Theoretical Analysis on an Inversion Phenomenon of Convergence Velocity in a Real-Coded GA

Hiroshi Someya, Member, IEEE

Abstract—The aims of this paper are to analyze an inversion phenomenon theoretically and discussion on appropriateness of combination of a crossover operator and a selection model. In the previous study, the author designed a crossover operator that worked well on various kinds of objective functions. One of the features of the objective functions is "the optimum exists near a boundary much more than the other." On such objective functions, with recommended selection model, the proposed crossover operator set with an appropriate parameter has shown the fastest convergence speed. However, with another selection model, its convergence speed has been the slowest. In order to understand this inversion phenomenon, a theoretical analysis quantified the selection pressures of the selection models and estimated the expected positions of the center of gravity of the population. The theoretical results corresponded to empirical verifications and successfully explained. Finally, a guideline for designing RCGAs was obtained.

I. INTRODUCTION

Recent years, several researchers have studied Real-Coded GAs (RCGAs) for objective functions where the optimum is located near a boundary [1], [2], [3], [4]. They have reported that most RCGAs prefer to search the center of the search space much more than the other. An example of such bias is shown in Fig.1. The left figure was obtained by an actual computer simulation. In this simulation, the individuals of the initial population were distributed randomly in the search space [−5.5]. In each generation, the parents randomly picked produced their children using Unimodal Normal Distribution Crossover (UNDX) [5]. The survival selection was performed as the random picking manner. Although all the selection procedures were completely at random, the searched points were clearly biased toward the center of the search space. This bias works advantageously only in the case that the optimum happens to be located in the center of the search space. In the other cases, it works vice versa. This paper claims that the robustness should be given priority over the performance in a specific case because no one knows where the optimum exists before a trial in real-world applications. The right figure in Fig.1 illustrates the theoretically obtained sampling bias of UNDX. Because the biased-curve is proper to the crossover operator, invention of an improved crossover operator would be believed to be the best promising approach for this problem.

In the previous study, the author designed a new crossover operator, Asymmetrical Normal Distribution Crossover (ANDX) [6]. One of the advantageous features of this crossover operator is flexibility. The shape of the probability density function (p.d.f.) of this crossover operator can be controlled in a continuous fashion. In a certain parameter setting, the theoretical biased-curve of ANDX is relatively flat.

In several experiments where the performance measure adopted was the number of finding the optimum, ANDX has worked well on all of a variety of test functions, such as $F_{\text{step}}, F_{\text{rastrigin}}, F_{\text{ikeda}}$ and $F_{\text{rosenbrock}}$. Their fitness landscapes are characterized by several different structures: discreteness, big valley, punch bowl, ridge structure, epistasis structure, UV structure, and structures where the optimum exists near a boundary rather than in the center of the search space. Therefore, the author concluded that ANDX would be suitable for searching near a boundary of the search space. Some parts of the experimental results are shown in Table I. Each test function except $F_{\text{rosenbrock}}$ has the optimum at the origin. The experiments on them were performed with the initial population distributed in [−0.62, 9.62]. $F_{\text{rosenbrock}}$ has the optimum at (1, ..., 1). The initial domain was set to be [-2.048, 2.048]. The $\zeta$ is a parameter of ANDX. The MGG and X-MGG are the selection models adopted. Then, the author examined the convergence curves of these experiments. The curves on $F_{\text{rosenbrock}}$ drew two curious graphs shown in Fig.2. ANDX$_{\zeta=0.4}$ with X-MGG rapidly heads toward the optimum. The convergence speed ranking of ANDXs with X-MGG is ANDX$_{\zeta=0.4}$, ANDX$_{\zeta=0.3}$, ANDX$_{\zeta=0.2}$. However, in the case of with MGG, the ranking is the opposite. The convergence velocity of ANDX$_{\zeta=0.4}$ with MGG is the slowest. Why?

The aims of this paper are to analyze this inversion phenomenon theoretically, and discussion on appropriateness of combination of a crossover operator and a selection model.

(a) Experimental

(b) Theoretical

Fig. 1. The sampling bias of UNDX [4].
The theoretical analysis quantifies the selection pressures of the selection models and estimates the expected positions of the center of gravity of the population.

II. ANALYSIS OBJECTS

A. UNDX with MGG

This method is widely referred and adopted in many literatures. It works well even though the landscapes of objective functions are characterized by multi-modal, discrete and high-dimensional.

UNDX produces two children in a single crossover operation around the middle point of the parents, Parent1 and Parent2. The children vectors, \( c^{(1)} \) and \( c^{(2)} \), are determined as follows:

\[
\begin{align*}
    c^{(1)} &= g + z_1 e_1 + \sum_{k=2}^{n} z_k e_k, \\
    c^{(2)} &= g - z_1 e_1 - \sum_{k=2}^{n} z_k e_k.
\end{align*}
\]

The parent vectors \( p^{(1)} \) and \( p^{(2)} \) introduce their middle point vector \( g = (p^{(1)} + p^{(2)})/2 \) and the orthogonal unit vectors \( e_1 = (p^{(2)} - p^{(1)})/(|p^{(2)} - p^{(1)})|, e_k (k = 2, \ldots, n) \), where \( n \) is the dimension of objective function, \( z_1 \sim N(0, \sigma_1^2) \) and \( z_k \sim N(0, \sigma_k^2) \) are normally distributed random numbers, where \( \sigma_1 = \alpha d_1 \) and \( \sigma_2 = \beta d_2 \sqrt{n} \), and \( \alpha \) and \( \beta \) are constants. \( d_1 \) is the distance between Parent1 and Parent2. \( d_2 \) is the distance of Parent3, the third parent, from the line connecting Parent1 and Parent2.

Minimal Generation Gap (MGG) model [7] is described in the followings:

1. Generate an initial population randomly.
2. Choose a pair of individuals as parents from the population randomly.
3. Make a certain number of children by a crossover.
4. Select the best individual out of the family, the parents and their children.
5. Choose an individual except the best out of the family randomly according to fitness-based or ranking-based wheel selection.
6. Replace the two selected individuals to the parents.
7. Iterate step2~step6 until a certain condition is satisfied.

In this model, evolutionary stagnation is avoided by selecting the best individual at step4. Early convergence is prevented by the roulette-wheel selection at step5.

B. ANDX with X-MGG

ANDX has been designed for searching the following promising search regions: (i) near a promising solution candidate (a parent), (ii) the inner region of a parent distribution, (iii) independent regions of the coordinate systems, and (iv) a region satisfying “preservation of statistics”. Examples in Fig.3 explain that ANDX searches the promising search regions.

ANDX produces children near each parent. Let \( g \) be the center of mass of \( m \) parents, \( p^{(1)}, \ldots, p^{(m)} \), and let \( d^{(j)} = p^{(j)} - g \). A child \( c^{(s)} \) is produced near \( p^{(s)} \), chosen from the parents randomly, in the following equation:

\[
c^{(s)} = p^{(s)} - \sum_{j \neq s}^{m} d^{(j)} \omega^{(j)} + q.
\]

The symbol \( \sum_{j \neq s}^{m} \) expresses the notation for summation of index \( j = 1, \ldots, m \) except for \( j = s \). Each \( \omega \) is an asymmetrical normal random variable defined as:

\[
\omega \sim \begin{cases} 
    2\zeta_o N(0, \sigma_o^2) & (x \geq 0) \\
    2\zeta_i N(0, \sigma_i^2) & (x < 0)
\end{cases}
\]

where \( \zeta_o + \zeta_i = 1 \), \( \sigma_o = \zeta_o \sigma \), and \( \sigma_i = \zeta_i \sigma \). The \( \zeta_o \) and \( \sigma \) are the parameters of ANDX. The former is set to satisfy...
\(\zeta_0 < \zeta_i\) in order to prefer the promising search region 2. When the equation,
\[
\sigma = \frac{2\sqrt{2\pi}(2\zeta_0 - 1)}{2(m - 2)(2\zeta_0 - 1)^2 - \pi(m - 1)(3\zeta_0^2 - 3\zeta_0 + 1)} \tag{3}
\]
is satisfied, the children are placed in the promising search region 4. Examples of p.d.f.s introduced by such parameters are demonstrated in Fig.4. Let \(W\) be the subspace spanned by \(d^{(1)}, \ldots, d^{(m)}\), and let \(W^\perp\) be the orthogonal complement. The \(q\) in \(W^\perp\) would be defined, for example, as the same manner of UNDX.

The eXclusive-MGG (X-MGG) [6] model has been introduced in order to equip MGG with exclusive replacement strategy. To avoid premature convergence of the population in the early stages, the population should contain a variety of characteristics of genes. For this, only one individual should survive from parents and their children, because each individual that belongs to the family has similar characteristics of genes. In this selection model only one individual is allowed to survive from a subfamily derived from dividing the family. Each subfamily consists of a parent and especially similar children to the parent. Procedure of a generation of this model using ANDX is described as follows:

**step 1** Choose \(m + 1\) parents, \(P = \{p^{(1)}, \ldots, p^{(m)}\}\) and \(p^{(m+1)}\), randomly from the population.

**step 2** Produce children, \(C = \{C^{(1)}, \ldots, C^{(m)}\}\) where \(C^{(j)}\) is a group including \(c^{(j)}\) produced near \(p^{(j)}\), using ANDX. Let \(S = \{S^{(1)}, \ldots, S^{(m)}\}\) where \(S^{(j)}\) is a subfamily that consists of \(p^{(j)}\) and \(C^{(j)}\).

**step 3** Select the best individual \(I_{\text{best}}\) in \(S\). Let \(S^{(\text{best})}\) be the \(S^{(j)}\) that includes \(I_{\text{best}}\), and let \(p^{(\text{best})}\) be the parent that belongs to \(S^{(\text{best})}\).

**step 4** Select one individual \(I_{\text{roulette}}\) by fitness-based or ranking-based roulette-wheel selection from the group \(S^{(\text{rest})}\) derived from \(S\) by removing \(S^{(\text{best})}\). Let \(p^{(\text{roulette})}\) be the parent that belongs to the \(S^{(j)}\) including \(I_{\text{roulette}}\).

**step 5** replace \(p^{(\text{best})}\) with \(I_{\text{best}}\).

**step 6** replace \(p^{(\text{roulette})}\) with \(I_{\text{roulette}}\).

### III. Theoretical Analysis

Firstly, this section discusses selection pressures of MGG model and X-MGG model. Secondly, expected values of children vectors selected are introduced. Finally, enlargement and acceleration of a population are quantified. This paper adopts ranking-based selection as the roulette-wheel selections of MGG and X-MGG.

**A. Terms**

The best evaluated member in a family is called the best member and a selected member by a roulette-wheel selection is called roulette selection member.

**B. The expected ranking in the ranking-based roulette-wheel selection**

Let consider a roulette-wheel selection for choosing out of a population consists of \(k\) individuals. The actual ranking to be selected is unknown because it would be determined by using a random number generator. Thus, we should consider the expected ranking, \(R(k)\).

The survival odds of the first, the second, the third, \ldots, and the worst member should be assigned as: \(\frac{c_k}{k}, \frac{c_{k-1}}{k-1}, \ldots, \frac{c_1}{1}\), where \(c_k\) is a constant that satisfies \(\sum_{r=1}^{k} \frac{c_1}{r} = 1\). Hence,

\[
c_k \cdot \left( \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{k-1} + \frac{1}{k} \right) = 1
\]

and

\[
c_{k-1} \cdot \left( \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{k-1} \right) = 1
\]

introduce

\[
c_k = \frac{k \cdot c_{k-1}}{k + c_{k-1}} . \tag{4}
\]

Arbitrary \(c_k\) can be derived from this recurrence formula and \(c_1 = 1\). Hence, the expected ranking is estimated as,

\[
R(k) = \sum_{r=1}^{k} r \cdot \frac{c_k}{r} = c_k \cdot k . \tag{5}
\]
C. Normalization

The ranking 10 is the worst in a 10 member group, but it is not bad in a 20 member group. This example explains that ranking value is often hard to be handled. This paper adopts a normalized ranking value to measure the quality of an individual. The value $Q(r,k)$ is defined as:

$$Q(r,k) = \frac{r - 0.5}{k},$$

(6)

where $k$ is the number of members within the focused group, and $r$ is the raw ranking value. For example, each member within a group that consists of four members is given the normalized ranking: 0.125, 0.375, 0.625 or 0.875 respectively. The values correspond to the positions uniformly distributed.

D. Selection Pressures

1) MGG model: The normalized ranking of the best member must be

$$Q_{\text{best}} = Q(1,t) = \frac{0.5}{t},$$

(7)

where $t$ is the family size. Because the roulette-wheel selection is performed among the rest members, the raw ranking value of the roulette selection member must be $1 + R(t-1)$. Thus, corresponding normalized ranking is derived from (6) as

$$Q^{\text{MGG}}_{\text{roulette}} = \frac{R(t-1) + 0.5}{t}. $$

(8)

2) X-MGG model: The $Q_{\text{best}}$ of X-MGG is the same as that of MGG. The roulette-wheel selection is performed among the rest subfamilies, $S^{(\text{rest})}$. The number of the rightful members is expressed as

$$t^{(\text{rest})} = \frac{t(m-1)}{m}. $$

(9)

Thus, the normalized ranking in $S^{(\text{rest})}$ is derived as:

$$Q^{\text{X-MGG}}_{\text{roulette}} = \frac{R(t^{(\text{rest})}) - 0.5}{t^{(\text{rest})}}. $$

(10)

E. Expected Position

For simple discussion, the population size is set to be two. Each of their initial positions is -1 or 1. The objective function is defined by $F(x) = x$. This optimization is treated as a one-dimensional maximization problem. Consequently, the optimum $x = \infty$ exists far from the initial population. Let $G(x)$ be the integral of the normal distribution $\int_{-\infty}^{x} N(0,1)$. The symbol $PDF(\text{crossover name})$ expresses the p.d.f. of a crossover operator. $x_Q$ is the position of the $x$ that satisfies $\int_{-\infty}^{x_Q} PDF(\text{crossover name}) = Q$.

1) UNDX with MGG: The $PDF(\text{UNDX})$ is just the $N(0,1)$. Therefore, $x_{Q,\text{m}}$, and $x_{Q,\text{MGG}}^{\min}$ are directly obtainable by the table of normal distribution.

2) ANDX with MGG: Let $H_{a,b}(x)$ be the integral $\int_{a}^{b} PDF(\text{ANDX})$. This is divided into the following four parts:

the outside of the right part of the $PDF(\text{ANDX})$

$$G_r(x) = \begin{cases} \xi_{c} \int_{x}^{\infty} N(1,\sigma_{x}) = \xi_{c} G(z_{r}), & (1 \leq x) \\ \frac{z_{r}}{2} & (x < 1), \end{cases}$$

(11)

the inside of the right part of the $PDF(\text{ANDX})$

$$G_r(x) = \begin{cases} 0 & (1 \leq x) \\ \xi_{i} \int_{x_{r1}}^{x} N(1,\sigma_{x}) = \xi_{i} \int_{x_{r1}}^{0} N(0,1) \\ \xi_{i} G(z_{r1}) - 0.5, & (x < 1), \end{cases}$$

(12)

the inside of the left part of the $PDF(\text{ANDX})$

$$G_l(x) = \begin{cases} \xi_{l} \int_{-\infty}^{x} N(-1,\sigma_{x}) = \xi_{l} G(z_{l}), & (-1 < x) \\ \frac{z_{l}}{2} & (x \leq -1), \end{cases}$$

(13)

the outside of the left part of the $PDF(\text{ANDX})$

$$G_l(x) = \begin{cases} 0 & (-1 < x) \\ \xi_{o} \int_{x_{l1}}^{x} N(-1,\sigma_{x}) = \xi_{o} \int_{x_{l1}}^{0} N(0,1) \\ \xi_{o} G(z_{l1}) - 0.5, & (x \leq -1), \end{cases}$$

(14)
The real-number was that the population may converge into a certain
Thus, arbitrary \( x_Q \), including \( x_{Q_{best}} \) or \( x_{Q_{roulette}} \), that satisfies
\[
H_{left}(x) = G_{ro}(x) + G_{ri}(x) + G_{li}(x) + G_{lo}(x) = Q
\]
is obtainable by the table of normal distribution.

3) ANDX with X-MGG: The \( x_{Q_{best}} \) is obtained by the same manner of the ANDX with MGG. In the case that the best is derived from the right asymmetrical normal distribution, the roulette-wheel selection is performed based on the left one. In this case, \( x_{Q_{roulette}} \) must satisfy
\[
H_{left}(x) = 2 \cdot (G_{li}(x) + G_{lo}(x)) .
\]
In the other case, it must satisfy
\[
H_{right}(x) = 2 \cdot (G_{ro}(x) + G_{ri}(x)) .
\]
Consequently, the expected value of \( x_{Q_{roulette}} \) is obtained by the weighted average, where the weight is defined by the ratio of these cases, \( \lambda = \frac{H_{right}(x_{Q_{best}})}{H_{left}(x_{Q_{best}})} \).

F. Convergence Speed

Let \( g_k \) be the center of gravity of parents after the k-th crossover operation. In cases of \( k \geq 2 \), the enlargement ratio of the distance between \( g_k \) and \( g_{k-1} \) per a single crossover operation, \( u_1 \), must be
\[
u_1 = \frac{g_k - g_{k-1}}{g_{k-1} - g_{k-2}} = \frac{g_{k-1} - g_{k-2}}{g_{k-2} - g_{k-3}} = \ldots .
\]
This equation and \( g_0 = 0 \) give
\[
g_k = \frac{g_1}{1 - u_1} (1 - u_1)^k .
\]
When \( k \rightarrow \infty \), \( u_1^k \rightarrow 0 \) if \( u_1 < 1 \). Hence, this result indicates that the population may converge into a certain point,
\[
g_\infty = \frac{g_1}{1 - u_1} ,
\]
before finding the optimum far from the initialized region. On the other hand, in the case of \( 1 < u_1 \) and \( 0 < g_1 \), the population would reach the optimum.

G. Numerical Example

An example case of \( t = 52 \) (adopted in [6]) is shown here. The selection pressures of MGG and X-MGG are
\[
Q_{best} = \frac{0.5}{52} = 0.0096, \\
Q_{MGG} = \frac{11.29 + 0.5}{52} \approx 0.23, \\
Q_{X-MGG} \approx 6.75 - 0.5 - 6.49 = 0.24.
\]

They give us Table II. Note, the values include a certain degree of computational errors, such as roundoff errors and interpolation errors [8]. All computation was performed by Java language version "1.5.0.10". The real-numbers were stored in Java’s double type. The integral values were obtained by the linear interpolation.

IV. EMPIRICAL VERIFICATION

This section compares the pure mathematical search process with the actual optimization search process.

A. Conditions

Section III assumes several impractical conditions. In experiments here, somewhat actual conditions are adopted. The objective function used is the Sphere function:
\[
F_{sphere} = \sum_{i=1}^{n} x_i^2 .
\]
The optimum of this function is just the origin. The individuals of the initial population were uniformly distributed in [9999, 10000]. Therefore, the population must go on a long long distance trip. The dimension \( n \) was set to be 100. In each generation, a set of two parents and one additional parent produce 50 children. The performance measures adopted were both the number of finding optimum and the convergence speed. They were evaluated on the average of 30 trials.

B. Experimental Results and Discussions

The first experiments were performed under the similar condition to the theoretical condition; the population size is only three. This size is the smallest for UNDX and ANDX. In this case, only ANDX_{c=0.4} with X-MGG reached the optimum and the others converged in a certain point. In the next case that the population size is four, all methods except ANDX_{c=0.4} with MGG reached the optimum. The transition curves in the case that the population size is five are shown in Fig. 5 (pop = 5). These graphs are similar to those in Fig. 2. This inversion phenomenon is successfully explained by the theoretical analysis. The larger \( g_\infty \) is better for
finding the outer optimum. In spite of the smaller number of generating children (not a little sampling error must happen), this convergence speed order approximately corresponds well to the ranking order based on the theoretical values.

The primary difference between the cases of Fig. 2 and Fig. 5 \((\text{pop} = 5)\) is the population size. The size 30 used on \(F_{\text{rosenbrock}}\) is far larger. The author continued the experiments on \(F_{\text{sphere}}\) with larger population sizes: \(6, 9, 12, \ldots, 30\). With some of these sizes, overlapped transition curves were observed. An example of such a case is shown in Fig. 5 \((\text{pop} = 15)\). Why did the overlapped curves appear only on \(F_{\text{sphere}}\)?

On \(F_{\text{rosenbrock}}\), Sakuma and Kobayashi have demonstrated the typical search process of the population in detail [3]. In the early search stage, the population extremely prefers to gather around the origin. Then the population moves along the curved sheer cliff toward the optimum. According to this demonstration, the main factor of that difference should be the extent of the population distribution. With larger population size, the combinatorial variations of parent pairs for a crossover operation are exponentially larger. Therefore, even if the distance between the survived children is short, the distance between the parents for the next crossover operation may have enough range. This means that larger population size weakens the influence of the distance between the survived children. On \(F_{\text{sphere}}\), the population distribution can easily enlarge if the crossover operator and the selection model allow. On the other hand, on \(F_{\text{rosenbrock}}\), the curved sheer cliff always prevents the population from rapid marching.

When the population size is 60, the transition curve of \(\text{ANDX}_{c=0.4}\) with \(X\text{-MGG}\) is the slowest, as shown in Fig. 5 \((\text{pop} = 60)\). In such case, the value of \(g_1\) would be more important to enlarge the population. This experiment suggests the important measure is not only one. The measures: \(u_1, g_1\) and \(g_{\infty}\), would have respective roles.

Finally, this paper qualitatively discusses the appropriateness of combination of a crossover operator and a selection model. We have learned an important guideline for designing a RCGA: “consider the total balance of the components,” through the above theoretical and empirical studies. The author had tried to discover only the relation between ANDX and MGG, or ANDX and X-MGG. However, we understand now that the appropriateness of the combination would be controlled by the population size parameter. Furthermore, the other parameter, the number of produced children per generation, also has an influence on the selection pressure, as shown in (6)∼(10). A GA that consists of several highly independent elements should be a better designed GA. Otherwise, designing a GA would continue being a kind of combinatorial optimization problems (hard work!).

V. CONCLUSIONS AND FUTURE WORKS

This paper theoretically analyzed an inversion phenomenon of the convergence velocities of ANDX with MGG and ANDX with X-MGG. The theoretical results corresponded to the empirical verification. The author has concluded that the main factor of the inversion phenomenon is the diversity change of the population. The diversity
change has been controlled by not only crossover operators and selection models but the population size parameter. The equations derived in the theoretical analysis informed that the number of produced children per a generation would be another factor to control the selection pressure. Designing improved genetic operators that have highly independence is an important future work. Comparison of ANDX with X-MGG and other RCGAs, such as introduced in [9] and [10], is also a future work.

ACKNOWLEDGMENT

This work was supported in part by the “Function and Induction research project” and by the Grant-in-Aid for Science Research (A), No.17200020.

REFERENCES


