Conditions for the Convergence of Evolutionary Algorithms

Jun He and Xinghuo Yu

Abstract

This paper presents a theoretical analysis of the convergence conditions for evolutionary algorithms. The necessary and sufficient conditions, necessary conditions, and sufficient conditions for the convergence of evolutionary algorithms to the global optima are derived, which describe their limiting behaviors. Their relationships are explored. Upper and lower bounds of the convergence rates of the evolutionary algorithms are given.

I. INTRODUCTION

Evolutionary Algorithms (EAs in short) are a kind of algorithms that simulate biological evolution. These algorithms have been applied to many optimization problems successfully, such as function optimization, combinatorial optimization and machine learning.

Convergence is an important issue in the theoretical study of EAs [1], [2]. In the past years, good results have been obtained about the convergence of EAs in finite space and discrete time, including both time-homogeneous transitions and time-inhomogeneous transitions [3], [4], [5], [6], [7]. The convergence theory of EAs beyond the finite space and discrete time has also been discussed [7]. The modeling and analysis of non-elitist evolutionary algorithms in terms of super-martingale has been suggested in [8]. Further research on convergence rates has been done [9], [7], [10]. A more detailed survey about the convergence theory of EAs can be found in [1]. Most of existing researches focus on the sufficient conditions for the convergence of EAs, but little work has been done in dealing with necessary conditions, and sufficient and necessary conditions. From the completeness of theory, these conditions are also important parts of the convergence conditions for EAs.

The following questions will be addressed in this paper:

- Which conditions are sufficient and necessary for the convergence of EAs?
- Which conditions are necessary conditions?
- Which conditions are sufficient conditions?

The discussion in this paper will be undertaken in the general measurable space whose operators and parameters could be time-varying. In order to make the results more general (instead of specific), we conduct our discussion from an abstract viewpoint.

We use the Markov process as the analysis tool. In our previous paper [11], we have discussed the convergence rate of genetic algorithms from the viewpoint of abstract framework. In the same way, we will undertake further investigations into the sufficient conditions, necessary conditions, and sufficient and necessary conditions for the convergence of EAs, and their relationship. We also obtain some results about the convergence rate of EAs based on these conditions.

This paper is organized as follows. Section 2 introduces the Markov process model for EAs. Section 3 gives different conditions about the convergence of EAs. Section 4 presents results about the convergence rate of EAs. Section 5 concludes the paper.

II. EAS AND THEIR MARKOV CHAIN MODEL

Consider the optimization problem:

$$\max\{f(s); s \in S\} \tag{1}$$

where $S \subset \Omega$ and Ω is assumed to be a measurable space, $f(s)$ the objective function, or the fitness of s where $\mid f(s) \mid < +\infty$. Denote the optimal solution set by $S^* = \{s^* \mid f(s^*) = \max f(s), s \in S\}$.

Let $\mu(\cdot)$ be a measure to space Ω. Sometimes $\mu(S^*) = 0$, which means S^* is a set with measure 0. It is not convenient to analyze such a set mathematically, so let's consider an expanded set $S^*_\delta = \{s \mid f(s^*) - f(s) < \delta\}$ where δ is a small positive value and S^*_δ will replace the set S^*. We can choose an appropriate δ and make $\mu(S^*_\delta) > 0$. In this sense, we always assumes that $\mu(S^*) > 0$ in this paper.

An EA for solving the above problem can be described briefly as:

1) Initialization: generate an initial population, denoted by $\xi(0)$, and let $t \leftarrow 0$.
2) Crossover: generate a new population from $\xi(t)$ by a crossover operator, denoted by $\xi_C(t)$.

Jun He is with Department of Computer Science, Northern Jiaotong University, Beijing, 100044, PR CHINA. (mailto:jun.he@ieee.org)
Xinghuo Yu is with Faculty of Informatics and Communication, Central Queensland University, Rockhampton, QLD 4702, AUSTRALIA

3) Mutation: generate a new population from $\xi_C(t)$ by a mutation operator, denoted by $\xi_M(t)$.
4) Selection: generate a new population from $\xi_M(t)$ (and $\xi(t)$) by a selection operator, denoted by $\xi_S(t)$.
5) Adaptation (or no adaptation): if there is an adaptation, the act of adjusting the algorithm's parameters (and operators) only depends on the states of $\xi(t)$ and $\xi_S(t)$.
6) if the EA satisfies the stopping criterion, then stop; else let $\xi(t+1)=\xi_S(t)$ and $t \leftarrow t+1$, then return to Step 2.

The adaptation (Step 5) is considered in the above algorithm, which is used to modify the values of strategy parameters (and the genetic operators) during the run of the algorithm. Markov chain is one of main mathematical analyzing tools in modeling EAs [2]. In this paper we use the theory of Markov process too. Now we give some definitions that will be used in the subsequent discussion.

Definition 1: The space S is called an *individual space*, where each state s in S is called an *individual*, the space $X=S^N$ consisting of N individuals is called a *population space*, where elements in X are called a *population*, denoted by $x=\{x_1,\cdots,x_N\}$, N is called the *population size*. The *optimal solution set* in the population space is denoted by $X^*=\{x^* \mid \exists x_i \in x : x_i \in S^*\}$.

Since the state of $\xi(t+1)$ is only dependent on the state of $\xi(t)$, then $\{\xi(t); t=0,1,\cdots\}$ can be modeled by a non-homogeneous Markov chain whose state space is X and the transition function is [11]:

$$\begin{aligned}&P(t;x,dy)=P(\xi(t+1)=dy \mid \xi(t)=x)\\&=\int_{u\in X}\int_{v\in X} P_C(t;x,du)P_M(t;u,dv)P_S(t;v,dy)\end{aligned} \tag{2}$$

There are different kinds of definitions of convergence in the Markov chain theory. The following definition of convergence will be used in this paper.

Definition 2: Let $\{\xi(t); t=0,1,\cdots\}$ be the Markov chain given by (2), then $\{\xi(t); t=0,1,\cdots\}$ converges to the optimal solution set X^*, if for any initial population $\xi(0)$,

$$\lim_{t\to+\infty}\mu_t(X^*)=1, \tag{3}$$

where $\mu_t(X^*)=\mu(\xi(t)\in X^*)$.

In the following, we will prove the convergence of EAs and estimate the convergence rate.

III. Convergence Conditions of EAs

First let's give sufficient and necessary conditions for the convergence of EAs.

Theorem 1: Let $\{\xi(t); t=0,1,\cdots\}$ be the Markov chain given by (2). Define

$$\begin{aligned}\alpha(t) \;&= \int_{x\notin X^*} P(\xi(t+1)\in X^* \mid \xi(t)=x)\mu_t(dx)\\&-\int_{x\in X^*} P(\xi(t+1)\notin X^* \mid \xi(t)=x)\mu_t(dx)\end{aligned} \tag{1}$$

then $\{\xi(t); t=0,1,\cdots\}$ converges to the optimal solution set X^* if and only if $\alpha(t)$ satisfies the following condition:

$$\sum_{t=0}^{+\infty}\alpha(t)=1-\mu_0(X^*) \tag{2}$$

Proof: For any given time t, we have

$$\begin{aligned}\mu_{t+1}(X^*) \;&= \int_{x\notin X^*} P(\xi(t+1)\in X^* \mid \xi(t)=x)\mu_t(dx)\\&+\int_{x\in X^*} P(\xi(t+1)\in X^* \mid \xi(t)=x)\mu_t(dx)\\&=\mu_t(X^*)-\int_{x\in X^*} P(\xi(t+1)\notin X^* \mid \xi(t)=x)\mu_t(dx)\\&+\int_{x\notin X^*} P(\xi(t+1)\in X^* \mid \xi(t)=x)\mu_t(dx)\end{aligned}$$

so we have

$$\mu_{t+1}(X^*)-\mu_t(X^*)=\alpha(t)$$

and then

$$\mu_{t+1}(X^*)-\mu_0(X^*)=\sum_{k=0}^{t}\alpha(k)$$

we get

$$\lim_{t\to\infty}\mu_{t+1}(X^*)=1 \Longleftrightarrow \sum_{k=0}^{+\infty}\alpha(k)=1-\mu_0(X^*)$$

which proves the conclusion. ■

Now we give some explanation to the above conditions. The meaning of $\alpha(t)$ can be viewed as the difference between the flow from the non-optimal solution set to the optimal solution set and vice versa. So the condition (2) is a constraint to the above flows. Usually the initial population is not optimal, that is $\mu_0(X^*) = 0$, then the condition (2) become $\sum\limits_{k=0}^{+\infty} \alpha(k) = 1$.

From the definition of the convergence: $\lim_{t\to+\infty} \mu_t(X^*) = 1$, we can get some other necessary and sufficient conditions, such as:

$$\lim_{t\to+\infty} \int_{x\in X} P(\xi(t+1) \in X^* \mid \xi(t) = x)\mu_t(dx) = 1 \tag{3}$$

or equivalently from $\lim_{t\to+\infty} \mu_t(X \backslash X^*) = 0$

$$\lim_{t\to+\infty} \int_{x\in X} P(\xi(t+1) \notin X^* \mid \xi(t) = x)\mu_t(dx) = 0 \tag{4}$$

The condition (3) can be rewritten as: there exists a sequence $\{\delta(t); t = 0, 1, \cdots\}$ such that $0 \leq \delta(t) \leq 1$, which satisfies

$$\int\limits_{x\in X} P(\xi(t+1) \in X^* \mid \xi(t) = x)\mu_t(dx) \geq \delta(t) \tag{5}$$

and

$$\lim_{t\to+\infty} (1 - \delta(t)) = 0 \tag{6}$$

If we weaken the above conditions, we will get some necessary conditions; on the other hand, sufficient conditions can be derived by strengthening the above conditions.

In the following, we consider the necessary conditions for the convergence of EAs.

Theorem 2: Let $\{\xi(t); t = 0, 1, \cdots\}$ be the Markov chain given by (2), if it converges to the optimal solution set X^*, then for any $\xi(0) \notin X^*$, the following necessary conditions hold:

(1) for any given time $t \geq 0$,

$$\lim_{t\to+\infty} \int\limits_{x\in X^*} P(\xi(t+1) \notin X^* \mid \xi(t) = x)\mu_t(dx) = 0 \tag{7}$$

(2) there is a sequence $\{\delta(t); t = 0, 1, \cdots\}$ such that $0 \leq \delta(t) \leq 1$, which satisfies

$$\int\limits_{x\notin X^*} P(\xi(t+1) \in X^* \mid \xi(t) = x)\mu_t(dx) \geq \delta(t)\mu_t(X \backslash X^*) \tag{8}$$

and

$$\prod_{t=0}^{+\infty} (1 - \delta(t)) = 0 \tag{9}$$

Proof: We first prove (1). Assume that (7) doesn't hold, that is, there exist some $\epsilon \in (0, 1)$ and a subsequence of time $\{t_k; k = 0, 1, \cdots\}$ such that

$$\int\limits_{x\in X^*} P(\xi(t_k+1) \notin X^* \mid \xi(t_k) = x)\mu_{t_k}(dx) > \epsilon \tag{10}$$

Since the chain converges to the optimal solution set, for the given ϵ, there exists some moment $\tau > 0$, such that for $t > \tau$, we have

$$\mu_t(X^*) > 1 - \epsilon$$

that is

$$\mu_t(X \backslash X^*) < \epsilon$$

Now let t be large enough, $t > \tau$, then we have

$$\begin{aligned} \epsilon &> \mu_{t_k+1}(X \backslash X^*) \\ &= \int\limits_{x\in X} P(\xi(t_k+1) \notin X^* \mid \xi(t_k) = x)\mu_t(dx) \\ &\geq \int\limits_{x\in X^*} P(\xi(t_k+1) \notin X^* \mid \xi(t) = x)\mu_t(dx) \end{aligned} \tag{11}$$

It can be seen that (10) contradicts (11), so the assumption doesn't holds, and therefore condition (7) holds.

We now prove (2). The discussion will be divided into two cases:

Case 1: there is some time τ such that: $\mu_\tau(X^*) = 1$. In this case, let's take $\delta(\tau) = 1$ and other $\delta(t) = 0$, then (8) and (9) hold.
Case 2: for any time $t \geq 0$, it holds: $\mu_t(X^*) < 1$. In this case, since:

$$\begin{aligned} \mu_{t+1}(X^*) &= \int\limits_{x \in X^*} P(\xi(t+1) \in X^* \mid \xi(t) = x)\mu_t(dx) \\ &+ \int\limits_{x \notin X^*} P(\xi(t+1) \in X^* \mid \xi(t) = x)\mu_t(dx) \\ &\leq \mu_t(X^*) + \int\limits_{x \notin X^*} P(\xi(t+1) \in X^* \mid \xi(t) = x)\mu_t(dx) \end{aligned} \tag{12}$$

define

$$\delta(t) = \frac{\int\limits_{x \notin X^*} P(\xi(t+1) \in X^* \mid \xi(t) = x)\mu_t(dx)}{1 - \mu_t(X^*)}$$

Since $\int_{x \notin X^*} P(\xi(t+1) \in X^* \mid \xi(t) = x)\mu_t(dx) \leq \mu_t(X \backslash X^*) = 1 - \mu_t(X^*)$, we have $0 \leq \delta(t) \leq 1$, and then (8) holds.

Substituting $\delta(t)$ into (12) yields

$$\delta(t) \geq \frac{\mu_{t+1}(X^*) - \mu_t(X^*)}{1 - \mu_t(X^*)}$$

or equivalently

$$1 - \delta(t) \leq \frac{1 - \mu_{t+1}(X^*)}{1 - \mu_t(X^*)}$$

so we have

$$\prod_{k=0}^{t} (1 - \delta(k)) \leq \frac{1 - \mu_{t+1}(X^*)}{1 - \mu_0(X^*)}$$

Since the chain converges to the optimal solution set, we have

$$\lim_{t \to \infty} \prod_{k=0}^{t} (1 - \delta(k)) \leq \lim_{t \to \infty} \frac{1 - \mu_{t+1}(X^*)}{1 - \mu_0(X^*)} = 0$$

so (9) holds. ■

The conditions in Theorem 2 can be explained as follows:

- Condition (7) means that: as the time step in EAs becomes large enough, once population $\xi(t)$ enters the optimal set X^*, then its offspring has little possibility to escape from the optimal set, that is, set X^* is attractive.
- Conditions (8) and (9) means that: if population $\xi(t)$ is not in the optimal set, then after a period of evolution, its offspring has a high possibility to enter the optimal set, that is, set X^* is accessible. In the finite space, the conditions mean that there exists a path from the initial state to the global optimal state.

The following examples illustrate whether an EA satisfies (7).

Example 1: For any EA with the elitist selection strategy (that is, the best individuals in the parent population will be kept in the offspring population in the selection), it satisfies (7).

Example 2: For an EA with a non-elitist selection strategy, if its evolutionary operators do not change in time, then it doesn't satisfy (7).

Example 3: For a simulated annealing like genetic algorithm [12], let x_{best} be the individual with the best fitness in population $\xi(t)$, if the possibility of x_{best} to be alive in the next generation population $\xi(t+1)$ will trend to 1 as time t increases, then this genetic algorithm satisfies (7).

The following examples illustrate whether an EA satisfies (8) and (9).

Example 4: For a genetic algorithm based on a binary state space $\{0,1\}^l$, let its mutation operator be: each site s_i in the individual $(s_1 \cdots s_l)$ become its complement with mutation rate $p_m : 0 < p_m < 0.5$; let its crossover operator be the one-point crossover; and the selection operator is an elitist selection strategy. Then the probability of individual s becoming the optimal individual $(s_1^* \cdots s_l^*)$ is not less than $(p_m)^l$. So let's choose $\delta(t) = (p_m)^l < 1$, then we have $\prod_{k=0}^{+\infty}(1 - \delta(k)) = 0$, that means (9) holds. For the given $\delta(t)$, condition (8) holds too.

A natural and interesting question is that: if an EA satisfies Condition (7), (8) and (9), then will the EA converge to the optimal set X^*? The answer to the question is negative. Let's see the following example.

Example 5: Let the state space consist of only two elements $\{0,1\}$, where $f(0) = 1, f(1) = 2$, and the (1+1) EA is:

1) Initialization: $\xi(0) = 0$ where consists of only one individual, and let $t = 0$.
2) Mutation: let mutation rate be e^{-t}, i.e. the probability for $\xi(t)$ becoming its complement is e^{-t}, then form $\xi_M(t)$.
3) Selection: select the offspring from $\xi(t)$ and $\xi_M(t)$ according to the proportional selection, then form $\xi_S(t)$.
4) let $\xi(t+1) = \xi_S(t)$ and $t \leftarrow t+1$, return to step 2 and repeat the loop.

In the above example, it is easy to see that if we let $\delta(0)=1$ and $\delta(t)=0,\ t\neq 0$, then

$$\int_{x\notin X^*} P(\xi(t+1)\in X^* \mid \xi(t)=x)\mu_t(dx) \geq \delta(t)\mu_t(X\backslash X^*)$$

and we also have $\prod_{k=0}^{+\infty}(1-\delta(k))=0$. So conditions (8) and (9) hold. Since the mutation rate $\lim_{t\to+\infty} e^{-t}=0$, then condition (7) holds too. But this EA is not convergent.

Last, let's consider some sufficient conditions for the convergence of EAs. There are already many good results on the sufficient conditions [1]. The sufficient conditions given here are obtained by strengthening conditions (3) and (4). The following results can also be drawn from Theorem 6.7 in [7] directly.

Theorem 3: Let $\{\xi(t); t=0,1,\cdots\}$ be the Markov chain given by (2). If it satisfies the following conditions:

(1) for any $t\geq 0$ and any $x\in X^*$,

$$P(\xi(t+1)\notin X^* \mid \xi(t)=x)=0 \tag{13}$$

(2) there exists a sequence $\{\delta(t); t=0,1,\cdots\}$ such that $0\leq\delta(t)\leq 1$ and it satisfies:

$$\int\limits_{x\notin X^*} P(\xi(t+1)\in X^* \mid \xi(t)=x)\mu_t(dx) \geq \delta(t)\mu_t(X\backslash X^*) \tag{14}$$

and

$$\prod_{t=0}^{+\infty}(1-\delta(t))=0 \tag{15}$$

Then starting from any $\xi(0)\notin X^*$, the chain converges to the optimal solution set X^*, and for any time t:

$$1-\mu_t(X^*) \leq \prod_{k=0}^{t-1}(1-\delta(k)) \tag{16}$$

Proof: Since

$$\begin{aligned}\mu_{t+1}(X^*) &= \int\limits_{x\in X^*} P(\xi(t+1)\in X^* \mid \xi(t)=x)\mu_t(dx)\\ &+ \int\limits_{x\notin X^*} P(\xi(t+1)\in X^* \mid \xi(t)=x)\mu_t(dx)\\ &\geq \mu_t(X^*)+\delta(t)\mu_t(X\backslash X^*)\end{aligned}$$

so

$$1-\mu_{t+1}(X^*) \leq (1-\delta(t))(1-\mu_t(X^*))$$

and then

$$1-\mu_t(X^*) \leq \prod_{k=0}^{t-1}(1-\delta(k))\mu_0(X^*)$$

if

$$\lim_{t\to\infty}\prod_{k=0}^{t-1}(1-\delta(k))=0$$

then $\lim_{t\to\infty}\mu_t(X^*)=1$, which proves the convergence. ■

Now we give some intuitive explanations for the above conditions.

- Condition (13) is stronger than Condition (7), that is, once population $\xi(t)$ enters the optimal set X^*, its offspring $\xi(t+1)$ has no possibility to escape from the optimal set.
- Conditions (14) and (15) are the same as Conditions (8) and (9).

The following example shows the application of Theorem 3:

Example 6: If the selection operator in a genetic algorithm is given as in Example 1, and the mutation and crossover operators are given as in Example 4, then from Theorem 3, we know that the genetic algorithm will converge to the optimal set.

IV. Convergence Rate

Based on conditions similar to the above conditions, we shall present some results on the convergence rate of EAs. The convergence rate refers to how fast the population state generated by EAs converges to the optimal set X^*. In this paper, the convergence rate is measured by $1-\mu_t(X^*)$.

The following lemma gives the lower bound of $1-\mu_t(X^*)$.

Lemma 1: Let $\{\xi(t); t=0,1,\cdots\}$ be the Markov chain given (2), if it converges and there exists a sequence $\{\beta(t); t=0,1,\cdots\}$ such that $0 \leq \beta(t) \leq 1$ and it satisfies:

$$\int_{x \notin X^*} P(\xi(t+1) \in X^* \mid \xi(t)=x)\mu_t(dx) \leq \beta(t)\mu_t(X \backslash X^*) \tag{1}$$

then starting from any $\xi(0) \notin X^*$, it holds for any time t:

$$1-\mu_t(X^*) \geq \prod_{k=1}^{t-1}(1-\beta(t)) \tag{2}$$

Proof: The proof is similar to the proof of Theorem 3. ■

From Lemma 1 and Theorem 3, we can get the following theorem.

Theorem 4: Let $\{\xi(t); t=0,1,\cdots\}$ be the Markov chain given by (2), if the following conditions hold:

(1) for any $t \geq 0$ and any $x \in X^*$,

$$P(\xi(t+1) \notin X^* \mid \xi(t)=x)=0 \tag{3}$$

(2) there exists a sequence $\{\delta(t); t=0,1,\cdots\}$ such that $0 \leq \delta(t) \leq 1$ and it satisfies:

$$\int_{x \notin X^*} P(\xi(t+1) \in X^* \mid \xi(t)=x)\mu_t(dx) \geq \delta(t)\mu_t(X \backslash X^*) \tag{4}$$

and

$$\prod_{t=0}^{+\infty}(1-\delta(t))=0 \tag{5}$$

(3) there exists a sequence $\{\beta(t); t=0,1,\cdots\}$ such that $0 \leq \beta(t) \leq 1$ and it satisfies:

$$\int_{x \notin X^*} P(\xi(t+1) \in X^* \mid \xi(t)=x)\mu_t(dx) \leq \beta(t)\mu_t(X \backslash X^*) \tag{6}$$

then starting from any $\xi(0) \notin X^*$, the chain converges to the optimal solution set X^* and the convergence rate is bounded by

$$\prod_{k=1}^{t-1}(1-\beta(t)) \leq 1-\mu_t(X^*) \leq \prod_{k=1}^{t-1}(1-\delta(t)) \tag{7}$$

Theorem 4 shows that the convergence rate of an EA can be obtained by estimating $\delta(t)$ and $\beta(t)$. However, in some sense, the result only indicates the existence of convergence rate, rather how to choose compute the convergence rate.

In the following we will give some further investigation into the estimation of $\delta(t)$ and $\beta(t)$. Let's consider the following EA which takes an elitist selection strategy:

1) Initialization: generate an initial population, denoted as $\xi(0)$, $t=0$.
2) Crossover: generate a new population $\xi_C(t)$ from $\xi(t)$ by crossover, and assume that all the individual(s) with the best fitness in the parent population are still kept in the population $\xi_C(t)$.
3) Mutation: generate a new population $\xi_M(t)$ from $\xi_C(t)$ by mutation, and assume that all the individual(s) with the best fitness in population $\xi_C(t)$ are still kept in the population $\xi_M(t)$.
4) Selection: generate a new population $\xi_S(t)$ from $\xi_M(t)$ by selection, and assume that all the individual(s) with the best fitness in population $\xi_M(t)$ are still kept in the population $\xi_S(t)$.
5) Adaptation: adjust the algorithm's parameters based on states $\xi(t)$ and $\xi_S(t)$.
6) let $\xi(t+1)=\xi_S(t)$, and $t \leftarrow t+1$.
7) if the EA satisfies the stopping criterion, then stops; else return to Step 2.

Now we analyze the bound of its convergence rate.

Definition 3: Let set D_C be the set of populations $D_C(t)=\{x \in X; P_C(\xi_C(t) \in X^* \mid \xi(t)=x)>0\}$.

Corollary 1: Let $\{\xi(t); t=0,1,\cdots\}$ be the Markov chain given by (2), if it satisfies the following conditions:

(1) there exits $0 \leq \delta_D(t) \leq 1$ and $0 \leq \beta_D(t) \leq 1$ such that

$$\delta_D(t) \leq \mu_t(D_C(t)) \leq \beta_D(t) \tag{8}$$

(2) there exits $0 \leq \delta_C(t) \leq 1$ and $0 \leq \beta_C(t) \leq 1$, for each $x \in D_C$ such that

$$\delta_C(t) \leq P(\xi_C(t) \in X^* \mid \xi(t)=x) \leq \beta_C(t) \tag{9}$$

(3) there exits $0 \le \delta_{_M}(t) \le 1$ and $0 \le \beta_{_M}(t) \le 1$, for any $x \notin X^*$ such that

$$\delta_{_M}(t) \le P(\xi_{_M}(t) \in X^* \mid \xi_{_C}(t) = x) \le \beta_{_M}(t) \tag{10}$$

then we have

$$\delta(t)\mu_t(X \backslash X^*) \le \int\limits_{x \notin X^*} P(\xi(t+1) \in X^* \mid \xi(t) = x)\mu_t(dx) \le \beta(t)\mu(X \backslash X^*) \tag{11}$$

where

$$\delta(t) = \delta_{_D}(t)\delta_{_C}(t) + \delta_{_M}(t) - \delta_{_M}(t)\delta_{_C}(t)\delta_{_D}(t) \tag{12}$$

$$\beta(t) = \beta_{_D}(t)\beta_{_C}(t) + \beta_{_M}(t) - \beta_{_M}(t)\beta_{_C}(t)\beta_{_D}(t) \tag{13}$$

Proof: Since the algorithm adopts the elitist selection strategy, so for any $x \notin X^*$, it holds:

$$P(\xi(t+1) \in X^* \mid \xi(t) = x) = P(\xi_{_M}(t) \in X^* \mid \xi(t) = x)$$

First let's consider $\beta(t)$. For any $x \notin X^*$, we have

$$\begin{aligned}
&\int\limits_{x \notin X^*} P(\xi(t+1) \in X^* \mid \xi(t) = x)\mu_t(dx) \\
&= \int\limits_{y \in X} \int\limits_{x \notin X^*} P(\xi_{_M}(t) \in X^* \mid \xi_{_C}(t) = y) P(\xi_{_C}(t) = dy \mid \xi(t) = x)\mu_t(dx) \\
&= \int\limits_{y \in X^*} \int\limits_{x \notin X^*} P(\xi_{_M}(t) \in X^* \mid \xi_{_C}(t) = y) P(\xi_{_C}(t) = dy \mid \xi(t) = x)\mu_t(dx) \\
&+ \int\limits_{y \notin X^*} \int\limits_{x \notin X^*} P(\xi_{_M}(t) \in X^* \mid \xi_{_C}(t) = y) P(\xi_{_C}(t) = dy \mid \xi(t) = x)\mu_t(dx)
\end{aligned} \tag{14}$$

For the first term in (14), since the all individuals with the best fitness will be kept in the mutation, we have

$$\begin{aligned}
&\int\limits_{y \in X^*} \int\limits_{x \notin X^*} P(\xi_{_M}(t) \in X^* \mid \xi_{_C}(t) = y) P(\xi_{_C}(t) = dy \mid \xi(t) = x)\mu_t(dx) \\
&= \int\limits_{x \notin X^*} P(\xi_{_C}(t) \in X^* \mid \xi(t) = x)\mu_t(dx)
\end{aligned} \tag{15}$$

For the second term in (14), from (10) we have

$$\begin{aligned}
&\int\limits_{y \notin X^*} \int\limits_{x \notin X^*} P(\xi_{_M}(t) \in X^* \mid \xi_{_C}(t) = y) P(\xi_{_C}(t) = dy \mid \xi(t) = x)\mu_t(dx) \\
&\le \beta_{_M}(t) \int\limits_{x \notin X^*} P(\xi_{_C}(t) \notin X^* \mid \xi(t) = x)\mu_t(dx) \\
&= \beta_{_M}(t)\mu_t(X \backslash X^*) - \beta_{_M}(t) \int\limits_{x \notin X^*} P(\xi_{_C}(t) \in X^* \mid \xi(t) = x)\mu_t(dx)
\end{aligned} \tag{16}$$

And from (15) and (16), we get

$$\begin{aligned}
&\int\limits_{x \notin X^*} P(\xi(t+1) \in X^* \mid \xi(t) = x)\mu_t(dx) \\
&\le \beta_{_M}\mu_t(X \backslash X^*) + (1 - \beta_{_M}) \int\limits_{x \notin X^*} P(\xi_{_C}(t) \in X^* \mid \xi(t) = x)\mu_t(dx) \\
&\le (\beta_{_M} + (1 - \beta_{_M}\beta_C)\mu_t(D_{_C}))\mu_t(X \backslash X^*) \\
&\le (\beta_{_D}\beta_{_C} + \beta_{_M} - \beta_{_M}\beta_{_C}\beta_{_D})\mu_t(X \backslash X^*) \\
&= \beta_{_M}(t)\mu_t(X \backslash X^*)
\end{aligned}$$

Note that in the derivation of the above last two inequalities, the conditions (8) and (9) are used. So we have finished the estimation of $\beta(t)$.

The estimation of $\delta(t)$ is similar to the above estimation of $\beta(t)$, here we ignore the detail. ■

Corollary 1 demonstrates that bounds do exist, but no further detail on practical estimation of the convergent rate is given. Accurate estimation will be dependent on the optimization function and the algorithm, that is, it is problem specific.

The following simple example shows an application of Corollary 1.

Example 7: Let the state space to $S = \{0, 1\}^l$, assume $f(0 \cdots 0)$ has the maximum value. The (1+1) EA for this problem is:

1) Initialization: assume $\xi(0) \ne (0 \cdots 0)$, and let $t = 0$.
2) Mutation: let mutation rate be $0 < p_m < 0.5$, that is, the probability for $\xi(t)$ becoming its complement is p_m, then form $\xi_{_M}(t)$.
3) Selection: select the better individual from $\xi(t)$ and $\xi_{_M}(t)$ as the next generation population, then form $\xi_{_S}(t)$.
4) let $\xi(t = 1) = \xi_{_S}(t)$ and $t \leftarrow t + 1$, return to step 2 and repeat the loop.

In this algorithm, there is no crossover, and we see from any $x \ne (0 \cdots 0)$:

$$(p_m)^l \le P(\xi_{_M}(t) \in X^* \mid \xi_{_C}(t) = x) \le (1 - p_m)^{(l-1)} p_m$$

where $\delta(t) = (p_m)^l$ and $\beta(t) = (1 - p_m)^{(l-1)} p_m$.

V. CONCLUSIONS

This paper has presented some theoretical results about the convergence of EAs in general measurable space and discrete time. By using the theory of Markov process, sufficient and necessary conditions, necessary conditions and sufficient conditions for the convergence of EAs have been obtained. These conditions describe the limiting behaviors of EAs. Different from the existing results, our results show the relationship between these conditions. Starting from the sufficient and necessary conditions (3) and (4), we can get the necessary conditions by weaken them, and the sufficient conditions by strengthening them. Based on some similar conditions, we also get some upper and lower bounds of the convergence rate of EAs with the elitist selection strategy.

Note that the conditions given in this paper are theoretical, the practical estimation is yet to be developed. Their applicability in practice problems will be further studied.

Acknowledgements: This work is partially supported by the Software Engineering State Key Laboratory of Wuhan University, China.

REFERENCES

[1] G. Rudolph. Finite Markov chain results in evolutionary computation: A tour d'Horizon. *Fundamenta Informaticae*, 35(1-4):67–89, 1998.
[2] A. E. Eiben and G. Rudolph. Theory of evolutionary algorithms: A bird's eye view. *Theoretical Computer Science*, 229(1-2):3–9, 1999.
[3] A. E. Eiben, E. H. L. Aarts, and K. M. Van Hee. Global convergence of genetic algorithm: a Markov chain analysis. In H.-P Schwefel and R. Männer, editors, *Parallel Problem Solving from Nature*, pages 4–12, Berlin, 1991. Springer.
[4] G. Rudolph. Convergence analysis of canonical genetic algorithms. *IEEE Trans. on Neural Networks*, 5(1):96–101, 1994.
[5] J. Suzuki. A Markov chain analysis on simple genetic algorithms. *IEEE Trans. on Systems, Man and Cybernetics*, 25:655–659, 1995.
[6] R. Cerf. An asymptotic theory of genetic algorithms. In J.-M. Alliot, E. Lutton, E. Ronald, M. Schoenauer, and D. Snyers, editors, *Artificial Evolution*. Springer Verlag, 1996.
[7] G. Rudolph. *Convergence Properties of Evolutionary Algorithms*. Kovač, Hamburg, Germany, 1997.
[8] G. Rudolph. Convergence of non-elitist strategies. In *Proc. of 1st IEEE Conference on Evolutionary Computation*, volume 1, pages 63–66, Piscataway, NJ, 1994. IEEE Press.
[9] H.-G. Beyer. Toward a theory of evolution strategies: On the benefits of sex -the (μ, λ)-theory. *Evolutionary Computation*, 3(1):81–111, 1995.
[10] G. Rudolph. Convergence rate of evolutionary algorithms for a class of convex objective functions. *Control and Cybernetics Journal*, 26(3):375–390, 1997.
[11] J. He and L. S. Kang. On the convergence rate of genetic algorithms. *Theoretical Computer Science*, 229(1-2):23–39, 1999.
[12] T. E. Davis. A simulated annealing like convergence theory for the simple genetic algorithm. In *Proc.of the Fourth Conference on Genetic Algorithms*, San Mateo, CA, 1991. Morgan Kauffman.