Complex Systems **10** (1996) 167–183

Theoretical Analysis of Genetic Algorithms with Infinite Population Size

Stefan Voget[*]
Institut für Mathematik, Universität Hildesheim,
Marienburger Platz 22,
D-31141 Hildesheim, Germany

Abstract. A genetic algorithm (GA) is a stochastic search and optimization algorithm that works by iterative application of several evolutionary operators. An approximation model is presented for the algorithm with the population size set to infinity. Results are given for the repeated, isolated application of the operators selection, crossover, and mutation. The speed of convergence to a limit distribution is examined. These examinations extend results of previous published results of GAs with an infinite population size.

1. Introduction

A genetic algorithm (GA) is a stochastic optimization algorithm that emulates processes of natural evolution in an abstract manner. The aim is to propagate a set of possible solutions such that an optimal or at least a very good solution can grow in a dynamic process. Drawing ideas from the natural prototypes of evolutionary operators (such as selection, mutation, crossover, etc.) the GA explores a search space from many different points simultaneously.

The algorithm starts with a set of N ($N \in \mathbb{N}$) configurations in a search space called *population*. A member of a population is called an *individual*. Taking such a population as the first generation, the algorithm produces new generations iteratively in the following way.

- First, randomly choose two parent individuals from a population of size N. This selection step realizes the exploitation of good solutions. Individuals with high fitness are preferred for selection.

- Evolutionary production operators are applied to generate new solutions inside the search space. Crossover and mutation realize an exploration through the search space.

[*] Electronic mail address: `voget@informatik.uni-hildesheim.de`.

- The first two steps are repeated until a new population of size N is filled.

A lot of successful applications of the GA are available [5, 6, 15]. But the algorithm is influenced by several parameters such as selection pressure, probability of crossover and mutation, and population size. The choice of these parameters makes the implementation difficult. The necessity of a mathematical description is obvious.

A lot of mathematical models suppose an infinite population size [16, 22]. They constitute an approximation of the exact model and allow for the precise study of isolated operators. These operators are defined by an iterative process. We present explicit formulas for the iterative processes of repeated selection, mutation, and crossover.

The description of the selection operator is known in the literature [1]. In section 3 we give some further results about speed of convergence. Often the mutation operator is used as a background operator and is not considered any further, we present a complete description of the mutation operator in section 4. The main production operator within a GA is crossover, this operator is discussed in section 5. A complete description for a kind of crossover arising in quantitative population genetics is given in [11], we transfer this model into the model of GAs. With this assignment we are able to give a comparison between different kinds of crossover. By the theoretical model a new kind of crossover, the so-called *halfcross*, is distinguished. In halfcross exactly one half of the coordinates are taken from each of the parents. This corresponds to the expected number of coordinates that uniform crossover takes from each parent.

2. Basic algorithm formulation

Let $K \in \mathbb{N}$ and Ξ_k $(k = 1, \ldots, K)$ be finite sets. With the definition of a search space $\Xi = \prod_{k=1}^{K} \Xi_k$ and an optimization function $f : \Xi \to \mathbb{R}^+$ the optimization task is given by

$$\arg\max_{i\in\Xi} f(i).$$

K denotes the dimension of the search space. A population is defined by $\boldsymbol{x} = (x_i)_{i\in\Xi}$ with the properties

$$x_i \in [0,1], \quad \sum_{i\in\Xi} x_i = 1.$$

Members of the population are called *individuals*. The coordinate with the index i denotes the relative proportion of the individual i in the population x. With $y \in \mathbb{R}^{\Xi}$ we will use the norm $\| y \| := \sum_{i\in\Xi} |y_i|$ on $\mathbb{R}^{\Xi}$.

A GA is given by iterative application of the following steps.

Step 1: *Selection.* Given a population $x(n)$ with $n \in \mathbb{N}$, members of the population are selected independently. In the evolution process individuals with higher fitness values are preferred, and individuals with low fitness get a lesser chance for selection. After the selection step, the distribution of a randomly chosen individual $I^S(n)$ is given by

$$\forall i \in \Xi : \; P(I^S(n) = i \mid X(n) = x(n)) := \frac{f(i) * x_i(n)}{\sum_{j\in\Xi} \; f(j) * x_j(n)}. \tag{1}$$

Usually this kind of selection is called proportional selection. Examples of this and other kinds of selection can be found in [1, 2, 7, 8, 12, 13, 14].

Step 2: *Crossover.* After the independent selection of two individuals, crossover is applied. A crossover schema is defined by a random variable U with values in $\{0,1\}^K$. By $\bar{u}$ we denote the bitwise complement of u and by $\odot$ and $\oplus$ we denote the componentwise binary multiplication and summation. The kth coordinate of the crossover schema U assigns the parent from which the kth coordinate is inherited. The distribution of a randomly produced individual $I^C(n)$ is defined by

$$P(I^C(n) = i \mid I^{S_1}(n) = j, \; I^{S_2}(n) = k) := P(\; (j \odot U) \oplus (k \odot \bar{U}) \; = i).$$

Provided that a coordinate of U takes the value 1, that coordinate of the offspring i is taken from the first selected individual $I^{S_1}(n) = j$. If a coordinate of U takes the value 0, that coordinate is taken from the other parent vector $I^{S_2}(n) = k$. In this model, two parent vectors only produce one offspring. Often two offsprings are produced, but the second offspring depends stochastically on the first. This leads to another, more complicated model.

Example 1.

1. *One-point crossover.* This kind of crossover divides a parent string into two substrings. The substrings of two parents are combined into one new individual, the offspring. In a mathematical description U is uniformly distributed on the set

$$\Omega := \{(0,\ldots,0,1), (0,\ldots,0,1,1), (0,\ldots,0,1,1,1), \ldots, (1,1,\ldots,1)\},$$

that is, the right substring is taken from $I^{S_1}(n)$ and the left substring is taken from $I^{S_2}(n)$.

2. *Two-point crossover.* One block of $I^{S_1}(n)$ is inherited and the remainder of the coordinates are taken from $I^{S_2}(n)$, that is, U is uniformly distributed on the set

$$\begin{aligned}\Omega := \{&(0,\ldots,0,0,1), (0,\ldots,0,0,1,0), \ldots, (1,0,0,\ldots,0),\\ &(0,\ldots,0,1,1), (0,\ldots,0,1,1,0), \ldots, (1,1,0,\ldots,0),\\ &\qquad\vdots \qquad\quad ,\ldots, \qquad\quad \vdots\\ &(1,\ldots,1)\}.\end{aligned}$$

3. *Generalized-uniform crossover.* Each coordinate is taken independently with probability $0 \leq a \leq 1$ from $I^{S_1}(n)$. With $u \in \{0,1\}^K$ and $|u| := \sum_{k=1}^{K} u_k$ the counting measure of U is given by

$$P(U = u) = a^{|u|} * (1-a)^{K-|u|}.$$

For $a := 0.5$ generalized uniform crossover reduces to the well known ordinary uniform crossover.

4. *Halfcross.* This new kind of crossover has not been mentioned in the literature before. Theoretical aspects (presented in section 5) and good experimental results in a lot of applications leads to the design of halfcross. For K even we take U uniformly distributed on the set

$$\Omega := \left\{ u \in \{0,1\}^K : \; |u| = \frac{K}{2} \right\}$$

and for K odd take U uniformly distributed on the set

$$\Omega := \left\{ u \in \{0,1\}^K : \; |u| = \frac{K-1}{2} \vee |u| = \frac{K+1}{2} \right\}.$$

This kind of crossover takes half of the coordinates from each parent string for the production of an offspring.

Further examples of crossover operators are given in [3, 4, 10, 17, 18, 19].

Given a population $x(n)$, it is not possible to produce all elements of Ξ with crossover. Crossover only mixes up the coordinates of individuals, it is not possible to produce coordinate values not present in individuals of $x(n)$. Therefore, the necessity of another production operator, mutation, is obvious.

Step 3: *Mutation.* The coordinates of an individual j are altered independently with a small probability $0 \leq \mu \leq 1$. By i_k we denote the kth coordinate of the individual i. If mutation occurs in the kth coordinate, the value of the coordinate is changed to one of the $|\Xi_k| - 1$ remaining values. Each of the values is taken with equal probability. The distribution of a randomly mutated individual $I^M(n)$ is given by

$$P(I^M(n) = i \mid I^C(n) = j) = \prod_{k=1}^{K} (1-\mu)^{\mathbf{1}_{\{i_k = j_k\}}} * \left(\frac{\mu}{|\Xi_k| - 1} \right)^{\mathbf{1}_{\{i_k \neq j_k\}}}$$

where we denote the indicator function of a set A by $\mathbf{1}_{\{A\}}$. If i_k is not equal to j_k, mutation has to occur in the kth coordinate and has to produce the value i_k. Otherwise the coordinate must not be changed.

Step 4: *Production of the next generation.* The next generation is produced by reiteration of the first three steps. N individuals are produced with selection, crossover, and mutation to fill a new population of size N. Because the iteration occurs independently the new population is given by a polynomial distributed random vector

$$p_M(n,i) := P(I^M(n) = i \mid X(n) = x(n)).$$

The distribution of the next population is

$$P_{N*X(n+1)|X(n)=x(n)} = M(N,\ (p_M(n,i))_{i\in\Xi}). \tag{2}$$

This step leads to an unwieldy mathematical description of the GA. This step is unique of the four because it needs a vector representation. The other steps are described by functions in which individuals appear, therefore authors often neglect this step. Production of the next generation corresponds to an approximation model with an infinite population size. The interrelations of the approximation model with the model of finite population size was first proved in [16]. The simplification leads to a mathematical model that is suitable to handle. In the remainder of this paper we consider the approximation model with infinite population size.

3. Repeated selection

In this section we consider the selection step without crossover and mutation. For the model of infinite population size and repeated application of selection the iteration of equation (1) is solved by

$$P(I^S(n) = i \mid X(1) = x(1)) = \frac{f^n(i) * x_i(1)}{\sum_{j\in\Xi}\ f^n(j) * x_j(1)}. \tag{3}$$

As this solution is already known [1, 8], we consider the solution more precisely. Infinite reiteration of the selection step often leads to convergence into a population consisting only of individuals with maximal fitness.

Theorem 1. *Denote the set of individuals which occur in the population* $x(1)$ *with* $\Xi_{x(1)} := \{i \mid x_i(1) > 0\}$ *and let* $\Xi^{\max}_{x(1)}$ *be the set of individuals with maximal fitness in the population* $x(1)$*, that is,*

$$\Xi^{\max}_{x(1)} := \{i \in \Xi_{x(1)} \mid f(i) = \max_{j\in\Xi_{x(1)}} f(j)\}.$$

The limit of $I^S(n)$ *with* $n \to \infty$ *is given by*

$$P_i(\infty) := \lim_{n\to\infty} P(I^S(n) = i \mid X(1) = x(1)\ ) = \begin{cases} \frac{x_i(1)}{\sum_{j\in\Xi^{\max}_{x(1)}} x_j(1)} & i \in \Xi^{\max}_{x(1)} \\ 0 & \textit{otherwise.} \end{cases}$$

Define the value of maximal fitness $f_{\max} := f(i^{\max})$ *with* $i^{\max} \in \Xi^{\max}_{x(1)}$*. The speed of convergence can be measured by the proportion of maximal fitness to second best fitness:*

$$\lambda := \min\left\{\frac{f_{\max}}{f(i)} \mid i \notin \Xi^{\max}_{x(1)},\ x_i(1) > 0\right\}.$$

There exists a constant $C > 0$*, such that*

$$\|\ P_{I^S(n)|X(1)=x(1)} - P(\infty)\ \| \le\ C * \lambda^{-n}.$$

Proof. See the appendix. ■

λ is an important constant for scaling the fitness function. If λ is too big, the GA converges into a population with N copies of one individual within a few steps. If λ is too small, individuals with high fitness may not be selected and the effect of evolution does not happen. Therefore it is necessary to scale the fitness function such that the selection operator coorperates with the production operators. In the theory of selection, how to choose a suitable λ is an open problem.

In most cases, the size of a finite population used in an application is essentially less than the size of the search space. Therefore there is a high probability that the individuals of the initial population are all different. Under this assumption the distance of the nth population from the limit distribution can be calculated exactly.

Corollary 1. *Let $x(1)$ be the uniform distribution on N elements of Ξ, that is,*

$$\exists j_1, ..., j_N \in \Xi : \ \forall i \in \Xi : \ x_i(1) = \begin{cases} \frac{1}{N} & \exists k \in \{1, ..., N\} : \ i = j_k \\ 0 & \textit{otherwise.} \end{cases}$$

The distance of the nth population from its limit distribution is given by

$$\| \ P_{I^S(n)|X(1)=x(1)} - P(\infty) \ \| \ = 2 - 2 * |\Xi_{x(1)}^{\max}| * \frac{f_{\max}^n}{\sum_{j \in \Xi_{x(1)}} f^n(j)}. \tag{4}$$

Proof. Consider the inequality

$$\frac{f_{\max}^n}{\sum_{j \in \Xi_{x(1)}} f^n(j)} = \frac{f_{\max}^n}{|\Xi_{x(1)}^{\max}| * f_{\max}^n + \sum_{j \notin \Xi_{x(1)}^{\max}} f^n(j)} \leq \frac{1}{|\Xi_{x(1)}^{\max}|}. \tag{5}$$

With this inequality the distance can be calculated directly:

$$\begin{aligned} & \| \ P_{I^S(n)|X(1)=x(1)} - P(\infty) \ \| \\ & = \sum_{i \notin \Xi_{x(1)}^{\max}} |P_{I^S(n)|X(1)=x(1)}(i)| + \sum_{i \in \Xi_{x(1)}^{\max}} \left| \frac{f_{\max}^n}{\sum_{j \in \Xi_{x(1)}} f^n(j)} - \frac{1}{|\Xi_{x(1)}^{\max}|} \right| \\ & \underbrace{=}_{(5)} \sum_{i \notin \Xi_{x(1)}^{\max}} | \ P_{I^S(n)|X(1)=x(1)}(i)| + |\Xi_{x(1)}^{\max}| * \left(\frac{1}{|\Xi_{x(1)}^{\max}|} - \frac{f_{\max}^n}{\sum_{j \in \Xi_{x(1)}} f^n(j)} \right) \\ & = 1 + \sum_{i \notin \Xi_{x(1)}^{\max}} |P_{I^S(n)|X(1)=x(1)}(i)| - \sum_{i \in \Xi_{x(1)}^{\max}} |P_{I^S(n)|X(1)=x(1)}(i)| \\ & = 2 - 2 * |\Xi_{x(1)}^{\max}| * \frac{f_{\max}^n}{\sum_{j \in \Xi_{x(1)}} f^n(j)}. \ \blacksquare \end{aligned}$$

Some results for the distance of a population from its limit distribution with different scalings are shown in Figure 1. The model with infinite population size seems to be a very good approximation for a real algorithm with a population size of 1000. The plots show that the scaling decisively changes the speed of convergence.

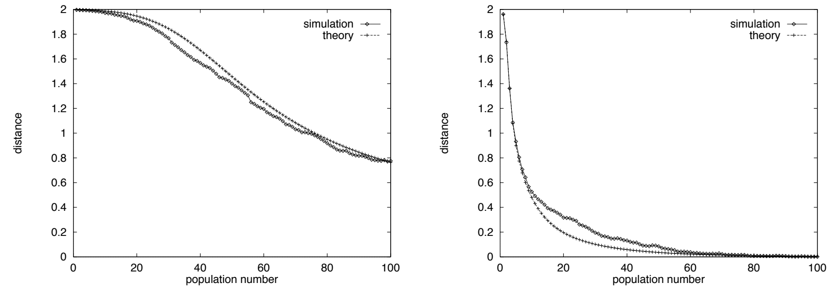

Figure 1: Distance of nth population from its limit distribution. The calculated values of equation (4) are compared with simulations. In the plots, populations of 1000 individuals are used. In the left plots, fitness values are randomly chosen such that $\lambda = 1.01$. In the right plots, the fitness values are transformed such that $\lambda = 2$, that is, the scaling has changed. The points of the simulation runs are means of 10 repetitions.

4. Repeated mutation

Mutation is often used as a background operator (e.g., [9]). Mutation is of essential importance because a GA without mutation may converge into a local maximum which is nonglobal. Without mutation the result of an optimization run with a GA depends heavily on the initial generation. Because of the necessity of its use, we consider the mutation operator theoretically. In this section we discuss repeated mutation without selection pressure, that is, individuals are selected uniformly distributed on Ξ:

$$\forall i \in \Xi : \; P(I^S(n) = i \mid X(n) = x(n)) = x_i(n).$$

Theorem 2 gives the solution of the mutation process.

Theorem 2. *Let $x(1)$ be a population and $n \in \mathbb{N}$:*

$$P(I^M(n) = i \mid X(1) = x(1)) = \sum_{j\in\Xi} x_j(1) * \prod_{k=1}^{K} \left[\frac{1}{|\Xi_k|} + \left(\mathbf{1}_{\{i_k=j_k\}} - \frac{1}{|\Xi_k|} \right) * \left(1 - \mu * \frac{|\Xi_k|}{|\Xi_k| - 1} \right)^n \right]. \quad (6)$$

Proof. See the appendix. ■

Coordinates are altered independently and all values are received with equal probability. Therefore the distribution of $I^M(n)$ converges into an uniform distribution on Ξ.

Corollary 2. *Let $x(1)$ be a population and $n \in \mathbb{N}$:*

$$\lim_{n\to\infty} P(I^M(n) = i \mid X(1) = x(1)) = \frac{1}{|\Xi|}.$$

The convergence is monotone in the sense

$$\| \; P_{I^M(n+1)|X(1)=x(1)} - \frac{1}{|\Xi|} \; \| \le \| \; P_{I^M(n)|X(1)=x(1)} - \frac{1}{|\Xi|} \; \| .$$

Proof. First we prove the assertion that mutation does not change an uniformly distributed population on Ξ. Let $x_i = \frac{1}{|\Xi|}$ for all $i \in \Xi$:

$$\begin{aligned}
&P(I^M(n) = i \mid X(1) = x(1)) \\
&\quad = \frac{1}{|\Xi|} \sum_{j \in \Xi} \prod_{k=1}^{K} \left[\frac{1}{|\Xi_k|} + \left(\mathbf{1}_{\{i_k = j_k\}} - \frac{1}{|\Xi_k|} \right) * \left(1 - \mu * \frac{|\Xi_k|}{|\Xi_k| - 1} \right)^n \right] \\
&\quad = \frac{1}{|\Xi|} \prod_{k=1}^{K} \sum_{j_k \in \Xi_k} \left[\frac{1}{|\Xi_k|} + \left(\mathbf{1}_{\{i_k = j_k\}} - \frac{1}{|\Xi_k|} \right) * \left(1 - \mu * \frac{|\Xi_k|}{|\Xi_k| - 1} \right)^n \right] \\
&\quad = \frac{1}{|\Xi|} \prod_{k=1}^{K} \left[1 + (1 - 1) * \left(1 - \mu * \frac{|\Xi_k|}{|\Xi_k| - 1} \right)^n \right] \\
&\quad = \frac{1}{|\Xi|}.
\end{aligned}$$

With these preliminaries we are now able to prove the corollary. From Theorem 2 and the inequality

$$\left| 1 - \mu * \frac{|\Xi_k|}{|\Xi_k| - 1} \right| < 1$$

the limit is obvious. The monotone convergence follows from

$$\begin{aligned}
&\sum_{i \in \Xi} \left| P(I^M(n+1) = i \mid X(1) = x(1)) - \frac{1}{|\Xi|} \right| \\
&\quad = \sum_{i \in \Xi} \left| \sum_{j \in \Xi} x_j(n) * \prod_{k=1}^{K} \left[\frac{\mu}{|\Xi_k| - 1} + \mathbf{1}_{\{i_k = j_k\}} \left(1 - \mu * \frac{|\Xi_k|}{|\Xi_k| - 1} \right) \right] - \frac{1}{|\Xi|} \right| \\
&\quad = \sum_{i \in \Xi} \left| \sum_{j \in \Xi} \left(x_j(n) - \frac{1}{|\Xi|} \right) * \prod_{k=1}^{K} \left[\frac{\mu}{|\Xi_k| - 1} + \mathbf{1}_{\{i_k = j_k\}} \left(1 - \mu * \frac{|\Xi_k|}{|\Xi_k| - 1} \right) \right] \right| \\
&\quad \le \sum_{j \in \Xi} \left| x_j(n) - \frac{1}{|\Xi|} \right| \sum_{i \in \Xi} \prod_{k=1}^{K} \left[\frac{\mu}{|\Xi_k| - 1} + \mathbf{1}_{\{i_k = j_k\}} \left(1 - \mu * \frac{|\Xi_k|}{|\Xi_k| - 1} \right) \right] \\
&\quad = \sum_{j \in \Xi} \left| x_j(n) - \frac{1}{|\Xi|} \right| . \; \blacksquare
\end{aligned}$$

The speed of convergence increases with increasing μ and decreasing $|\Xi_k|$. The higher the speed of convergence the more disorderly the coordinates get mixed. In a GA with a high rate of mixing no genetic material survives over several populations and the algorithm tends to a pure random search. Figure 2 shows a simulation result. As for the case of selection, the theoretical result seems to be a good approximation for a real algorithm with a population size of 1000.

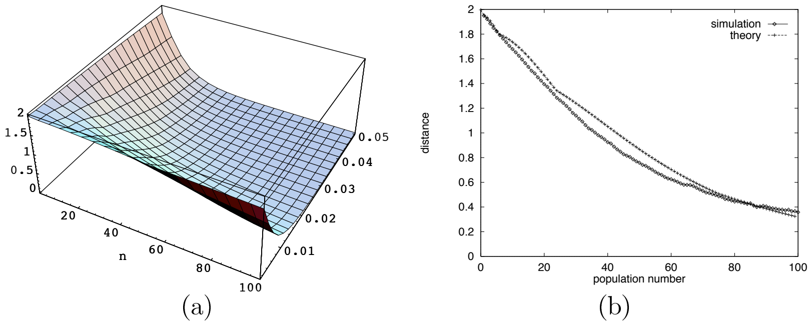

Figure 2: Distance of population from the limit distribution with $K = 9$ and $\Xi = \{0,1\}^K$. As $x(1)$, a population with one point distribution on the individual $\vec{0}i$ is taken. The points of the simulation runs are means of 10 repetitions. (a) With the help of equation (5), the calculated distance in dependence of n and μ. (b) Comparison of the calculated distance with simulations for $\mu = 0.01$.

5. Repeated crossover

The crossover operator is the most important production operator. In the literature about GAs there is no solution for repeated crossover without selection pressure and without mutation that can be found up to now. For all that, the solution can be given, by applying a result found in biology. In [11] a crossover process that arises in quantitative population genetics is considered and solved. The solution is not given in closed mathematical form and as it would waste too much space to include the complete solution here, we present a convergence result which is important for the comparison of different kinds of crossover operators.

Theorem 3. *Let U be a random variable with values in $\{0,1\}^K$ and*

$$P(U_k \neq U_l) > 0 \tag{7}$$

for all $k, l \in \{1, \ldots, K\}$ and $k \neq l$. For $n \to \infty$ the limit distribution of a randomly chosen individual after crossover is calculated by

$$\begin{aligned} P_i(\infty) &:= \lim_{n\to\infty} P(I^C(n) = i \mid X(1) = x(1)) \\ &= \prod_{k=1}^{K} P(I_k^S(1) = i_k \mid X(1) = x(1)). \end{aligned}$$

The speed of convergence can be computed by

$$\| \ P_{I^C(n)|X(1)=x(1)} - P(\infty) \ \| = C * (\lambda(P_U, K))^n$$

with the abbreviation

$$\lambda(P_U, K) = \max\{P(U_k = U_l) \mid k, l \in \{1, \ldots, K\}, \ k \neq l\}.$$

Proof. The proof is given in [20]. ■

With the assumption $P(U_k \neq U_l) > 0$ it is possible with a positive probability that two coordinates are taken from different parents. All coordinates can be separated from the others. Without the assumption, coordinates which cannot be separated may be combined into equivalent classes. The speed of convergence is influenced by the value of $\lambda(P_U, K)$, that is, the maximal interdependence between two coordinates. In some problems it may be useful to reach a maximal speed of mixing. Define

$$\lambda_K := \min_{P_U} \lambda(P_U, K), \tag{8}$$

where P_U runs over all possible kinds of crossover. Corollary 3 gives a maximal speed of convergence (maximal λ_K) dependent on the dimension K.

Corollary 3. *Let $\lfloor z \rfloor$ be the biggest integer not bigger than $z \in \mathbb{R}$. The value of λ_K defined in equation (8) is given by*

$$\lambda_K = \frac{\lfloor \frac{K-1}{2} \rfloor}{2 * \lfloor \frac{K-1}{2} \rfloor + 1} = \begin{cases} \frac{K-1}{2*K} & K \text{ odd} \\ \frac{K-2}{2*(K-1)} & K \text{ even.} \end{cases}$$

Proof. The proof is given in [11]. ■

Example 2. The speed of convergence for different kinds of crossover are given in Table 1.

In generalized-uniform crossover the speed of convergence is independent of the dimension. With $a = 0.5$ the ordinary uniform crossover reaches the minimal value $\lambda(P_U, K) = 0.5$ in this kind of crossover. The theoretical significance of halfcross is given in the result that this kind of crossover has maximal speed of convergence. This can be interpreted to mean that halfcross mixes the coordinates best.

Table 1: Speed of convergence for different kinds of crossover.

Kind of crossover	$\lambda(P_U, K)$
one-point	$1 - \frac{1}{K}$
two-point	$1 - \frac{2}{K-1}$
generalized-uniform	$1 - 2a + 2a^2$
uniform	$\frac{1}{2}$
halfcross	λ_K

Halfcross is not the only kind of crossover that fulfills

$$\lambda(P_U, K) = \lambda_K. \tag{9}$$

Theorem 4 characterizes all kinds of crossover fulfilling equation (9). We formulate and prove the result for even dimensions K. For odd dimensions the transfer is obvious.

Theorem 4. *Let K (even) and $u \in \{0,1\}^K$. With a crossover schema U consider the following three conditions.*

$$1. \lambda(P_U, K) = \lambda_K.$$
$$2. \forall k,\ l \in \{1, \dots, K\},\ k \neq l :\ P(U_k = U_l) = \lambda_K.$$
$$3. P\left(|U| \neq \frac{K}{2}\right) = 0.$$

The following relations appear

$$1. \Leftrightarrow 2. \mathrel{\substack{\Rightarrow \\ \not\Leftarrow}} 3.$$

Proof. The proof is given in the appendix. ■

Condition 1, which says that a kind of crossover takes the maximal speed of convergence, is equivalent to condition 2, which says that the interdependences between all coordinates are equal and take the value λ_K. In uniform crossover the probabilities $P(U_k = U_i)$ are all equal, but $\lambda(P_U, K)$ takes the nonmaximal value 0.5. For example, halfcross fulfills the condition. The third one is a necessary but not sufficient condition. A kind of crossover with maximal speed of convergence only consists of crossover schemas which take half of the coordinates from each of the parents. An example against the implication $3 \not\Rightarrow 2$ is given in Example 3.

Example 3. Let U be one point distributed on

$$\{ \underbrace{0, \ \dots \ , 0}_{K/2}, \underbrace{1, \ \dots \ , 1}_{K/2} \}.$$

This distribution fulfills condition 3 but not 1.

Figure 3 shows a comparison of different kinds of crossover. In finite populations perturbations of the realization of I^C arise. Therefore the distance does not converge to zero. Because of the random perturbations not all elements of the search space are produced for one population simultaneously, other individuals are more often present than the probability of their production is. Nevertheless the figure shows the different speeds of convergence which influence the first 15 generations.

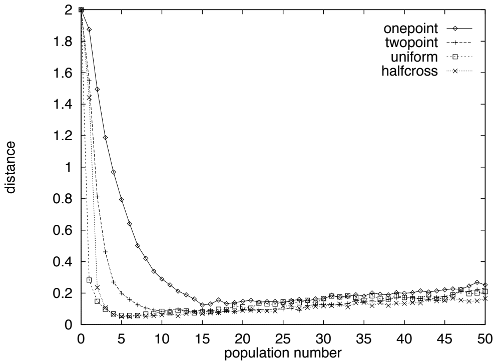

Figure 3: Distance of nth population from its limit distribution. Simulation runs with different kinds of crossover are compared. As search space Ξ, the set of bitstrings of length 8 is used, that is, $|\Xi| = 256$. An initial population of size 1000 is filled with 500 copies of the individual 00000000 and with 500 copies of the individual 11111111. Therefore the limit distribution is the uniform distribution on Ξ. The graphed values are means of 10 repetitions.

6. Summary and future research

The results presented in this paper help to gain insight into the GA. We explain some dependences of the operators from several parameters. Although we consider the operators separately, the need for a simultaneous tuning of all parameters is obvious. The three parameters:

$$\lambda_{\text{sel}} := \min\left\{\frac{f^{\max}}{f(i)} \mid i \notin \Xi^{\max},\ x_i(1) > 0\right\}$$

for selection,

$$\lambda_{\text{mut}} := \left| 1 - \mu * \frac{|\Xi_k|}{|\Xi_k| - 1} \right|$$

for mutation, and

$$\lambda_{\text{cross}} := \max\{P(U_k = U_l) \mid k, l \in \{1, ..., K\},\ k \neq l\}$$

for crossover are of significant importance for the algorithm. The choice of these three parameters strongly influences the GA. The difficulty in the choice of the parameters lies in the different attributes of the three operators. λ_{sel} influences the dominance of some individuals. This dominance may be destroyed by mutation influenced by λ_{mut}. On the other hand mutation produces new individuals without any regard to their fitness. The operators should constitute a tension of exploration and exploitation. These relations may differ for different applications of the GA. The kind of disorderly mixing by crossover is of different importance for different applications, too. In some applications the kind of crossover which should be used may depend on the choice of population size or other parameters [21]. Therefore the examinations in this paper may only constitute a foundation for the understanding of the GA. In the future, research may result in more structural considerations for the interconnection of the parameters.

Appendix

In this appendix we present the remaining proofs.

Proof (Theorem 1). First we calculate the limit:

$$\begin{aligned}
\lim_{n\to\infty} P(I^S(n)=i \mid X(1)=x(1)\ ) &= \lim_{n\to\infty} \frac{(f(i))^n * x_i(1)}{\sum_{j\in\Xi}(f(j))^n * x_j(1)} \\
&= \lim_{n\to\infty} \frac{x_i(1)}{\sum_{j\in\Xi} \frac{(f(j))^n}{(f(i))^n} * x_j(1)} \\
&= \frac{x_i(1)}{\sum_{j\in\Xi} \left(\lim_{n\to\infty} \frac{(f(j))^n}{(f(i))^n}\right) * x_j(1)} \\
&= \begin{cases} \frac{x_i(1)}{\sum_{j\in\Xi^{\max}_{x(1)}} x_j(1)} & i \in \Xi^{\max}_{x(1)} \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

In the same manner we calculate the speed of convergence:

$$\begin{aligned}
&\sum_{i\in\Xi} \mid P(I^S(n)=i \mid X(1)=x(1)\ ) - P_i(\infty) \mid \\
&= \sum_{i\in\Xi^{\max}_{x(1)}} \left| \frac{(f^{\max}_{x(1)})^n * x_i(1)}{\sum_{j\in\Xi}(f(j))^n * x_j(1)} - \frac{x_i(1)}{\sum_{j\in\Xi^{\max}_{x(1)}} x_j(1)} \right| \\
&\quad + \sum_{i\notin\Xi^{\max}_{x(1)}} \left| \frac{(f(i))^n * x_i(1)}{\sum_{j\in\Xi}(f(j))^n * x_j(1)} \right| \\
&= \sum_{i\in\Xi^{\max}_{x(1)}} \left| \frac{(f^{\max}_{x(1)})^n * x_i(1)}{\sum_{j\in\Xi^{\max}_{x(1)}}(f^{\max}_{x(1)})^n * x_j(1) + \sum_{j\notin\Xi^{\max}_{x(1)}}(f(j))^n * x_j(1)} - \frac{x_i(1)}{\sum_{j\in\Xi^{\max}_{x(1)}} x_j(1)} \right| \\
&\quad + \sum_{i\notin\Xi^{\max}_{x(1)}} \left| \frac{x_i(1)}{\sum_{j\in\Xi^{\max}_{x(1)}}(\frac{f^{\max}_{x(1)}}{f(i)})^n x_j(1) + \sum_{j\notin\Xi^{\max}_{x(1)}}(\frac{f(j)}{f(i)})^n x_j(1)} \right| \\
&\le \sum_{i\in\Xi^{\max}_{x(1)}} \left| \frac{\sum_{j\notin\Xi^{\max}_{x(1)}}(\frac{f(j)}{f^{\max}_{x(1)}})^n x_j(1)}{\sum_{j\in\Xi^{\max}_{x(1)}} x_j(1) + \sum_{j\notin\Xi^{\max}_{x(1)}}(\frac{f(j)}{f^{\max}_{x(1)}})^n x_j(1)} \right| * \frac{x_i(1)}{\sum_{j\in\Xi^{\max}_{x(1)}} x_j(1)} \\
&\quad + \sum_{i\notin\Xi^{\max}_{x(1)}} \left| \frac{x_i(1)}{\lambda^n * \sum_{j\in\Xi^{\max}_{x(1)}} x_j(1) + \sum_{j\notin\Xi^{\max}_{x(1)}}(\frac{f(j)}{f(i)})^n x_j(1)} \right| \\
&\le \lambda^{-n} * \left(2 * \frac{\sum_{j\notin\Xi^{\max}_{x(1)}} x_j(1)}{\sum_{j\in\Xi^{\max}_{x(1)}} x_j(1)}\right). \ \blacksquare
\end{aligned}$$

Proof (Theorem 2). The proof follows with induction. For $n = 1$ equation (6) holds. Let the assertion be valid for $n \in \mathbb{N}$:

$$\begin{aligned}
&P(I^M(n+1) = z \mid X(1) = x(1)\ ) \\
&= \sum_{i\in\Xi} P(I^M(n+1) = z \mid I^M(n) = i) * P(I^M(n) = i \mid X(1) = x(1)\ ) \\
&= \sum_{j\in\Xi} x_j(1) \prod_{k=1}^{K} \sum_{i_k} \left[\frac{\mu}{|\Xi_k|-1} + \mathbf{1}_{\{i_k=z_k\}}\left(1-\mu*\frac{|\Xi_k|}{|\Xi_k|-1}\right)\right] \\
&\qquad\qquad * \left[\frac{1}{|\Xi_k|} + \left(\mathbf{1}_{\{i_k=j_k\}} - \frac{1}{|\Xi_k|}\right) * \left(1-\mu*\frac{|\Xi_k|}{|\Xi_k|-1}\right)^{n}\right] \\
&= \sum_{j\in\Xi} x_j(1) \prod_{k=1}^{K} \left\{\mathbf{1}_{\{j_k=z_k\}}\left[(1-\mu)*\left(\frac{1}{|\Xi_k|} + \left(1-\frac{1}{|\Xi_k|}\right)*\left(1-\mu*\frac{|\Xi_k|}{|\Xi_k|-1}\right)^{n}\right)\right.\right. \\
&\qquad\qquad \left. +\mu*\left(\frac{1}{|\Xi_k|} - \frac{1}{|\Xi_k|}*\left(1-\mu*\frac{|\Xi_k|}{|\Xi_k|-1}\right)^{n}\right)\right] \\
&\qquad\qquad +\mathbf{1}_{\{j_k\neq z_k\}}\left[(1-\mu)*\left(\frac{1}{|\Xi_k|} - \frac{1}{|\Xi_k|}*\left(1-\mu*\frac{|\Xi_k|}{|\Xi_k|-1}\right)^{n}\right)\right. \\
&\qquad\qquad +\frac{\mu}{|\Xi_k|-1}\left(\frac{1}{|\Xi_k|} + \left(1-\frac{1}{|\Xi_k|}\right)*\left(1-\mu*\frac{|\Xi_k|}{|\Xi_k|-1}\right)^{n}\right) \\
&\qquad\qquad \left.\left. +(|\Xi_k|-2)\frac{\mu}{|\Xi_k|-1}\left(\frac{1}{|\Xi_k|} - \frac{1}{|\Xi_k|}*\left(1-\mu*\frac{|\Xi_k|}{|\Xi_k|-1}\right)^{n}\right)\right]\right\} \\
&= \sum_{j\in\Xi} x_j(1) \prod_{k=1}^{K} \left\{\frac{1}{\Xi_k} + \mathbf{1}_{\{j_k=z_k\}}\left[\left(1-\mu-\frac{1}{|\Xi_k|}\right)*\left(1-\mu*\frac{|\Xi_k|}{|\Xi_k|-1}\right)^{n}\right]\right. \\
&\qquad\qquad \left. +\mathbf{1}_{\{j_k\neq z_k\}}\left[-\frac{1}{|\Xi_k|}*\left(1-\mu*\frac{|\Xi_k|}{|\Xi_k|-1}\right)^{n+1}\right]\right\} \\
&= \sum_{j\in\Xi} x_j(1) \prod_{k=1}^{K} \left\{\frac{1}{\Xi_k} + \mathbf{1}_{\{j_k=z_k\}}\left[\left(1-\frac{1}{|\Xi_k|}\right)*\left(1-\mu*\frac{|\Xi_k|}{|\Xi_k|-1}\right)^{n+1}\right]\right. \\
&\qquad\qquad \left. +\mathbf{1}_{\{j_k\neq z_k\}}\left[-\frac{1}{|\Xi_k|}*\left(1-\mu*\frac{|\Xi_k|}{|\Xi_k|-1}\right)^{n+1}\right]\right\} \\
&= \sum_{j\in\Xi} x_j(1) \prod_{k=1}^{K} \left[\frac{1}{\Xi_k} + \left(\mathbf{1}_{\{j_k=z_k\}} - \frac{1}{|\Xi_k|}\right)*\left(1-\mu*\frac{|\Xi_k|}{|\Xi_k|-1}\right)^{n+1}\right]. \blacksquare
\end{aligned}$$

Proof (Theorem 4). For fixed u the sum $\sum_{k\neq l} P(U_k = U_l \mid U = u)$ counts the number of possibilities for choosing two coordinates of u with the same value:

$$\sum_{k\neq l} P(U_k = U_l \mid U = u) = \frac{|u|*(|u|-1)}{2} + \frac{(K-|u|)*(K-|u|-1)}{2}$$

$$\begin{aligned}
&= \frac{K*(K-1)}{2} - \underbrace{|u|*(K-|u|)}_{\leq\left(\frac{K}{2}\right)^2} \\
&\geq \frac{K*(K-2)}{4}.
\end{aligned}$$

From this consideration we get the inequalities:

$$\begin{aligned}
\frac{K*(K-1)}{2}*\lambda &\geq \sum_{k\neq l} P(U_k = U_l) \\
&= \sum_{u\in\{0,1\}^K} \sum_{k\neq l} P(U_k = U_l \mid U = u) * P(U = u) \\
&\geq \min\left\{ \sum_{k\neq l} P(U_k = U_l \mid U = u) \mid u \in \{0,1\}^K \right\} \\
&= \frac{K*(K-2)}{4}.
\end{aligned}$$

1. ⇒ 2. Let condition 1 be true. In this case, the inequalities given above are equalities. The sum $\sum_{k\neq l}$ consists of $K*(K-1)/2$ terms and the maximum of $P(U_k = U_l)$ is given by λ. From this consideration condition 2 follows.

1. ⇒ 3. Let condition 1 be true. In this case, the inequalities given above are equalities. Especially with a constant $\sum_{k\neq l} P(U_k = U_l \mid U = u) = c$ follows for all $u \in \{0,1\}^K$. This is only possible if the minimum is received for all U with positive probability. Condition 3 follows.

2. ⇒ 1. Follows from the definition of λ. ∎

References

[1] C. A. Ankenbrandt, "An Extension to the Theory of Convergence and a Proof of the Time Complexity of Genetic Algorithms," in *Foundations of Genetic Algorithms* edited by G. J. E. Rawlins (Morgan Kaufmann, San Mateo, CA, 1991).

[2] J. E. Baker, "Adaptive Selection Methods for Genetic Algorithms," in *Proceedings of the International Conference on Genetic Algorithms* edited by J. J. Grefenstette (Lawrence Erlbaum, Hillsdale, NJ, 1985).

[3] L. B. Booker, "Recombination Distributions for Genetic Algorithms," in *Foundations of Genetic Algorithms 2* edited by L. D. Whitley (Morgan Kaufmann, San Mateo, CA, 1993).

[4] Y. Davidor, "Analogous Crossover," in *Proceedings of the International Conference on Genetic Algorithms* edited by D. Schaffer (Morgan Kaufmann, San Francisco, CA, 1989).

[5] L. Davis, *Handbook of Genetic Algorithms* (Van Nostrand Rheinhold, New York, 1990).

[6] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison Wesley, New York, 1989).

[7] D. E. Goldberg, "A Note on Boltzmann Tournament Selection for Genetic Algorithms and Population-oriented Simulated Annealing," *Complex Systems*, **4** (1990) 445–460.

[8] D. E. Goldberg and K. Deb, "A Comparative Analysis of Selection Schemes used in Genetic Algorithms," in *Foundations of Genetic Algorithms* edited by G. J. Rawlins (Morgan Kaufmann, San Mateo, CA, 1991).

[9] J. H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, 1975).

[10] K. A. De Jong, "An Analysis of the Interacting Roles of Population Size and Crossover in Genetic Algorithms," in *Parallel Problem Solving from Nature* edited by H. P. Schwefel and R. Maenner (Springer Verlag, New York, 1990).

[11] Y. I. Lyubich, *Mathematical Structures in Population Genetics* (Springer Verlag, 1992).

[12] S. W. Mahfoud, "Finite Markov Chain Models of an Alternative Selection Strategy for the Genetic Algorithm," Illigal Report, 91007 (Illinois Genetic Algorithms Laboratory, Urbana, 1991).

[13] S. W. Mahfoud, "An Analysis of Boltzmann Tournament Selection," Illigal Report, 94007 (Illinois Genetic Algorithms Laboratory, Urbana, 1994).

[14] M. de Maza and B. Tidor, "An Analysis of Selection Procedures with Particular Attention Paid to Proportional and Boltzmann Selection," in *Proceedings of the International Conference on Genetic Algorithms* edited by S. Forrest (Morgan Kaufmann, San Francisco, CA, 1993).

[15] K. Nachtigall and S. Voget, "A Genetic Algorithm Approach to Periodic Railway Synchronization," *Computers and Operations Research*, **23** (1996) 453–463.

[16] A. E. Nix and M. D. Vose, "Modeling Genetic Algorithms with Markov Chains," *Annals of Mathematics and Artificial Intelligence*, **5** (1992) 79–88.

[17] P. W. Poon and J. N. Carter, "Genetic Algorithm Crossover Operators for Ordering Applications," *Computer Operations Research*, **22** (1995) 135–147.

[18] W. M. Spears and K. A. De Jong, "An Analysis of Multipoint Crossover," in *Foundations of Genetic Algorithms* edited by L. D. Whitley (Morgan Kaufmann, San Mateo, CA, 1993).

[19] G. Syswerda, "Uniform Crossover in Genetic Algorithms," in *Proceedings of the International Conference on Genetic Algorithms* edited by D. Schaffer (Morgan Kaufmann, San Francisco, CA, 1989).

[20] S. Voget, *Aspekte genetischer Optimierungsalgorithmen: mathematische Modellierung und Einsatz in der Fahrplanerstellung*, Thesis, Universität Hildesheim, 1995.

[21] S. Voget, "Speeding up of a Genetic Algorithm for the Optimization of Periodic Networks," *Hildesheimer Informatik Berichte*, **21** (1995) 1–19.

[22] M. D. Vose, "Modeling Simple Genetic Algorithms," in *Foundations of Genetic Algorithms 2* edited by L. D. Whitley (Morgan Kaufmann, San Francisco, CA, 1993).