On Convergence Rate of a Class of Genetic Algorithms

Liang Ming, Yuping Wang, and Yiu-ming Cheung

Abstract—Convergence rate study for genetic algorithms is a very important but very difficult task. The existing results can be classified into two types. One type is based on Doeblin condition in which some parameters should be estimated. The other type needs to estimate the eigenvalues of the state transition matrix. However, these parameters are difficult to estimate. In this paper, we first formulate a model for a class of genetic algorithms, then we analyze the convergence rate of this class of genetic algorithms in a different way, and prove the convergence rate is linear based on property of Markov chain.

Index Terms—Genetic Algorithms, convergence rate, Markov chain.

I. INTRODUCTION

In the past, Genetic Algorithms (GAs) have been extensively applied to a variety of fields such as function optimization, adaptive control, machine learning, neural networks, fuzzy systems, and so forth. It is a complicated and important task to study the convergence rate of GAs in the field of Evolutionary Computation. To the best of our knowledge, the existing results on the convergence rate of GAs can be divided into two groups. One group is based on the estimation of eigenvalues of state transition matrix. However, it is difficult to give a good estimation to them. For example, [1] derived a low bound of convergence rate for the only simple GAs based on the eigenvalue analysis of the state transition matrix. Similarly, [2] extended previous work of Suzuki [1] and got that the convergence rate of a GA was determined by the second largest eigenvalue of the transition matrix. Unfortunately, a proper estimation on the eigenvalues is difficult. The other group is based on the Doeblin condition. For example, [3] discussed the convergence rate of a GA in which only the crossover operator is used with a elitist selection scheme. Moreover, [4] further discussed the global convergence based on [3], but their results contain some parameters need to be further estimated. The local convergence rate of a GA with Cauchy mutation operator was analyzed in [5]. The convergence rate of a GA for a set of convex objective functions was discussed in [6]. Recently, [7]–[9] discussed the convergence time of evolutionary algorithms. In this paper, the convergence rate of a class of GAs is discussed from a new way different from the above two types. This new way is based on the properties of Markov chain. There is no any parameter which needs further estimation in our results. Our results show this class of genetic algorithms are linearly convergent. This class of GAs is applicable to arbitrary coding, arbitrary crossover, arbitrary mutation (the probability from one to another by mutation for any two individuals is positive), arbitrary selection (the probability that each individual is chosen is positive). Of course, this class of GAs cover many genetic algorithms and are globally convergent. For example, the canonical genetic algorithm [10] is included in this class of GAs.

This paper is organized as follows. Firstly, some necessary concepts are presented in Section II. In Section III the frame of one class of genetic algorithms is given. Subsequently, the global convergence of this class of GAs is discussed in Section IV. Furthermore, the convergence rate of this class of GAs is obtained by a new way in Section V. There exists no parameter needed to be further estimated in our result. Finally, the conclusions are drawn in Section VI.

II. BASIC CONCEPTS

Definition 1: (Finite Markov Chain)

\[ X = \{X_n, n = 0, 1, \ldots \} \]

is a stochastic discrete process on \( (\Omega, \mathcal{F}, P) \), and its state space \( S \) is finite. \( X \) is finite Markov chain, for \( \forall \) nonnegative integer \( n \) and arbitrary state \( i \), \( i_1, \ldots, i_n+1 \in S \),

\[ P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) > 0, \]

if

\[ P(X_{n+1} = i_{n+1}|X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) \]

\[ = P(X_{n+1} = i_{n+1}|X_n = i_n). \]

Definition 2: (State Transition Probability)

Given Markov chain \( X = \{X_n, n = 0, 1, \ldots \} \), after \( n \) steps transition from state \( i \) at time \( m \), the condition probability that the chain is state \( j \) at time \( m + n \) is called \( n \) steps transition probability of \( X \), denoted as \( m p_{ij}^{(n)} \). When \( m = 1 \), it can be simplified as \( p_{ij}^{(n)} \).

III. A CLASS OF GENETIC ALGORITHMS

In this paper, suppose that the number of all possible individuals under a specific encoding is finite. The set of all possible individuals is denoted as \( E \) whose cardinality is denoted as \( |E| = N \).

Algorithm 1 (model for a class of Genetic Algorithms)

1) Set \( k = 0 \) and generate initial population \( P(0) \). The best individual in \( P(t) \) is called the super individual, denoted as \( x_0(t) \);

2) Determine the fitness of each individual;
3) Perform crossover. After crossover the population \( P(t) \) is transited to \( P_1(t) \);
4) Perform mutation on \( P_1(t) \). Independently mutate each individual in \( P_1(t) \). The mutation showed satisfied that the probability that any individual in \( P_1(t) \) turns to be any individual in \( E \) is positive, i.e., for \( \forall a, b \in E, \text{prob}\{a \xrightarrow{M} b\} > 0 \), where \( \text{prob}\{a \xrightarrow{M} b\} \) represents the probability that \( a \) is changed to \( b \) by mutation. After mutation, the population \( P_1(t) \) is changed to \( P_2(t) \);
5) Perform selection. \( P_2(t) \) turns to be \( P(t + 1) \) after selection, where the probability that each individual is chosen will be positive;
6) Upgrade super individual. Find the best one in \( x_0(t) \cup P(t + 1) \), denoted as \( x_0(t + 1) \);
7) Stop if some stopping criterion is met, otherwise, go to 2).

Above model covers many genetic algorithms. For example, the canonical GA [10] with binary coding is a special case of it. Furthermore, the canonical genetic algorithms [10] with elitist selection is also in this class.

In this paper we consider the case that the population size is finite and denoted as \( n \). If each population is seen as a state and the state space consists of all possible populations, the cardinality of the state space, denoted as \( E^n \), is obviously \( |E^n| = |E|^n = N^n \). Note that the best individual will maintain in the population after selection and will not participate in crossover and mutation. The crossover can be seen as a random transition from \( P(t) \) to \( P_1(t) \), and is not changed along with the time. So crossover, mutation and selection can be seen as random transition from \( E^n \) to \( E^n \). Since the intermediate transitions caused by crossover, mutation and selection are independent from the time, they have Markov properties and are homogeneous. Thus the genetic process (crossover, mutation and selection) of Algorithm 1 can be described as a finite Markov chain, and its state transition matrix, denoted as \( P \) can be calculated by \( P = CMS \), where \( C, M \) and \( S \) are the state transition matrices caused by crossover, mutation and selection, respectively. It can be easily seen that the following properties are true.

**Property 1:** \( M \) is positive, i.e., any element of \( M \) is positive.

**Property 2:** \( S \) is column allowable, i.e., in each column of \( S \), there exists at least one positive element.

### IV. THE CONVERGENCE OF ALGORITHM 1

For convenience, we suppose that there exists only one global optimum. In the following, the extended-population \( x_0(t + 1) \cup P(t + 1) \) is regarded as a state, each state consists of \( n + 1 \) individuals (the super individual and individuals in the population). Therefore, the size of the state space is changed to \( |E^{n+1}| = |E|^{n+1} = N^{n+1} \). In order to make the discussion easier, we queue all states according to the following rules:

1) The better the super individual, the smaller the number of the state;
2) For states with the same super individual, the better the best individual in the population, the smaller the number of the state. If the super individual and the best individual are same in two states, then queue these two states according to the quality of their second best individuals, and so on.

For denotation convenience, we denote \( x_0(i) \) as the super individual of the state \( i \), and \( x_1(i) \cdots x_n(i) \) denote the individuals of the population in state \( i \). State \( i \) can be denoted as \( i = x_0(i)x_1(i) \cdots x_n(i) \). According to the rules for queuing all states aforementioned, we have

1) Each super individual is included in \( N^n \) states;
2) If the super individual is the global optimum, the corresponding states will ranked in the first \( N^n \) states.

In fact, every individual in \( E \) can be as a super individual, thus there are \( N \) all probable super individuals. For each super one, it may be the super individual of any population, so each super individual corresponds to \( N^n \) states. These \( N^n \) states can be seen as a state group. It is easy to get that there are \( N \) state groups, denoted as \( A_1, A_2, \cdots, A_N \). Furthermore, the super individual does not participate in crossover, mutation and selection, so the \( N^n \) states corresponding to each super individual include all possible \( N^n \) extended-populations with this fixed super individual. Thus, it can be easily gotten that the outcome of a state in a state group after crossover, mutation and selection will be still in this state group. The intermediate state transition matrices caused by crossover, mutation and selection in Algorithm 1, denoted as \( C, M, S \), respectively, are

\[
\tilde{C} = \begin{pmatrix}
  C & & & \\
  C & & & \\
  & C & & \\
  & & \ddots & \\
\end{pmatrix}, \\
\tilde{M} = \begin{pmatrix}
  M & & & \\
  M & & & \\
  & M & & \\
  & & \ddots & \\
\end{pmatrix}, \\
\tilde{S} = \begin{pmatrix}
  S & & & \\
  S & & & \\
  & S & & \\
  & & \ddots & \\
\end{pmatrix}.
\]

The state transition matrix caused by upgrading super individual is

\[
\tilde{\bar{P}} = \begin{pmatrix}
  U_{11} & U_{12} & \cdots & U_{1N} \\
  U_{21} & U_{22} & \cdots & U_{2N} \\
  \vdots & \vdots & \ddots & \vdots \\
  U_{N1} & U_{N2} & \cdots & U_{NN}
\end{pmatrix},
\]

where \( U_{11} \) is unit matrix, \( U_{rk} \neq 0 \) (\( r \leq k \)).

Therefore, Algorithm 1 can be described as a finite Markov chain whose state transition matrix is

\[
\bar{P} = \begin{pmatrix}
  P & & & \\
  \cdots & P & & \\
  \cdots & & \ddots & \\
  \cdots & & & P
\end{pmatrix},
\]

\[
= \begin{pmatrix}
  PU_{11} & PU_{12} & \cdots & PU_{1N} \\
  PU_{21} & PU_{22} & \cdots & PU_{2N} \\
  \vdots & \vdots & \ddots & \vdots \\
  PU_{N1} & PU_{N2} & \cdots & PU_{NN}
\end{pmatrix},
\]

where \( PU_{11} = P \) is a positive stochastic matrix. Denote
\[ R = \begin{pmatrix} PU_{21} \\ \vdots \\ PU_{N1} \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} PU_{22} \\ \vdots \\ PU_{NN} \end{pmatrix}. \]

Since \( U_{rk} \neq 0 (2 \leq r \leq k) \), we get \( R \neq 0 \) and \( T \neq 0 \).

Therefore \( \bar{P} = \begin{pmatrix} P \\ 0 \\ R \ T \end{pmatrix}. \)

Definition 3: Let \( Z_t = \max \{ f(a^{(t)}(i)) | k = 1, \ldots, n \} \) be a sequence of random variable representing the best fitness within a population represented by state \( i \) at generation \( t \). A genetic algorithm is called to converge to the global optimum if and only if

\[
\lim_{t \to \infty} P[Z_t = f^*] = 1,
\]

where \( f^* \) is the global optimum.

Lemma 1: Let \( C, M, S \) be stochastic matrices, where \( M \) is positive and \( S \) is column allowable. Then the product \( CMS \) is positive.

Lemma 2: Let \( P \) be an \( N \times N \) stochastic matrix and a state transition matrix of homogeneous finite Markov chain. If \( P \) has the form \( P = (P \ R \ T) \) in which both \( C \) and \( T \) are square matrices, and \( C : m \times m \) is a primitive stochastic matrix and \( R, T \neq 0 \), then \( P^k \) converges to a unique stable matrix \( P^\infty \), i.e., \( P^k \to P^\infty = (\pi^T, \pi^T, \ldots, \pi^T)^T \), where \( \pi = (p_1, p_2, \ldots, p_m, 0, \ldots, 0) \), \( p_1 > 0 \) \((t = 1, \ldots, m)\), and \( p^\infty \) is unique regardless of the initial distribution,

\[ p^k = p^0 P^k \to P^\infty = (p_1^\infty, p_2^\infty, \ldots, p_m^\infty 0, \ldots, 0) \],

where \( \sum_{i=1}^{m} p_i^\infty = 1 \).

Theorem 1: Algorithm 1 converges to the global optimum.

Proof: It can be seen from Properties 1, 2 and Lemma 1 that \( P > 0 \). From Lemma 2, we get that for arbitrary initial distribution \( p^0 \), there is a unique limit distribution \( p^\infty = \lim_{t \to \infty} p^0 \bar{P}^t = (p_1^\infty, p_2^\infty, \ldots, p_m^\infty 0, \ldots, 0) \), where \( p_i^\infty \) is the probability that the limit state is at the \( i \)-th state of \( N \) final optimal states. Therefore, the probability that the limit state is any non-globally optimal state is zero. It follows that the probability of the limit state being in globally optimal states is one:

\[ p_1^\infty + p_2^\infty + \cdots + p_N^\infty = 1, \]

i.e., \( \lim_{t \to \infty} P[Z_t = f^*] = 1 \). The proof is completed.

V. CONVERGENCE RATE OF ALGORITHM 1

In the former section, we discussed the global convergence of Algorithm 1. Now, we will estimate its convergence rate by use of the properties of Markov chain. First, we introduce a property of Markov chain [11] as follows.

Lemma 3: [11] Let a finite Markov chain has the state space containing \( N \) states numbered as \( \{1, \ldots, N\} \), and its one step transition matrix is \( P = [p_{ij}] \), where \( p_{ij} \) is the probability of the transition from state \( i \) to state \( j \). For any state \( i \), if \( \exists \) integers \( m \) and \( L \) such that

\[
\min_{1 \leq j \leq N} \{ p_{ij}^{(m)} \} \geq \delta > 0, (j = 1, \ldots, L) \tag{2}
\]

then there exists a \( \pi_j \) such that \( \lim_{k \to \infty} p_{ij}^{(k)} = \pi_j \), \( j = 1, \ldots, N \), where \( p_{ij}^{(m)} \) is the element of the \( i \)-th row and the \( j \)-th column of matrix \( P^m \). Moreover, when \( k > m \), it follows that

\[
|p_{ij}^{(k)}(k) - \pi_j| \leq (1 - L\delta)^{\frac{1}{L}} \tag{3}
\]

where \( \frac{1}{L} \) is the largest integer no larger than \( \frac{k}{L} \).

**Theorem 2:** For Algorithm 1, the convergence rate that the distribution corresponding to the \( k \)-th generation extended-population converges to the limit distribution \( p^\infty \) can be estimated as follows

\[
\|p^k - p^\infty\|_\infty \leq \{1 - [N^2 - (N^2 - N)](p_m p_0^2)\}^{\frac{k}{2}} \tag{4}
\]

where

\[
p_m = \min_{a_k \in E} \text{prob}\{a^M \to b\}, \quad p_0 = \min_{X \in E} \{ \min_{a_k \in E} \text{prob}\{a^S \to a|a \in X\} \}
\]

\(\lfloor \frac{k}{2} \rfloor\) is the largest integer no larger than \( \frac{k}{2} \).

Proof: By equation (1), we get the one step transition matrix \( \bar{P} \) by

\[
\bar{P} = \begin{pmatrix} PU_{11} & PU_{12} & \cdots \\ PU_{21} & PU_{22} & \cdots \\ \vdots & \vdots & \ddots \\ PU_{N1} & PU_{N2} & \cdots & PU_{NN} \end{pmatrix}
\]

It can be easily obtained the formula of \( \bar{P}^2 \) which is given at the top of the next page.

We first prove all elements in the first \( N \) columns of \( \bar{P}^2 \) are positive. Because \( P > 0 \) and \( U_{1j} \neq 0 \) \( (2 \leq r \leq N) \), \( PU_{1j}P > 0 \) \( (2 \leq r \leq N) \). Actually, there is at least one nonzero element \( u_{ij} \) \( = 0 \) in \( U_{1j} \), then this leads to all elements in the \( j \)-th column in \( PU_{1} \) are positive. Therefore, all elements in \( PU_{ij} \) are positive. Note that \( P \) is a positive matrix, there must exist a positive number \( \delta > 0 \) such that

\[
\min_{1 \leq j \leq N} \{ p_{ij}^{(2)} \} \geq \delta > 0, j = 1, \ldots, N \).
\]

By taking \( m = 2 \) in Lemma 3, we get that for \( \forall i \in [1, N^{n+1}] \), there exists a \( \pi_j \) such that \( \lim_{k \to \infty} p_{ij}^{(k)} = \pi_j \) \((j = 1, \ldots, N^n)\). Furthermore, when \( k > 2 \), the following is always true:

\[
|\tilde{p}_{ij}^{(k)} - \pi_j| \leq (1 - N^\delta)^{\frac{k}{2}} \tag{5}
\]

For arbitrary initial distribution \( p^0 = (p_0^0, p_0^1, \ldots, p_0^{N^{n+1}}) \), \( 0 \leq p_0^j \leq 1 \) \((j = 1, \ldots, N^n)\), the distribution corresponding to the \( k \)-th generation is \( p^k = p^0 P^k \). From Lemma 1, there must exist one unique limit distribution \( p^\infty \) satisfying \( p^k \to p^\infty \) \((k \to \infty)\). Let \( \tilde{p}_{ij}^{(k)} \) denote the \( j \)-th vector in \( \tilde{p}^k \), then

\[
\tilde{p}_{ij}^{(k)} \to \tilde{p}_{ij}^{(\infty)} = (\pi, \pi, \ldots, \pi)^T \to \infty \to \infty,
\]

\[
\|p^k - p^\infty\|_\infty = \|p^0 \tilde{p}^k - p^0 \tilde{p}^{(\infty)}\|_\infty \leq \max_{j} |p_0| \max_{i,j} |\tilde{p}_{ij}^{(k)} - \pi_j|.
\]
Thus denote the probability that the population evolves to the $k$-th generation population $X(k)$ after crossover, mutation and selection, where the population after crossover is denoted as $\tilde{X}$, and the probability that $X(k)$ evolves to $\tilde{X}$ after crossover is expressed as $P(X(k) \xrightarrow{C,M,S} \tilde{X})$. Thus, for all possible $\tilde{X} \in E^n$, there must be $\sum_{\tilde{X} \in E^n} P(X(k) \xrightarrow{C,M,S} \tilde{X}) = 1$. Let $P(X(k + 1) \xrightarrow{M} X(k + 1))$ denote the probability that $X$ becomes $X(k + 1)$ after mutation, where $p_m = \min_{a \in E} \text{prob}(a \xrightarrow{M} b)$. Let $P(X(k + 1) \xrightarrow{S} X(k + 1))$ denote the probability that the population $X(k+1)$ is kept after selection. Obviously we have $P(X(k + 1) \xrightarrow{S} X(k + 1)) \geq p_s^n$, where $p_s = \min_{\tilde{X} \in E^n} \{\min_{a \in E} \text{prob}(a \xrightarrow{S} a | a \in \tilde{X})\}$. Thus

$$P(X(k) \xrightarrow{C,M,S} X(k + 1)) \geq \sum_{\tilde{X} \in E^n} P(X(k) \xrightarrow{C} \tilde{X}) P(\tilde{X} \xrightarrow{M} X(k + 1)) \cdot P(X(k + 1) \xrightarrow{S} X(k + 1)) \geq 1 \cdot p_m^n \cdot p_s^n \tag{7}$$

It follows from equation (7) that the elements of matrix $P$ satisfy

$$\min_{1 \leq i,j \leq N^n} p_{ij} \geq (p_m p_s)^n \tag{8}$$

Note that all the elements in the first $N^n$ columns in $\tilde{P}^2$ are positive. Now we will estimate the lower bound on these elements. For notation convenience, denote $\tilde{P}_1 = P^2 = [p_{ij}^{(2)}]$ and $P_1 = PU_{ij}P + PU_{ij}PU_{ij} + \cdots + PU_{ij}PU_{ij}PU_{ij} = [q_{ij}] (2 \leq r \leq N)$ in $P^2$, then $\tilde{P}_1$ and $P_1$ have the following properties:

(I) $p_{ij}^{(2)} \geq \sum_{k=1}^{N^n} p_{ik} p_{kj} \geq (\min_{1 \leq i,j \leq N^n} p_{ij})^2 \geq n_1 (p_m p_s)^{2n}$, where $n_1 = N^n - (N - 1)^n$.

(II) $q_{ij} = \sum_{k=1}^{N^n} p_{ik} p_{kj} \geq (\min_{1 \leq i,j \leq N^n} p_{ij})^2 \geq n_1 (p_m p_s)^{2n}$, and $q_{ij} \geq \min_{1 \leq i,j \leq N^n} (p_m p_s)^{2n}$.

It can be seen from properties (I) and (II) that the elements in the first $N^n$ columns in $\tilde{P}^2$ are $p_{ij}^{(2)} \geq n_1 (p_m p_s)^{2n}$. Therefore, there exists $\delta = \min_{1 \leq i,j \leq N^n} \{n_1 (p_m p_s)^{2n}\}$ satisfying equation (6), i.e., when $k \geq 2$, there must be

$$\|p^k - p^\infty\|_\infty \leq (1 - \delta (N^n - (N - 1)^n)(p_m p_s)^{2n})^{1/2} \tag{9}$$

when $0 < (N^2 - (N - 1)^n)(p_m p_s)^{2n} < 1$, this class of algorithms are linearly convergent.

VI. CONCLUSIONS

In this paper, the convergence rate of a class of GAs was discussed by a new way. The discussion is based on the properties of Markov chain. This class of GAs is applicable to arbitrary coding, arbitrary crossover, arbitrary mutation (the probability from one to another by mutation for any two individuals is positive), arbitrary selection (the probability that each individual is chosen is positive). There exists no parameter needed for further estimation in our result, and the convergence rate is linear.

REFERENCES


