction

From the beginnings of the evolutionary computation, many papers have been centered in the study of crossover operator. This operator is the most innovative, as the mutation operator inherits many of the features of the operators of random search methods [1]. The crossover operator combines the features of two or more individuals to create a, possibly better, offspring. The operator is based on the idea that the combination of good individuals will produce better individuals.

Initially, the theoretical studies were developed for binary crossovers from the schema theory [13]. Nevertheless, genetic algorithms (GAs) that use binary coding find certain difficulties when they are used in multidimensional continuous search spaces where a great numerical precision is required [17]. That is the reason why in problems of optimization in continuous domains real-coded genetic algorithms (RCGAs) are used [3, 7, 9, 11, 18, 23]. The first studies that justified, in a formal way, the power of RCGAs appeared in 1991 [8, 22, 23].

Most of the studies developed for crossover of real-coded individuals are based on empirical results, however there are a few theoretical studies worth to mention. In [16] it is proved that the distribution of the offspring generated by the unimodal normal distribution crossover (UNDX) keeps the vector mean and covariance matrix of the distribution of the parents. This is called *statistics preservation*. In [2] it is made a theoretical study of the self-adapted features of the crossover operator that is centered in statistical features of the population such as the mean and the standard deviation of the population. Its theoretical conclusions are validated in the experimental phase. There are also other works with similar features [5, 15].

This paper is focused on the analysis of the confidence interval based crossover using L_2 norm (CIXL2), whose interesting performance in problems in continuous domains has been established in previous works [12, 20]. Firstly, we will carry out a theoretical study of the behavior of the mean and variance of the distribution of the offspring with regard to the distribution of its parents. Secondly, we will corroborate the theoretical conclusions in practice with several experiments of optimization of real valued functions with different features.

The remainder of this paper is organized as follows. In Sect. 2 we present the definition of the CIXL2 crossover operator, and in Sect. 3 we made a theoretical study of its behavior. The theoretical conclusions are verified in the experimental stage whose results are

C. Hervás-Martínez (✉) · D. Ortiz-Boyer
Univ. of Córdoba Campus de Rabanales,
Dept. Computer Science,
Edificio Albert Einstein 3a planta,
14071 Córdoba, Spain
E-mail: chervas@uco.es

shown in Sect. 4. Finally, Sect. 5 states the final conclusions of our work.

2 Confidence interval based crossover

Let β be the set of N individuals that form the population, $\beta^* \subset \beta$ the subset of the n fittest individuals, and q the number of genes on each chromosome. Let us assume that the genes, β_i , of the chromosomes of the individuals in β^* are independent random variables with a normal distribution and with a localization parameter μ_{β_i} . Then we have the model $\beta_i = \mu_{\beta_i} + e_i$, for all $i=1, \dots, q$, being e_i a random variable.

If we assume that the n fittest individuals form actually a simple random sample $(\beta_{i_1}, \beta_{i_2}, \dots, \beta_{i_n})$ of the normal distribution of the fittest individuals of the population β_i^b , the model can be written:

$$\beta_{ij}^b = \mu_{\beta_i^b} + e_{ij}, \quad \text{for } j = 1, 2, \dots, n \text{ and } i = 1, \dots, q. \quad (1)$$

Let us consider the L_2 norm, defined as $\|\beta_i^b\|_2 = \sqrt{\sum_{j=1}^n (\beta_{ij}^b)^2}$, then we can define the dispersion function, D_2 , induced by the L_2 norm as: $D_2(\mu_{\beta_i^b}) = \sqrt{\sum_{j=1}^n (\beta_{ij}^b - \mu_{\beta_i^b})^2}$. The estimator, using this dispersion function, of the localization parameter is the mean of the distribution β_i^b , that is, $\bar{\beta}_i^b = \hat{\mu}_{\beta_i^b}$.

The sample mean estimator is a linear estimator, so it has the properties of unbiasedness and consistency, and when the distribution of the genes is normal, it follows a normal distribution $N(\mu_{\beta_i^b}, \sigma_{\beta_i^b}^2/n)$. Under these assumptions we can construct a bilateral confidence interval for the localization of the genes using the sample mean as localization parameter. This confidence interval, for a confidence coefficient $1 - \alpha$, has the form:

$$I_{1-\alpha}(\mu_{\beta_i^b}) = \left[\bar{\beta}_i^b - t_{n-1, \alpha/2} \frac{S_{\beta_i^b}}{\sqrt{n}}; \bar{\beta}_i^b + t_{n-1, \alpha/2} \frac{S_{\beta_i^b}}{\sqrt{n}} \right], \quad (2)$$

where $S_b = [\sum_{j=1}^n (\beta_{ij}^b - \bar{\beta}_i^b)^2 / (n-1)]^{1/2}$ is the standard deviation, and t_{n-1} is a Student t of $n-1$ degrees of freedom.

2.1 Hypothesis of normality

In all the above framework, we have assumed that the values of the genes of the best individuals are normally distributed. With that hypothesis we can use 2 for obtaining the confidence interval of every gene.

However, due to the special features of the evolutionary process, it is plausible that the hypothesis of normality would not be fulfilled in all cases. Even when the hypothesis is fulfilled, its degree of fulfillment may vary depending on the problem or the evolutionary process. If we initialize the values following and uniform distribution, the most common situation, in the first stages of the evolution the

distributions of the genes will have long tails that will tend to diminish along the evolutionary process. Eventually, the initial distribution will produce a distribution with very short tails and almost all the individuals will be concentrated in a small interval within the search space. It is also possible to have a distribution with more than one peak when the population is moving from a local minimum to another local minimum or when the population is distributed among several local minima.

In this scenario, we have decided to assume an a priori distribution of the genes of the best individuals of the population, accepting that any hypothesis on the distribution would not be fulfilled up to the same degree during all the evolutionary process. In our case it would be advisable to choose a distribution where the amplitude of the interval is big enough to contain the confidence intervals of as many distributions as possible. In this way we guarantee that at least the $100(1 - \alpha)\%$ of the values of the genes are within the confidence interval. So, we will choose the distribution whose value of $t_{n-1}\alpha/2$ is the maximum or, at least, is near the maximum. With certain restrictions, the normal distribution is the distribution that best reflects these features. That is the reason why we have chosen the normal distribution to model the distribution of the values of the genes of the best individuals of the population.

2.2 Crossover operator method

From the confidence interval of (2) we build three individuals that are considered the parents in the proposed crossover. These three parents are formed by: all the lower limit values of the confidence intervals of the genes, individual CILL; all the upper limit values of the confidence intervals of the genes, individual CIUL; and all the means of the confidence intervals of the genes, individual CIM. These individuals divide the domain of the gene values, D_i , into three subintervals: $I_i^L \equiv [a_i, CILL_i]$, $I_i^{CI} \equiv [CILL_i, CIUL_i]$, and $I_i^R \equiv [CIUL_i, b_i]$, being a_i and b_i the bounds of the search domain.

The interval based crossover operator using L_2 norm, CIXL2, creates an offspring β^s , from an individual of the population, $\beta^f = (\beta_1^f, \beta_2^f, \dots, \beta_p^f)$, and the three individuals, CILL, CIUL, and CIM, obtained from the confidence interval. We consider these four individuals and their fitness (being $f(\beta)$ the fitness value of individual β) and distinguish three cases depending on the position of β^f in one of the three subintervals defined above. These three cases are:

- Case 1: $\beta_i^f \in I_i^{CI}$. If $f(\beta^f) \geq f(CIM)$ then $\beta_i^s = r(\beta_i^f - CIM_i) + \beta_i^f$ else $\beta_i^s = r(CIM_i - \beta_i^f) + CIM_i$.
- Case 2: $\beta_i^f \in I_i^L$. If $f(\beta^f) \geq f(CILL)$ then $\beta_i^s = r(\beta_i^f - CILL_i) + \beta_i^f$ else $\beta_i^s = r(CILL_i - \beta_i^f) + CILL_i$.
- Case 3: $\beta_i^f \in I_i^R$. If $f(\beta^f) \geq f(CIUL)$ then $\beta_i^s = r(\beta_i^f - CIUL_i) + \beta_i^f$ else $\beta_i^s = r(CIUL_i - \beta_i^f) + CIUL_i$ (Fig. 1).

where r is a uniform random number belonging to $[0, 1]$.

3 Theoretical analysis of CIXL2 crossover

This analysis is centered in determining the behavior of the mean and the standard deviation of the distribution of the offspring with regard to the distribution of its parents. As the operator distinguishes three different cases we must make a separate analysis for each one.

3.1 Case 1

Let us assume that the gene involved in the crossover, β_i^f , is within the confidence interval, $\beta_i^f \in I_i^{CI}$. We consider that in the j -th generation $\beta_i^f \in N(\mu, \sigma^2)$, and that, provided that the population is large enough, the sample of the fittest individuals of the population is also normally distributed, $\beta_i^b \in N(\mu_{\beta_i^b}, \sigma_{\beta_i^b}^2)$; and that the two distributions are independent. With these assumptions the distribution of the mean of the confidence interval, CIM_i , follows a normal distribution, $CIM_i = \beta_i^b \in N(\mu_{\beta_i^b}, \sigma_{\beta_i^b}^2/n)$.

We have two possible events, $f(\beta^f) \geq f(CIM)$ with a probability p , and $f(\beta^f) < f(CIM)$ with a probability $1 - p$. In the first event the distribution of the offspring is normal, as it is a linear combination of normal distributions, $\beta_i^s = (1 + r)\beta_i^f - rCIM_i$:

$$\beta_i^s \in N\left((1 + r)\mu - r\mu_{\beta_i^b}, (1 + r)^2\sigma^2 + r^2\frac{\sigma_{\beta_i^b}^2}{n}\right). \quad (3)$$

In general, it can be expected that at the beginning of the evolution $\sigma^2 \geq \sigma_{\beta_i^b}^2$ and $\mu \neq \mu_{\beta_i^b}$, due to the fact that the subpopulation of the n fittest individuals is a subset of the whole population. Along the evolution, due to the selection pressure, $\mu \rightarrow \mu_{\beta_i^b}$ and $\sigma^2 \rightarrow \sigma_{\beta_i^b}^2$, yielding:

$$\beta_i^s \rightarrow N\left(\mu_{\beta_i^b}, \sigma_{\beta_i^b}^2\left((1 + r)^2 + \frac{r^2}{n}\right)\right). \quad (4)$$

In the second event, $f(\beta^f) < f(CIM)$, the distribution of the offspring is normal, as it is a linear combination of normal distributions, $\beta_i^s = (1 + r)CIM_i - r\beta_i^f$:

$$\beta_i^s \in N\left((1 + r)\mu_{\beta_i^b} - r\mu, (1 + r)^2\frac{\sigma_{\beta_i^b}^2}{n} + r^2\sigma^2\right), \quad (5)$$

and, following the same reasoning above:

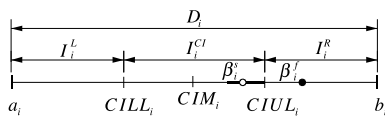


Fig. 1 Graphic representation of case 3 of the crossover based on confidence intervals

$$\beta_i^s \rightarrow N\left(\mu_{\beta_i^b}, \sigma_{\beta_i^b}^2\left(\frac{(1 + r)^2}{n} + r^2\right)\right). \quad (6)$$

So, the distribution of the generated offspring by means of this crossover will be a mixture of normal distributions, and so a normal itself, of mean:

$$E(\beta_i^s) = p\mu_{\beta_i^b} + (1 - p)\mu_{\beta_i^b} = \mu_{\beta_i^b}, \quad (7)$$

and variance:

$$\begin{aligned} V(\beta_i^s) &= p^2\sigma_{\beta_i^b}^2\left((1 + r)^2 + \frac{r^2}{n}\right) \\ &\quad + (1 - p)^2\sigma_{\beta_i^b}^2\left(\frac{(1 + r)^2}{n} + r^2\right) \\ &= V(\beta_i^b)\left[p^2\left((1 + r)^2 + \frac{r^2}{n}\right) \right. \\ &\quad \left. + (1 - p)^2\left(\frac{(1 + r)^2}{n} + r^2\right)\right]. \end{aligned} \quad (8)$$

We can conclude that, if $\beta_i^f \in I_i^{CI}$, the distribution of the offspring obtained using a CIXL2 crossover is normal. The mean is the same as the mean of the fittest individuals, and the variance depends on the choice of the parameter n , the random value r and the fitness landscape that will determine the value of p .

3.1.1 Analysis of the variance

In this section we will analyze the relation among n , r and p , and the influence of these three values over the variance of the offspring. We can distinguish three cases:

a) If $p = P(f(\beta^f) \geq f(CIM)) = 0$, substituting in (8)

$$V(\beta_i^s) = \left(\frac{(1 + r)^2}{n} + r^2\right)V(\beta_i^b) \quad (9)$$

so, if $n \geq \frac{1+r}{1-r}$ then $V(\beta_i^s) \leq V(\beta_i^b)$, that is, the variance of the offspring is less or equal than the variance of the best individuals of the population in the previous generation.

If $r = 0$ then $n \geq 1$, and if $r = 1$ then $n \rightarrow \infty$. Considering that r is a random value in the interval $[0, 1]$ the probability that $V(\beta_i^s) \leq V(\beta_i^b)$ will depend on n and r .

b) If $p = P(f(\beta^f) \geq f(CIM)) = 1$, following the same reasoning, if $n \geq -\frac{r}{2+r}$ then $V(\beta_i^s) \leq V(\beta_i^b)$.

If $r = 0$ then $n \geq 0$, and if $r = 1$ then $n \geq -1/3$. So, for any value of r and n we will always have $V(\beta_i^s) \leq V(\beta_i^b)$.

c) If $p = P(f(\beta^f) \geq f(CIM)) = 1/2$, if $n \geq \frac{1+2r+2r^2}{3-2r-2r^2}$ then $V(\beta_i^s) \leq V(\beta_i^b)$. If $r = 0$ then $n \geq 1/3$, and if $r = 1$ then $n \geq -5$. So the condition $V(\beta_i^s) \leq V(\beta_i^b)$ will be met for any value of r and n .

Case a), for certain values of n , favors the dispersion of the offspring around the mean of the best individuals, propitiating the avoidance of local minima. However, cases b) and c) will narrow the search to a reduced area

around the optimum favoring the precision of the solution and the convergence rate.

3.2 Case 2

We assume that the gene involved in the crossover, β_i^f , is outside the confidence interval, $\beta_i^f \in I_i^L$. With the same hypothesis of the previous case, we have that $\beta_i^f \in N(\mu, \sigma^2)$, $\beta_i^b \in N(\mu_{\beta_i^b}, \sigma_{\beta_i^b}^2)$ and that the two distributions are independent. The mean and variance of the lower bound of the confidence interval, $CILL_i$, is

$$E(CILL_i) = E\left(\bar{\beta}_i^b - t_{n-1, \alpha/2} \frac{S_{\beta_i^b}}{\sqrt{n}}\right) \\ = \mu_{\beta_i^b} - t_{n-1, \alpha/2} \frac{E(S_{\beta_i^b})}{\sqrt{n}} \quad (10)$$

As the distribution of the statistic $\frac{(n-1)S_{\beta_i^b}^2}{\sigma_{\beta_i^b}^2}$, under the hypothesis of normality, follows a distribution χ_{n-1}^2 , then

$$E(S_{\beta_i^b}^2) = \sigma_{\beta_i^b}^2 \quad (11)$$

so, ignoring terms of order $n^{\frac{1}{2}}$ [14]

$$E(S_{\beta_i^b}) = \sigma_{\beta_i^b} \quad (12)$$

Substituting this result in (10)

$$E(CILL_i) = \mu_{\beta_i^b} - t_{n-1, \alpha/2} \frac{\sigma_{\beta_i^b}}{\sqrt{n}} \quad (13)$$

$\bar{\beta}_i^b$ and $S_{\beta_i^b}$ are independent random variables and the variance of $CILL_i$ is obtained from

$$V(CILL_i) = V(\bar{\beta}_i^b) + t_{n-1, \alpha/2}^2 \frac{V(S_{\beta_i^b})}{n} \quad (14)$$

If we consider that

$$V(\beta_i^b) = \frac{\sigma_{\beta_i^b}^2}{n}; \\ V(S_{\beta_i^b}^2) = 2(n-1) \frac{\sigma_{\beta_i^b}^4}{(n-1)^2} = \frac{2\sigma_{\beta_i^b}^4}{n-1}, \quad (15)$$

and

$$V(S_{\beta_i^b}) = \frac{V(S_{\beta_i^b}^2)}{4\sigma_{\beta_i^b}^2}, \quad (16)$$

then

$$V(S_{\beta_i^b}) = \frac{\sigma_{\beta_i^b}^2}{2(n-1)}. \quad (17)$$

Substituting in (14)

$$V(CILL_i) = \frac{\sigma_{\beta_i^b}^2}{n} + t_{n-1, \alpha/2}^2 \frac{\sigma_{\beta_i^b}^2}{2n(n-1)}. \quad (18)$$

Depending on the crossover dynamic we have two possible events, $f(\beta^f) \geq f(CILL)$ with a probability p , and $f(\beta^f) < f(CIM)$ with a probability $1 - p$. In the first

case we have $\beta_i^s = (1+r)\beta_i^f - rCILL_i$ and the mean of the distribution of the offspring is

$$E(\beta_i^s) = (1+r)E(\beta_i^f) - rE(CILL_i) \quad (19)$$

and the variance is

$$V(\beta_i^s) = (1+r)^2V(\beta_i^f) + r^2V(CILL_i). \quad (20)$$

Using the equations (19) y (20) and considering a high selective pressure, $E(\beta_i^f) \rightarrow E(CILL_i)$ and $V(\beta_i^f) \rightarrow V(CILL_i)$, we have

$$E(\beta_i^s) = E(\beta_i^f) = E(CILL_i) = \mu_{\beta_i^b} - t_{n-1, \alpha/2} \frac{\sigma_{\beta_i^b}}{\sqrt{n}} \quad (21)$$

and

$$V(\beta_i^s) = V(CILL_i)(1+2r+2r^2) \\ = \left(\frac{\sigma_{\beta_i^b}^2}{n} + t_{n-1, \alpha/2}^2 \frac{\sigma_{\beta_i^b}^2}{2n(n-1)}\right)(1+2r+2r^2) \\ = \sigma_{\beta_i^b}^2 \left(\frac{2(n-1) + t_{n-1, \alpha/2}^2}{2n(n-1)}\right)(1+2r+2r^2) \quad (22)$$

In the second case, $f(\beta^f) < f(CILL)$ with probability $1 - p$, and being $\beta_i^s = (1+r)CILL_i - r\beta_i^f$, the mean and variance of the distribution of the offspring are

$$E(\beta_i^s) = E(CILL_i) = \mu_{\beta_i^b} - t_{n-1, \alpha/2} \frac{\sigma_{\beta_i^b}}{\sqrt{n}} \quad (23)$$

and

$$V(\beta_i^s) = \sigma_{\beta_i^b}^2 \left(\frac{2(n-1) + t_{n-1, \alpha/2}^2}{2n(n-1)}\right)(1+2r+2r^2) \quad (24)$$

We have, as in case 1, the linear combination of two distributions, its mean is given by

$$E(\beta_i^s) = pE(CILL_i) + (1-p)E(CILL_i) = E(CILL_i) \\ = \mu_{\beta_i^b} - t_{n-1, \alpha/2} \frac{\sigma_{\beta_i^b}}{\sqrt{n}} \quad (25)$$

and its variance

$$V(\beta_i^s) = p^2\sigma_{\beta_i^b}^2 \left(\frac{2(n-1) + t_{n-1, \alpha/2}^2}{2n(n-1)}\right) \\ \times (1+2r+2r^2) \\ + (1-p)^2\sigma_{\beta_i^b}^2 \left(\frac{2(n-1) + t_{n-1, \alpha/2}^2}{2n(n-1)}\right) \\ \times (1+2r+2r^2) \\ = V(\beta_i^b) \left(\frac{2n + t_{n-1, \alpha/2}^2}{2n^2}\right) \\ \times (1+2r+2r^2)(1-2p+2p^2) \quad (26)$$

We can conclude that if $\beta_i^f \in I_i^L$, the mean of the distribution of the offspring obtained using a CIXL2 crossover is , if $n \rightarrow \infty$ or $1 - \alpha \rightarrow 0$ or $\sigma_{\beta_i^b} \rightarrow 0$, $E(\beta_i^s) = \mu_{\beta_i^b}$. So, $E(\beta_i^s) \leq \mu_{\beta_i^b}$ and the difference among

the means will be larger if the values of n are small or the values of $1 - \alpha$ are large. This effect is specially significant in the first stages of the evolution when $V(\beta_i^b)$ is very large. The variance will depend on the choice of n and $1 - \alpha$, that will set the value of $t_{n-1, \alpha/2}$, the random quantity r and the fitness landscape that will determine p . The effect of these parameters will be analyzed in the next section.

3.2.1 Analysis of the variance

In this section we analyze the relation among $1 - \alpha$, r and p and their influence over the variance of the offspring. We can distinguish three cases:

a) and b) If $p = P(f(\beta^f) \geq f(CILL))$ is 0 or 1 then

$$V(\beta_i^s) = V(\beta_i^b) \left(\frac{2(n-1) + t_{n-1, \alpha/2}^2}{2n(n-1)} \right) (1 + 2r + 2r^2) \quad (27)$$

so, if

$$\left(\frac{2(n-1) + t_{n-1, \alpha/2}^2}{2n(n-1)} \right) (1 + 2r + 2r^2) \leq 1 \quad (28)$$

then $V(\beta_i^s) \leq V(\beta_i^b)$.

If we suppose that $r = 0$ the condition will be

$$\frac{2(n-1) + t_{n-1, \alpha/2}^2}{2n(n-1)} \leq 1 \quad (29)$$

which is fulfilled if

$$n \geq 1 + \frac{t_{n-1, \alpha/2}^2}{\sqrt{2}} \quad (30)$$

This last condition is true for the pairs $\{(1 - \alpha \leq 0.99, n > 4), (1 - \alpha \leq 0.70, n > 2), \dots\}$. If we suppose $r = 1$ the condition is

$$\frac{10(n-1) + 5t_{n-1, \alpha/2}^2}{2n(n-1)} \leq 1 \quad (31)$$

which is true when

$$n \geq 3 + \sqrt{4 + \frac{5}{2} t_{n-1, \alpha/2}^2} \quad (32)$$

This condition is true for the pairs $\{(1 - \alpha \leq 0.99, n > 8), (1 - \alpha \leq 0.70, n > 5), \dots\}$. Considering that r is a random number within the interval $[0, 1]$, for $V(\beta_i^s) \leq V(\beta_i^b)$ we must take pairs fulfilling $\{(1 - \alpha \leq 0.99, n > 8), (1 - \alpha \leq 0.70, n > 5), \dots\}$, and for $V(\beta_i^s) > V(\beta_i^b)$ the pair must fulfill $\{(1 - \alpha \geq 0.99, n \leq 4), (1 - \alpha \geq 0.90, n \leq 2), \dots\}$, in other case it will depend on the value of r .

c) If $p = P(f(\beta^f) \geq f(CIM)) = 1/2$ then

$$V(\beta_i^s) = \frac{1}{2} V(\beta_i^b) \left(\frac{2(n-1) + t_{n-1, \alpha/2}^2}{2n(n-1)} \right) \times (1 + 2r + 2r^2) \quad (33)$$

if the condition

$$\left(\frac{2(n-1) + t_{n-1, \alpha/2}^2}{4n(n-1)} \right) (1 + 2r + 2r^2) \leq 1 \quad (34)$$

is met, then $V(\beta_i^s) \leq V(\beta_i^b)$. For $r = 0$ the condition is reduced to

$$\frac{2(n-1) + t_{n-1, \alpha/2}^2}{4n(n-1)} \leq 1 \quad (35)$$

that it is fulfilled if

$$n \geq \frac{3}{4} + \frac{\sqrt{1 + t_{n-1, \alpha/2}^2}}{4} \quad (36)$$

that will be fulfilled for the pairs $\{(1 - \alpha \leq 0.99, n > 3), (1 - \alpha \leq 0.70, n > 1), \dots\}$. For $r = 1$ the condition

$$\frac{10(n-1) + 5t_{n-1, \alpha/2}^2}{4n(n-1)} \leq 1 \quad (37)$$

must be met. This condition if true when

$$n \geq \frac{7 + \sqrt{9 + t_{n-1, \alpha/2}^2}}{4} \quad (38)$$

which is fulfilled for the pairs $\{(1 - \alpha \leq 0.99, n > 8), (1 - \alpha \leq 0.70, n > 7), \dots\}$. Considering that r is a random number in the interval $[0, 1]$, for $V(\beta_i^s) \leq V(\beta_i^b)$ we must take pairs that fulfill $\{(1 - \alpha \leq 0.99, n > 8), (1 - \alpha \leq 0.70, n > 7), \dots\}$ and for $V(\beta_i^s) > V(\beta_i^b)$ we must take pairs that verify $\{(1 - \alpha \geq 0.99, n \leq 3), (1 - \alpha \geq 0.89, n \leq 2), \dots\}$, in other case it will depend on the value of r .

As we can see, if $\beta_i^f \in I_i^L$ the mean of the genes of the offspring will be shifted from $\mu_{\beta_i^b}$ to β_i^f , the wider the confidence interval the more drastic this effect. This will happen more frequently in multimodal functions and in the first stages of evolution. On the other hand, as the evolution progresses the variance of the values of the genes of the offspring will tend to be smaller than the variance of the best individuals of the population, except to a small range of values of $1 - \alpha$ and n . This way the offspring will tend to be clustered in an area far from the best individuals and nearer to the lower bound of the confidence interval. This will allow changes in the tendency of the search very useful in multimodal functions or in the first stages of the evolution.

3.3 Case 3

For this case, $\beta_i^f \in I_i^U$, the conclusions are the same of case 2. The only difference is that the mean will be shifted to the right, to β_i^f , following

$$E(\beta_i^s) = E(CIUL_i) = \mu_{\beta_i^b} + t_{n-1, \alpha/2} \frac{\sigma_{\beta_i^b}}{\sqrt{n}} \quad (39)$$

3.4 Summary of cases

From the previous study we can conclude that in case 1, CIXL2 will create individuals with a strong exploratory component centered in the mean of the values of the genes of the best individuals. Nevertheless, the amplitude of the exploitation region will increase if the dispersion of the best individuals increases. This will happen mainly in three situations: if the fitness landscape is highly multimodal, in the first stages of the evolutive process, and with large values of n .

Cases 2 and 3 allow the generation of individuals with a strong exploratory component that tend to place themselves in the bounds of the confidence interval. The variance of these individuals will be less than the variance of the best individuals, this will favor the exploration and shifting of the population to the lower bound of the confidence interval. The displacement will be more drastic if the confidence interval is wider.

So, the balance between exploration and exploitation near the best individuals of the population will highly depend on the fitness landscape that will determine the features of the best individuals of the population. CIXL2 will be more exploratory as the multimodality degree of the problem increases, helping the process to avoid being trapped in local minima. On the other hand, in unimodal problems the exploratory component will be small encouraging a fast convergence.

4 Experimental analysis of CIXL2 crossover

This section will experimentally analyze the evolution of the mean and standard deviation of the individuals generated by CIXL2 crossover regarding their parents and the best individuals of the population. We will also study the evolution of the fitness of the offspring with regard to its parents and the best individuals of the population.

We have chosen for the parameters of the CIXL2 crossover $n = 5$ and $1 - \alpha = 0.7$. As it is common in GAs, there can be several copies of the best individuals in the population; when we select n individuals we take care that these individuals are different.

The values of n and $1 - \alpha$ used in this paper have been obtained in a previous work [19] where the sensibility of the crossover to these parameters is analyzed. In that previous work we used the set of test functions proposed in [6]. We considered the values for $n = \{5, 10, 30\}$ and $1 - \alpha = \{0.70, 0.90, 0.95\}$; the conclusion of that work was that the values $n = 5$ and $1 - \alpha = 0.7$ establish a balance between exploration and exploitation that is very adequate for most of the problems. However, higher values of n and $1 - \alpha$ favor the exploratory component of the crossover making the operator more robust when optimizing multimodal functions, but also slower in terms of convergence. Similar results have been obtained in [21], where it is proposed the utilization of a RCGA using CIXL2 as an alternative to classic least

squares estimation methods for non-linear models, and in [20] where CIXL2 is applied to the optimization of functions with restrictions.

The RCGA uses a fixed size population of 100 individuals, randomly initialized, a crossover probability of 0.6, a probability of gene mutation of 0.05, and a tournament selection method with 2 opponents, and elitism. Each experiment will be repeated 10 times with different random seeds and the population will evolve for 100 generations. As mutation operator we have chosen the non-uniform mutation with parameter $b = 5$ [18] as its dynamic nature make it very suitable for a wide variety of problems [10].

We have analyzed the behavior of CIXL2 using three functions [4] with different features, as this behavior depends on the fitness landscape of each function. These functions are shown on the following table:

Function	Expression
Sphere	$f_1(\mathbf{x}) = \sum_{i=1}^q x_i^2$ $x_i \in [-5.12, 5.12]$
Sphere + 4	$f_2(\mathbf{x}) = \sum_{i=1}^q (x_i + 4)^2$ $x_i \in [-5.12, 5.12]$
Rastrigin	$f_3(\mathbf{x}) = 10q + \sum_{i=1}^q (x_i^2 - 10 \cos(2\pi x_i))$ $x_i \in [-5.12, 5.12]$

Sphere is a simple, continuous, strongly convex, unimodal and separable function. Rastrigin is a continuous, scalable, multi-modal and separable function, its contour is made up by a large number of local minima whose values increase with the distance to the global minimum. The minimum of both functions is 0 and it is located in the center of their domains, $x_i = \{0, 0, \dots, 0\}$. Sphere+4 is identical to Sphere but its optimum is shifted from the center of the domain in order to avoid that the RCGA will be favored by this fact. This function will illustrate the behavior of CIXL2 when the optimum is near the bounds of the domain of the function.

The results that are shown in the following sections are obtained for a dimensionality of $q = 2$ for interpretability's sake. Optimization of functions with a higher q would require more generations for the convergence of the GA and the visualization of the behavior of the mean and standard deviation of the offspring with regard to their parents and the best individuals of the population in each generation would be hardly possible. We will study the first of the two genes, as their behavior is similar.

4.1 Sphere

Figure 2 shows the evolution of the mean of the value of gene 0 of the offspring with regard to its parents in one of the 10 runs of the experiment. It show how the mean

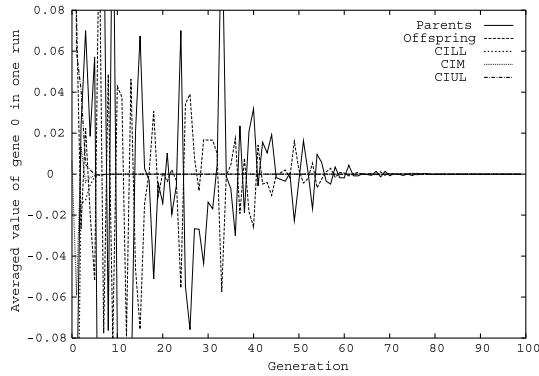


Fig. 2 Values of gene 0 in one of the 10 runs of the RCGA for CILL, CIM and CIUL individuals, and averaged values of parents and offspring for Sphere function

of the offspring is always less than the mean of its parents and tends to 0, the mean value of the 0 allele of the best individuals (CIM_0), as it was proved in the theoretical study. This behavior is most marked in the last generations when the individuals of the population are concentrated in the region bounded by the confidence interval.

We can observe in the figure that, in most generations, when the parents are placed in one side of the optimum the offspring is situated in the opposite side but nearer the optimum. This is due to the attraction effect of CILL, CIM and CIUL individuals when they have better fitness than the parents. This way, the searching process tends to be centered around the optimum with a search domain increasingly restricted. As the optimum is in the center of the domain an effect of zig-zag is observed around the optimum that it is shown on Fig. 2. So, as the evolutive process proceeds the search becomes more exploitative in contrast with the more exploratory search of the first stages. In both cases, the search is driven by the best n individuals of the population.

The Fig. 2 represent the values of gene 0 of CILL, CIM and CIUL individuals. Their values converge very fast towards the optimum from a large initial amplitude to a small final amplitude that favors the fine tuning of the solution. This behavior is quite reasonable as Sphere

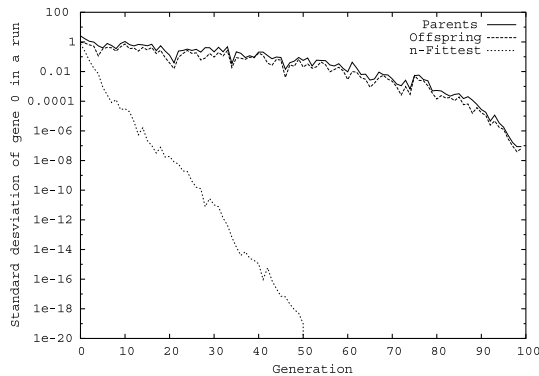


Fig. 3 Standard deviation in logarithmic scale of gene 0 in one of the 10 runs of the RCGA for all the parents, offspring, and n best individuals in a generation for Sphere function

is an unimodal and strongly convex function, and CIXL2 centers the search in the subset of the best individuals of the population. Also, its exploitative component is emphasized in this kind of functions as we have stated in the theoretical study.

Figure 3 shows, in logarithmic scale, the standard deviation of the values of gene 0 of parents and offspring on every generation and the value of the standard deviation of gene 0 of the best n individuals. We can observe that the variance of the offspring is always less or equal than the variance of the parents with a strong tendency to converge towards the best individuals of the population, which favors the convergence of the algorithm. However, due to the features of the function, the best individuals quickly concentrate around the optimum and their standard deviation sensibly differs from the deviation of the offspring which has larger diversity. Although, the standard deviation of the offspring tends to approach the deviation of the best individuals.

It is noticeable the exponential decrease of the standard deviation of the best individuals for this functions. Such exponential decrease helps the convergence and the fine tuning of the solution.

We can see in Fig. 4, in logarithmic scale, how the averaged fitness of the best individuals on the 10 runs of the RCGA of the generated offspring from the parents is always better and tends to approach to the best individuals of the population. This results is coherent with Fig. 2 where the mean of the genes of the offspring is nearer the optimum than its parents, and along the evolution this mean tends to the mean of the best individuals. This feature, together with the fact that the function is convex and unimodal, guarantees in all cases an improvement in the fitness of the offspring with regard to its parents and a progressive tendency to the best individuals.

Figure 4 also shows how the averaged value of the best n individuals and the best individual of the population decreases exponentially, the later being slightly better. This verifies the fast convergence and the fine tuning of the solution of CIXL2 in unimodal functions similar to Sphere.

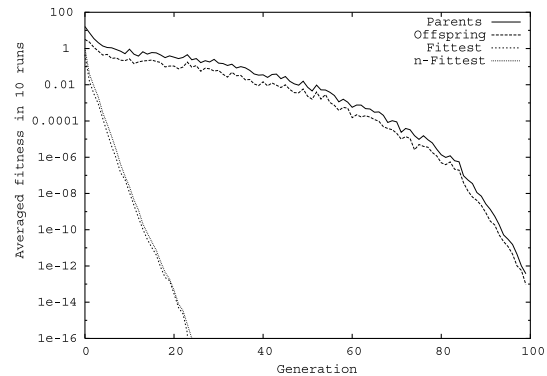


Fig. 4 Averaged fitness in logarithmic scale in 10 runs of parents, offspring, the best and the n best individuals of the population for Sphere function

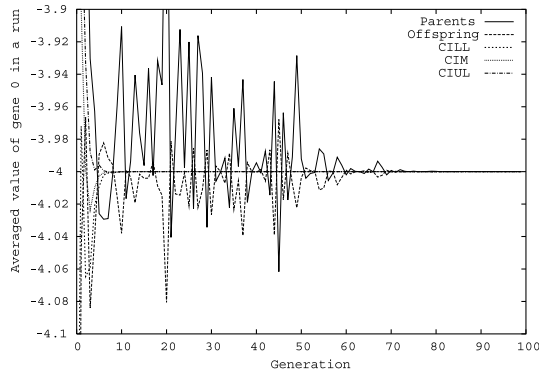


Fig. 5 Values of gene 0 in one of the 10 runs of the RCGA for CILL, CIM and CIUL individuals, and averaged values of parents and offspring for Sphere+4 function

4.2 Sphere + 4

The behavior of CIXL2 in this function is somewhat different from the behavior for the original Sphere function. On Fig. 5 we can see how, as in Sphere, the mean of gene 0 of the generated offspring is closer to the mean of the best n individuals than the mean of its parents, and tends to the mean of the best n individuals that is located around the optimum in -4 . However, the alternation of sign between parents and offspring does not happen, in contrast with Sphere function. In this case, the mean of the parents is usually above the optimum value and the mean of the offspring is below the optimum value. This is due to the fact that CIXL2 tends to balance the population around the mean of the best individuals of the population; as the optimum is shifted to the bound of the domain and the initialization of the population is uniform, most of the parents have values above -4 . So, the crossover generates children in the opposite side, the interval $[-4, 5.12]$, in order to balance the population and make the mean of the population tend to the best individuals.

The speed of convergence, as in Sphere, is very high. The wideness of the confidence interval, delimited by $CILL_0$ and $CIUL_0$, diminishes quickly and the individuals of the population concentrate speedily around the confidence interval.

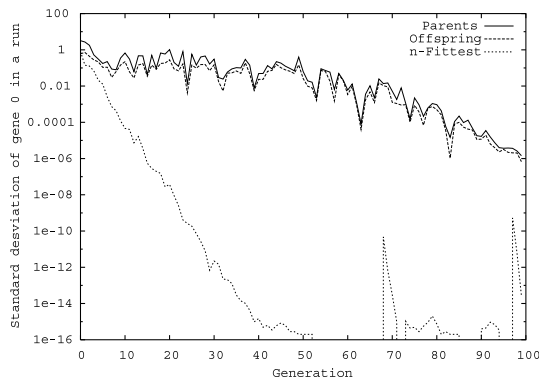


Fig. 6 Standard deviation in logarithmic scale of one of the 10 runs of parents, offspring and best n individuals in a generation for function Sphere+4

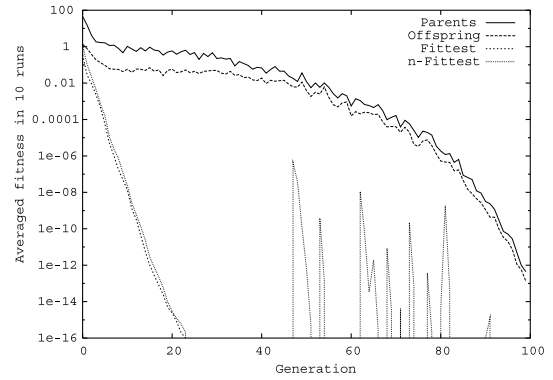


Fig. 7 Averaged fitness in logarithmic scale in 10 runs of parents, offspring, best and n best individuals of the population for function Sphere+4

The behavior of parents and offspring in terms of standard deviation, Fig. 6, is similar to their behavior in Sphere function. Nevertheless, the standard deviation of the best individuals suffers wide oscillations for a short period of time once it reaches its minimum. One possible reason of this behavior is the existence in the population of several copies of the best individuals. In such cases, the selection of n different individuals could lead to consider individuals far from the optimum value. This suboptimal individuals will contribute to an exploration of the search space even in advanced stages of evolution, which could be very useful in multimodal functions.

Figure 7 shows the behavior of parents and offspring in terms of fitness value. This behavior is almost identical to the case of Sphere, with more differences between parents and offspring at the beginning of the evolution. The behavior of the averaged fitness of the best individuals reflects the effect in the standard deviation that we have discussed in the previous paragraph.

4.3 Rastrigin

Figure 8 shows how once the population is stabilized around the mean of the best individuals, the mean of the offspring is near than its parent of the mean of the best

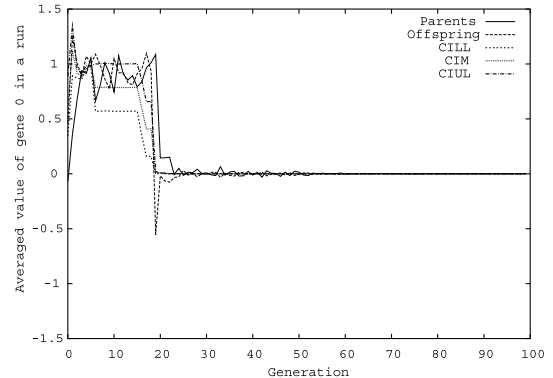


Fig. 8 Values of gene 0 in one of the 10 runs of the RCGA for CILL, CIM and CIUL individuals, and averaged values of parents and offspring for Rastrigin function

individuals of the population. But, during the evolution, it can happen that the best individuals would concentrate around a local optimum and the amplitude of the confidence interval would drastically diminish. This stage will end if the confidence interval widens and the mean of the best individuals is shifted to the optimum, which will happen when some of the best individuals were outside the neighborhood of the local minima. This fact would happen if the status of the evolution corresponds with one of the cases of the theoretical study where the mean of the offspring is shifted to the bounds of the domain or the standard deviation increases, propitiating the creation of individuals outside the neighborhood of the local minima. The mutation operator will also contribute to the creation of such individuals.

The convergence of the best individuals to local minima will be less probable with a larger n . But, once the best individuals are trapped in local minima is easier to get out of them if n is small, because the impact of an individual far from the local minima will be more important. Moreover, it is more probable that in a few generations there will be several copies of the best individuals and the algorithm must look for alternative individuals that help to abandon the local minima.

Figure 8 shows how in this transition periods, when the mean is shifted and the confidence interval becomes very wide, the offspring does not have a clear tendency change. One possible explanation is that, due to the multimodality of the function, the individuals CILL, CIM and CIUL do not always have a better fitness than the parents. The offspring is not strongly attracted to the confidence interval if these individuals are worse than the parents.

On Fig. 9 we can see how the standard deviation of the offspring is not always less than the deviation of its parents, specially in the transition periods when the standard deviation of the best individuals is very large, even larger than the deviation of the parents and offspring. This is a reasonable behavior if we take into account the multimodality of the function and Fig. 8. On the other hand, in more advanced stages of evolution the standard deviation hit high peaks after reaching a

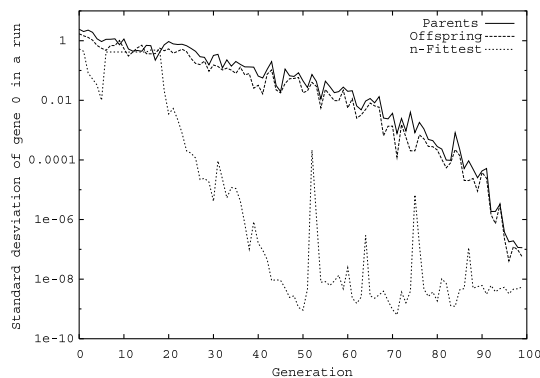


Fig. 9 Standard deviation in logarithmic scale of gene 0 in one of the 10 runs of parents, offspring, and best n individuals in a generation for Rastrigin function

minimum, as in Sphere+4. In this occasion the peaks are more numerous as the function is multimodal. Nevertheless, the reason why these peaks appear is the same that in Sphere+4.

On Fig. 10 we can see how the averaged fitness of the offspring is not always better than the fitness of its parents, due to the multimodality of the function. In contrast with the two previous functions, at the beginning the value of the function decreases slowly till the optimum is targeted and the descent becomes exponential. The descent in the value of the best individuals is clearly stepped in the transition stage with abrupt falls in the transitions between different local minima. The fitness of the best individuals also suffers oscillations in the last stages of the evolutionary process in concordance with the behavior of the standard deviation.

5 Conclusions and future work

In this paper we have shown the features of CIXL2 crossover, both theoretically and experimentally. As the theoretical study assumes several hypothesis, such as normality and independence of the distributions, that are difficult to verify in real problems, we have carried out the experimental study to support that the theoretical conclusions have their correspondence in the experiments. Nevertheless, as the hypothesis are not always fulfilled in the experiments, the behavior of the mean and standard deviation developed in the theoretical study must be apprehended in terms of tendencies in the experimental study.

In the experimental study, we have corroborated that CIXL2 generates an offspring that tends to approach CILL, CIM and CIUL individuals, that is, the mean of the best individuals of the population within a certain confidence interval, provided that the fitness of these individuals is better than the fitness of the rest of individuals that participate in the crossover. These virtual parents summarize the features of the best individuals of the population, and transmit such features to the offspring if they mean an evolutive improvement. In unimodal functions, such as Sphere and Sphere+4, where

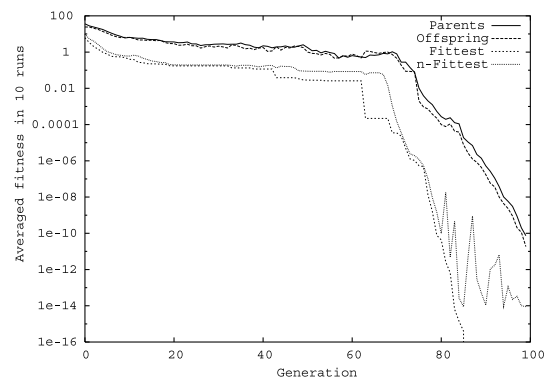


Fig. 10 Averaged fitness in logarithmic scale of parents, offspring, best and n best individuals for Rastrigin function

the fitness of the virtual parents is better than the fitness of the parent from the population, the generated individuals are always closer to the best individuals than its parents, approaching quickly the mean of the best individuals. But, in multimodal functions, such as Rastrigin, the fitness of CILL, CIM and CIUL individuals is not always better than the fitness of the parent and the offspring does not tends clearly to these individuals till their fitness is improved.

We have also shown that in unimodal functions CIXL2 crossover reduces the standard deviation of the offspring with regard to the deviation of its parents. The deviation of the offspring tends to the deviation of the best individuals. This tendency is not so clear in multimodal functions where, during certain periods of time, the standard deviation of the offspring is larger than the deviation of its parents and even larger than the deviation of the best individuals of the population.

So, we can conclude that CIXL2 establishes a balance between exploration and exploitation around the best individuals of the population. This balance depends on the shape of the fitness landscape, as this shape determines the features of the distributions of the best individuals. CIXL2 will be more exploratory as the multimodality of the function increases, diminishing the risk of being trapped in local minima. In unimodal functions the exploitative component will be predominant allowing a fast convergence. In both cases, at the last stages of the evolution, when the best individuals are concentrated around the optimum, the crossover will be mainly exploitative, favoring a local fine tuning of the solution.

The theoretical study states that the statistical features of the offspring depend not only on the fitness landscape, but also on the parameters of the crossover. Although we have made some efforts in this direction, [19], one of our future research lines is an exhaustive study of the optimal values for the number of best individuals, n , the confidence coefficient, $1 - \alpha$, and the range of r for each kind of problem.

It will also be interesting a study of some dynamic strategies of modification of the values of the parameters of the crossover along the evolution. The aim would be to increase the exploratory component of the crossover at the beginning of the evolution and the exploitative component at the end of the evolution, in order to improve the robustness, speed of convergence and local fine tuning to the RCGA.

Acknowledgements This work has been financed in part by the project TIC2002-04036-C05-02 of the Spanish CICYT and FEDER funds.

References

- Bäck JH (1996) *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, Oxford
- Beyer HG, Deb K (2001) On self-adapting features in real-parameter evolutionary algorithms. *IEEE Trans Evol Comput* 5(3): 250–270
- Davis L (1991) *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, New York
- De Jong KD (1975) An analysis of the behavior of a class of genetic adaptive systems. PhD thesis, Department of Computer and Communication Sciences, University of Michigan, Ann Arbor
- Deb K, Agrawal RB (1995) Simulated binary crossover for continuous search space. *Complex Systems*, 9:115–148
- Eiben AE, Bäck Th (1997) Empirical investigation of multi-parent recombination operators in evolution strategies. *Evol Comput* 5(3):347–365
- Eshelman LJ, Schaffer JD (1993) Real-coded genetic algorithms and interval-schemata. In: Darrell Whitley L (ed), *Foundation of Genetic Algorithms 2*, San Mateo Morgan Kaufmann pp 187C3.3.7:1–C3.3.7:8.-202
- Goldberg DE (1991) Real-coded genetic algorithms, virtual alphabets, and blocking. *Complex Syst* (5):139–167
- Herrera F, Herrera-Viedma E, Lozano E, Verdegay JL (1994) Fuzzy tools to improve genetic algorithms. In: *Second European Congress on Intelligent Techniques and Soft Computing*, pp 1532–1539
- Herrera F, Lozano M (2000) Gradual distributed real-coded genetic algorithms. *IEEE Trans. Evol Comput* 4(1):43–63
- Herrera F, Lozano M, Verdegay JL (1995) Tuning fuzzy logic controllers by genetic algorithms. *Int J Approx Reasoning*, (12):299–315
- Hervás C, Ortiz D (2002) Operadores de cruce basados en estadísticos de localización para algoritmos genéticos con codificación real In: Gomez JA, Herrera F, Hidalgo JI, Lanchares J, Merelo JJ, Sánchez JM, Alba E, Fernandez F (eds), *Primer Congreso Español De Algoritmos Evolutivos y Bioinspirados (AEB'02)*, Mérida, pp 1–8
- Holland JH (1975) *Adaptation in natural and artificial systems*. The University of Michigan Press, Ann Arbor, MI
- Kendall M, Stuart A (1979) *Inference and relationship*, volume 2 of *The advanced theory of statistics*. Charles Griffin and Co Ltd, 4 edn
- Kita H (2001) A comparison study of self-adaptation in evolution strategies and real-code genetic algorithms. *Evol Comput* 9(2):223–241
- Kita H, Ono I, Kobayashi S (1998) Theoretical analysis of the unimodal normal distribution crossover for real-coded genetic algorithms. In: *IEEE International Conference on Evolutionary Computation IEC'98*, Anchorage, pp 529–534
- Koza JR (1992) *Genetic Programming*. The MIT Press
- Michalewicz Z (1992) *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, Berlin Heidelberg New York
- Ortiz D (2001) Operadores de cruce basados en intervalos de confianza en algoritmos genéticos con codificación real. PhD thesis, E.T.S.I. Informática, Málaga
- Ortiz D, Hervás C, García N (2002) Crossover operator effect in function optimization with constraints. In: Merello JJ, Adamidis P, Beyer H.-G, Fernandez JL, Schwefel HP (ed) *The 7th Conference on Parallel Problem Solving from Nature*, vol. 2439 of *Lecture Notes in Computer Science*, Granada Springer, Berlin Heidelberg New York, pp 184–193
- Ortiz D, Hervás C, Muñoz J (2001) Genetic algorithm with crossover based on confidence interval as an alternative to traditional nonlinear regression methods. In: Verleysen M. (ed), *The 9th European Symposium on Artificial Neural Networks*, Bruges, Belgium, D-Facto, pp 193–198
- Radcliffe NJ (1991) Equivalence class analysis of genetic algorithms. *Complex Systems*, 2(5):183–205
- Wright A (1991) Genetic algorithms for real parameter optimization. In: Rawlin GJE (ed) *Foundations of Genetic Algorithms 1*, San Mateo, Morgan Kaufmann, pp 205–218