



On asymptotic behaviour of a binary genetic algorithm

Witold Kosiński^{1,2*}, Stefan Kotowski³, Jolanta Socła⁴

¹*Research Center, Department of Intelligent Systems, Polish-Japanese Institute of Information Technology, Koszykowa 86, 02-008 Warszawa, Poland*

²*Institute of Environmental Mechanics and Applied Computer Science, Kazimierz Wielki University, Chodkiewicza 30, 85-064 Bydgoszcz, Poland*

³*Institute of Fundamental Technological Research, IPPT PAN, Świętokrzyska 21, 00-950 Warszawa, Poland*

⁴*Institute of Mathematics, Silesian University, Bankowa 14, 40-007 Katowice, Poland*

Abstract

The simple genetic algorithm (SGA) and its convergence analysis are main subjects of the article. A particular SGA is defined on a finite multi-set of individuals (chromosomes) together with mutation and proportional selection operators, each of which with some prescribed probability. The selection operation acts on the basis of the fitness function defined on individuals. Generation of a new population from a given one is made by iterative actions of those operators. Each iteration is written in the form of a transition operator acting on probability vectors which describe probability distributions of all populations. The transition operator is power of Markovian matrix. Based on the theory of Markov operators [1-3] new conditions for asymptotic stability of the transition operator are formulated.

1. Introduction

In the last two decades there has been growing interest in universal optimization methods realized by genetic and evolutionary algorithms. These algorithms use only limited knowledge about problems to be solved and are constructed on the basis of some similarity to the processes in nature. Extensive application of those methods in practical solutions of complex optimal problems cause a need to their develop theoretical foundations. The question of their convergence properties is one of the most important issues [4-9].

2. Preliminaries

The genetic (GA) as well as the evolutionary algorithms (EG) perform multi-directional search by maintaining a population of potential solutions, called

*Corresponding author: *e-mail address*: wkos@pjwstk.edu.pl

individuals, and encourage information formation and exchange between these directions. A population, i.e. a set of individuals, undergoes a simulated evolution with a number of steps. In the most general case the evolution is due to an iterative action – with some probability distributions – of a composition of three operators: mutation, crossover and selection. If a population is regarded as a point in the space Z of (encoded) potential solutions then the effect of one iteration of this composition is to move that population to another point. In this way the action of GA as well as EA is a discrete (stochastic) dynamical system. In the paper we use the term *population* in two meanings; in the first it is a finite multi-set (a set with elements that can repeat) of individuals, in the second they are frequency vector components of which fractions are composed, i.e. the ratio of the number of copies of each element $z_k \in Z$ to the total population size. The action of that composition is a random operation on populations.

In the paper we deal with a particular case of the simple genetic algorithm (SGA) in which the mutation follows the fitness proportional selection and the crossover is not present. In the case of the binary genetic algorithm (BGA) the mutation can be characterized by the bitwise mutation rate μ – the probability of the mutation of one bit of a chromosome. In SGA with the known fitness function the fitness proportional selection can be treated as a multiplication of each component of the frequency vector by the quotient of the fitness of the corresponding element to the average fitness of the population. This allows to write the probability distribution for the next population in the form of the product of the diagonal matrix with the population (frequency) vector. Moreover, results of the mutation can also be written as a product of another matrix with the population (probability) vector. Finally the composition of both operations is a matrix (cf.(10)), which leads to the general form of the transition operator (cf.(12)) acting on a new probability vector representing a probability distribution of appearance of all populations of the same size equal to the population size *PopSize*. The matrix appearing there turns to be Markovian and each subsequent application of SGA is the same as the subsequent composition of that matrix with itself (cf.(13)). In the paper owing to the well-developed theory of Markov operator([1-3,10]) new conditions for the asymptotic stability of the transition operator are formulated and some conclusions are drawn.

3. Frequency and population vector

In the case of BGA the set of individuals

$$Z = \{z_0, \dots, z_{s-1}\}$$

are *chromosomes* and they form all binary l -element sequences. For a better description one orders them and the set Z with $s = 2^l$, becomes a list, in which its typical element (chromosome) is of the form $z_j = \{0, 0, 1, 0, \dots, 1, 0, 0\}$.

At first by a *population* we understand any multi-set of r chromosomes from Z , then r is the population size: *PopSize*.

Definition 1. By a frequency vector of population we understand the vector

$$p = (p_0, \dots, p_{s-1}), \quad \text{where } p_k = \frac{a_k}{r} \quad (1)$$

where a_k is a number of copies of the element z_k .

The set of all possible populations now understood in the other meaning as frequency vectors is

$$\Lambda = \left\{ p \in \mathbb{R}^s : p_k \geq 0, p_k = \frac{d}{r}, d \in \mathbb{N}, \sum_{k=0}^{s-1} p_k = 1 \right\}. \quad (2)$$

When GA is realized by an action of the so-called transition operator on a given population, a new population is generated. Since the transition between two subsequent populations is random and is realized by a probabilistic operator, then if one starts with a frequency vector, a probabilistic vector can be obtained, in which p_i may not be rational any more. Hence for our analysis the closure of the set Λ , namely

$$\overline{\Lambda} = \left\{ x \in \mathbb{R}^s : \forall k, x_k \geq 0, \text{ and } \sum_{k=0}^{s-1} x_k = 1 \right\}, \quad (3)$$

is more suitable.

4. Selection operator

The optimization problem at hand is characterized by a goal (or cost) function. If we transform it by a standard operation to a nonnegative function we will get the so-called *fitness function* $f : Z \rightarrow \mathbb{R}^+$. If we assume the first genetic operator is the *fitness proportional selection*, then the probability that the element z_k from a given population p will appear in the next population equals

$$\frac{f(z_k) p_k}{\overline{f}(p)}, \quad (4)$$

where $\overline{f}(p)$ is the *average population fitness* denoted by

$$\overline{f}(p) = \sum_{k=0}^{s-1} f(z_k) p_k. \quad (5)$$

Then the transition from the population p into the new one, say q , can be given by

$$q = \frac{1}{\overline{f}(p)} S p, \quad (6)$$

where the matrix S of the size s , has on its main diagonal the entries

$$S_{kk} = f(z_k). \quad (7)$$

Matrix S describes the selection operator [7-9].

5. Mutation operator

The second genetic operator considered the binary uniform mutation with a parameter μ as the probability of changing bits 0 into 1 or vice versa. If the chromosome z_i differs from z_j at c positions then the probability of mutation of the element z_j into the element z_i is

$$U_{ij} = \mu^c (1 - \mu)^{l-c}. \quad (8)$$

Then we may define a matrix

$$U = [U_{ij}] ,$$

with U_{ij} as in (8) and U_{ii} – the probability of the surviving of the element (individual) z_i . In general one requires

$$U_{ij} \geq 0 ,$$

$$\sum_{i=0}^{s-1} U_{ij} = 1, \quad \text{for all } j. \quad (9)$$

6. Transition operator

When we have the specific population p , then it means p is a frequency vector and $p \in \Lambda$. If the mutation and selection (random) operators are applied to it they could lead p out the set Λ . The action of the genetic algorithm in the first and all subsequent steps is the following: if we have a given population p then we sample with returning r -elements from the set Z , and the probability of sampling the elements z_0, \dots, z_{s-1} is described by the vector $G(p)$, where

$$G(p) = \frac{1}{f(p)} USp. \quad (10)$$

This r -element vector is our new population q .

Let us denote by W the set of all possible r -element populations composed of the elements selected from the set Z , where elements in the population could be repeated. This set is finite and let its cardinality be M . It can be shown that the number M is given by a combinatoric formula, cf. [12]. Let us order all populations, then we identify the set W with the list $W = \{w^1, \dots, w^M\}$. Typical w^k , $k = 1, 2, \dots, M$ is some population for which we used the notation p in the previous section. That population will be identified with its frequency vector or probabilistic vector. This means that for a given population $p = w^k = (w_0^k, \dots, w_{s-1}^k)$, the number w_i^k , for $i \in \{0, \dots, s-1\}$, denotes the probability of sampling from the population w^k the individual z_i . If p is a frequency vector then the number w_i^k is the fraction of the individual z_i in the population w^k .

Beginning our implementation of BGA from an arbitrary population $p = w^k$ in the next stage each population w^1, \dots, w^M can appear with the probability, which can be determined from our analysis. In particular, if in the next stage the population has to be q , with the position l on our list W (it means $q = w^l$), then this probability [8,11,12] is equal to

$$r! \prod_{j=0}^{s-1} \frac{(\mathcal{G}(p)_j)^{rq_j}}{(rq_j)!}. \quad (11)$$

After two steps, every population w^1, \dots, w^M will appear with some probability, which is a double composition of this formula. It will be analogously in the third step and so on. This formula gives a possibility of determining all elements of a matrix T which defines the probability distribution of appearance of populations in the next steps, if we have current probability distribution of the populations. With our choice of denotations for the populations p and q , the element (l,k) of the matrix will give transition probability from the population with the number k into the population with the number l . It is important that elements of the matrix are determined once forever, independently of the number of steps. The transition between elements of different pairs of populations is described by different probabilities (11) represented by different elements of the matrix. We can see that the nonnegative, square matrix T of dimension M , with elements p_{lk} , $l,k = 1,2,\dots,M$ has the property: the probability distribution of all M populations in the step t is given by the formula

$$T^t u = 0, 1, 2, \dots$$

Let us denote by

$$\Gamma = \{x \in \mathbb{R}^M : \forall k x_k \geq 0 \text{ oraz } \|x\| = 1\},$$

where $\|x\| = x_1 + \dots + x_M$, for $x = (x_1, \dots, x_M)$, the set of new M -dimensional probabilistic vectors. A particular component of the vector x represents the probability of the appearance of this population from the list W of all M populations. The set Γ is composed of all possible probability distributions for M populations. Then the described implementation transforms, in every step, the set Γ into the same.

Note that if at the beginning we start our SGA at a specific population p , which attains the place j -th on our list W , i.e. $p = w^j$, then the vector u will denote the particular situation of the population distribution in the step zero 0 if

$$u = (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^m.$$

On the set Γ the basic, fundamental *transition operator*,

$$T(\cdot) : \mathbb{N} \times \Gamma \rightarrow \Gamma \quad (12)$$

is defined. According to the above remark, the transition operator $T(t)$ is linked with the above matrix by the dependence

$$T(t) = T'. \quad (13)$$

If $u \in \Gamma$, then $T(t)u = ((T(t)u)_1, \dots, (T(t)u)_M)$ is the probability distribution for M populations in the step number t , if we have begun our implementation of SGA given by $\mathcal{G}(10)$ from the probability distribution $u = (u_1, \dots, u_M) \in \Gamma$, by t -application of this method. The number $(T(t)u)_k$ for $k \in \{1, \dots, M\}$ denotes the probability of appearance of the population w^k in the step of number t . By the definition $\mathcal{G}(p)$ in (10), (11), and the remarks made at the end of the previous section, the transition operator $T(t)$ is linear for all natural t .

Notice that though formula (11) determining individual entries (components) of the matrix T is a dependent population, and hence nonlinear, the transition operator $T(t)$ is due thanks to the order relation introduced in the set W of all M populations. The multi-index l, k of the component p_{lk} kills, in some sense, this nonlinearity, since it tells (is responsible for) a pair of populations between which the transition takes place. The matrix T is a Markovian matrix. This fact permits us to apply the Theory of Markov operators to analyze the convergence of genetic algorithms [1-3, 10].

Note that the action of the matrix T can be seen as follows. In the space of all possible populations there is a walking point, which attains its next random position numbered by $1, 2, \dots, M$, as an action of SGA on the actual population, with probabilities u_1, u_2, \dots, u_M . We know that if at the moment t (in the generation number t) we had population p with the position k on our list, i.e. the population w^k , then the probability that at the moment $t+1$ (in the generation number $t+1$) it will attain population q with the position l , on our list, i.e. the population w^l , is p_{lk} , and this probability is independent of the number of steps in which it is realized. With this denotation the probability p_{lk} is given by formula (11).

Let $e_k \in \Gamma$ be a vector which at the k -th position has one and zeroes at the other positions. Then e_k describes the probability distribution in which the population w^k is attained with the probability 1.

By the notation $T(t)w^k$ we will understand

$$T(t)w^k = T(t)e_k \quad (14)$$

which means that we begin the GA at the specific population w^k .

Further on we will assume $U_{jj} > 0$ for $j \in \{0, \dots, s-1\}$. Note that in the case of binary mutation (8) this condition will be satisfied if $0 \leq \mu < 1$.

Definition 2. We will say that the model is asymptotically stable if there exists $u^* \in \Gamma$ such that:

$$T(t)u^* = u^* \quad \text{for } t = 0, 1, \dots \quad (15)$$

$$\lim_{t \rightarrow \infty} \|T(t)u - u^*\| = 0 \quad \text{for all } u \in \Gamma. \quad (16)$$

Since for $k \in \{1, \dots, M\}$ we have

$$\left| (T(t)u)_k - u_k^* \right| \leq \|T(t)u - u^*\| \quad (17)$$

then (16) gives

$$\lim_{t \rightarrow \infty} (T(t)u)_k = u_k^*. \quad (18)$$

It means that probability of appearance of the population w^k in the step number t converges to a certain fixed number u_k^* independently of the initial distribution u . It is realized in a special case, when our implementation begins at one specific population $p = w^j$.

We can say that from the chromosome z_a it is possible to obtain z_b in one mutation step with a positive probability if $U_{ba} > 0$ and that from the chromosome z_a it is possible to get the chromosome z_b with a positive probability in n -step mutation if there exists a sequence of chromosomes z_{i_0}, \dots, z_{i_n} , such that $z_{i_0} = z_a$, $z_{i_n} = z_b$ and any z_{i_j} for $j = 1, \dots, n$ is possible to be obtained from $z_{i_{j-1}}$ in one step with a positive probability.

Definition 3. Model is pointwise asymptotically stable if there exists such a population w^j that

$$\lim_{t \rightarrow \infty} (T(t)u)_j = 1 \text{ for } u \in \Gamma. \quad (19)$$

Condition (19) denotes that in successive steps the probability of appearance of other population than w^j tends to zero. It is a special case of the asymptotic stability for which

$$u^* = e_j.$$

Theorem 1. Model is pointwise asymptotically stable if and only if there exists exactly one chromosome z_a with such a property that it is possible to attain it from any chromosome in a finite number of steps with a positive probability. In this situation the population w^j is exclusively composed of the chromosomes z_a and

$$T(t)w^j = w^j \quad (20)$$

holds. Moreover, the probability of appearance of other population than w^j tends to zero in the step number t with a geometrical rate, i.e. there exists $\lambda \in (0, 1)$, $D \in \mathbb{R}_+$ that

$$\sum_{\substack{i=1 \\ i \neq j}}^M (T(t)u)_i \leq D \cdot \lambda^t \cdot \Delta \quad (21)$$

The proofs of our theorems and auxiliary lemmas are presented in original articles [12, 13].

Numbers λ and D could be determined for a specific model. This will be the subject of next articles.

Theorem 1 states that the convergence to one population could occur only under specific assumptions. This justifies the investigation of asymptotic stability in Definition 2.

Definition 4. By an *attainable chromosome* we denote $z_a \in Z$ such that it is possible to attain it from any other chromosome in a finite number of steps with a positive probability. Let us denote by Z^* the set of all z_a with this property.

Theorem 2. Model is asymptotically stable if and only if $Z^* \neq \emptyset$. Δ

Theorem 3. Let us assume that the model is asymptotically stable. Then the next relationship holds:

(war) $u_k^* > 0$ if and only if the population w^k is exclusively composed of chromosomes belonging to the set Z^* . Δ

7. Conclusions

Here we present the summary of our results obtained in this and our other papers [11-13]:

1. If $Z^* = \emptyset$ then there is a lack of asymptotic stability.
2. If $Z^* \neq \emptyset$ then asymptotic stability holds but:
3. If cardinality $(Z^*) = 1$ then pointwise asymptotic stability (in some sense convergence to one population) holds.
4. If cardinality $(Z^*) > 1$ then asymptotic stability holds, but there is no pointwise asymptotic stability.
5. If $Z^* = Z$ then $u_k^* > 0$ for all $k \in \{1, \dots, M\}$.

REMARK. In SGA with a positive mutation probability, it is possible to attain any individual (chromosome) from any other individual. Then there is more than one chromosome which is possible to attain from any other in a finite number of steps with a positive probability. Hence, from Theorem 1, it is impossible to get the population composed exclusively of one type of chromosomes.

The last conclusion means that if any chromosome is possible to attain from any other in a finite number of steps with a positive probability then in the limit (probability distribution) of infinite number of generations each population (has a positive probability) may be reached with a positive probability.

Theorem 2 is an extension of Th.4.2.2.4_4 from [9] for the case when it is possible to attain any population in a finite number of steps, (not only in one step). It means that transition operator does not need to be positively defined, but there exists such k , that the k -th power of the transition operator possesses a column which is strongly positive. The same concerns Th.4.2.2_1 of [9] which is true only for a positively defined transition matrix.

Acknowledgement

The research presented in the paper was partially done by W.K. and S.K. in the framework of the KBN Project (State Committee for Scientific Research) No. 3 T11 C007 28. The authors are grateful to Professor Zbigniew Michalewicz for inspiration and discussions.

References

- [1] Lasota A., Yorke J.A., *Exact dynamical systems and the Frobenius-Perron operator*. Trans. Amer. Math. Soc., 273 (1982) 375.
- [2] Rudnicki R., *On asymptotic stability and sweeping for Markov operators*. Bull. Polish Acad. Sci. Math., 43 (1995) 245.
- [3] Socala J., *Asymptotic behaviour of the iterates of nonnegative operators on a Banach lattice*. Ann. Polon. Math., 68(1) (1998) 1.
- [4] Kieś P., Michalewicz Z., *Podstawy algorytmów genetycznych*. Matematyka Stosowana, Matematyka dla Społeczeństwa, 1(44) (2000) 68, in Polish.
- [5] Michalewicz Z., *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, Berlin, (1996).
- [6] Cytowski J., *Algorytmy genetyczne: podstawy i zastosowania*. Seria: Problemy Współczesnej Nauki – Teoria i Zastosowania, Akademicka Oficyna Wydawnicza PLJ, Warszawa, 18 (1996), in Polish.
- [7] Vose M.D., *The Simple Genetic Algorithm: Foundation and Theory*. MIT Press, Cambridge, MA, (1999).
- [8] Rowe J.E., *The dynamical system models of the simple genetic algorithm*. in Theoretical Aspects of Evolutionary Computing, Leila Kallel, Bart Naudts, Alex Rogers (Eds.), Springer, (2001) 31.
- [9] Schaefer R., *Podstawy genetycznej optymalizacji globalnej*. Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków, (2002), in Polish.
- [10] Lasota A., *Asymptotyczne własności półgrup operatorów Markowa*. Matematyka Stosowana, Matematyka dla Społeczeństwa, 3(45) (2002) 39, in Polish.
- [11] Socala J., Kosiński W., Kotowski S., *O asymptotycznym zachowaniu prostego algorytmu genetycznego*. Matematyka Stosowana, Matematyka dla Społeczeństwa, 6(47) (2005) 70, in Polish.
- [12] Socala J., *Markovian approach to genetic algorithms*. Under preparation, (2006).
- [13] Kotowski S., Socala J., Kosiński W., Michalewicz Z., *Markovian model of simple genetic algorithms and its asymptotic behaviour*. Under preparation, (2006).
- [14] Kotowski S., *Klasyfikacja algorytmów genetycznych*. Under preparation, (2006), in Polish.