An Analysis About the Asymptotic Convergence of Evolutionary Algorithms

Lixin Ding[1] and Jinghu Yu[2]

[1] State Key Lab of Software Engineering, Wuhan University,
Wuhan 430072, China
lxding@whu.edu.cn
[2] Department of Mathematics, School of Natural Sciences,
Wuhan University of Technology,
Wuhan 430070, China
yujh@wipm.ac.cn

Abstract. This paper discusses the asymptotic convergence of evolutionary algorithms based on finite search space by using the properties of Markov chains and Perron-Frobenius Theorem. First, some convergence results of general square matrices are given. Then, some useful properties of homogeneous Markov chains with finite states are investigated. Finally, the geometric convergence rates of the transition operators, which is determined by the revised spectral of the corresponding transition matrix of a Markov chain associated with the EA considered here, are estimated by combining the acquired results in this paper.

1 Introduction

Evolutionary algorithms(EAs for brevity) are a class of useful optimization methods based on a biological analogy with the natural mechanisms of evolution, and they are now a very popular tool for solving optimization problems. An EA is usually formalized as a Markov chain, so one can use the properties of Markov chains to describe the asymptotic behaviors of EAs, i.e., the probabilistic behaviors of EAs if never halted. Asymptotic behaviors of EAs has been investigated by many authors [1−12]. Due to the connection between Markov chains and EAs, a number of results about the convergence of EAs have been obtained by adopting the limit theorem of the corresponding Markov chin in the above works. In this paper, we will make further research on this topic, especially on convergence rate of EAs by using Perron-Frobenius Theorem and other analytic techniques.

The remaining parts of this paper are organized as follows. In section 2, we apply some basic matrix theory, such as Jordan Standard Form Theorem and Perron-Frobenius Theorem etc., to study the convergence of general square matrix A. We obtain that A^n converge with geometric convergence rate defined by the revised spectral of A. In section 3, we concern on homogeneous Markov chains with finite states. We give the relations among states classification, geometric convergence rate and eigenvalues of transition matrix. In section 4, we combine the results in section 2 and section 3 to investigate the limit behaviors

Y. Wang, Y. Cheung, and H. Liu (Eds.): CIS 2006, LNAI 4456, pp. 156–166, 2007.

of EAs. Under some mild conditions, we get that EAs converges to the optimal solution set related to the given problem with geometrical rate which is determined by the revised spectral of corresponding transition matrix of a Markov chain associated with the EA considered in this paper. Finally, we conclude this paper with a short discussion in section 5.

2 Preliminaries

In this section, we need to collect a number of definitions and elementary facts with respect to matrix classification, matrix decomposition and matrix convergence which will be useful throughout the whole paper. For a detailed reference on matrix theory, see the monograph by Steward[13]

Definition 1. *A $m \times m$ square matrix $\mathbf{A}$ is said to be*
(1) nonnegative($\mathbf{A} \geq 0$), if $a_{ij} \geq 0$ for all $i, j \in \{1, 2, \cdots, m\}$,
(2) positive($\mathbf{A} > 0$), if $a_{ij} > 0$ for all $i, j \in \{1, 2, \cdots, m\}$.

A nonnegative matrix $\mathbf{A} : m \times m$ is said to be
(3) primitive, if there exists a positive integer k such that $\mathbf{A}^k$ is positive,
(4) reducible, if there exists a permutation matrix $\mathbf{B}$ such that

$$\mathbf{BAB}^T = \begin{pmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{R} & \mathbf{T} \end{pmatrix},$$

where square matrix $\mathbf{C}$ and $\mathbf{T}$ are square matrices,
(5) irreducible, if it is not reducible,
(6) stochastic, if $\sum\limits_{j=1}^{m} a_{ij} = 1$ for all $i \in \{1, 2, \cdots, m\}$.
A $m \times m$ stochastic matrix $\mathbf{A}$ is said to be
(7) stable, if it has identical rows.

Definition 2. *For a square matrix $\mathbf{A} : m \times m$ with eigenvalues $\lambda_1, \cdots, \lambda_m$, its revised spectral gap is usually defined as $r(\mathbf{A}) = \max\{|\lambda_i| : |\lambda_i| \neq 1, i = 1, \cdots, m\}$, and its norm is defined as $||\mathbf{A}|| = \max\{|a_{ij}| : i, j = 1, \cdots, m\}$.*

The following two Lemmas are well-known and can be found in many literatures of matrix theory.

Lemma 1 (Jordan Standard Form Theorem). *Suppose that square matrix $\mathbf{A} : m \times m$ has r different eigenvalues $\lambda_1, \cdots, \lambda_r$. Then there exists an invertible matrix $\mathbf{B}$ such that*

$$\mathbf{B}^{-1}\mathbf{AB} = \mathbf{J} \equiv diag[\mathbf{J}(\lambda_1), \cdots, \mathbf{J}(\lambda_r)],$$

where

$$\mathbf{J}(\lambda_i) = \begin{pmatrix} \lambda_i & 0 & \cdots & 0 & 0 \\ 1 & \lambda_i & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 1 & \lambda_i & 0 \\ 0 & 0 & \cdots & 1 & \lambda_i \end{pmatrix}$$

$$\in \mathbf{C}^{n(\lambda_i)\times n(\lambda_i)}, 1 \leq i \leq r,$$

and $\sum_{i=1}^{r} n(\lambda_i) = m.$

Lemma 2 (Perron-Frobenius Theorem). *For any nonnegative square matrix* $\mathbf{A} : m \times m$*, the following claims are true.*

(1)There exists a non-negative eigenvalue λ *such that there are no other eigenvalues of* $\mathbf{A}$ *with absolute values greater than* λ*;*

(2) $\min_i(\sum_{j=1}^{m} a_{ij}) \leq \lambda \leq \max_i(\sum_{j=1}^{m} a_{ij})$.

By using the above matrix theorems, we can get the following convergence results about $\mathbf{A}^n$ as n tends to infinity.

Proposition 1. *Suppose that* 1 *is a simple eigenvalue of square matrix* $\mathbf{A} : m \times m$ *and all other eigenvalues have absolute values less than* 1*. Then* $\lim_{n\to\infty} \mathbf{A}^n$ *exists and has geometric convergence rate.*

Proof. Let $\lambda_1, \lambda_2, \cdots, \lambda_{m-1}$ be those eigenvalues with absolute values less than 1. By Lemma 1, we know that the Jordan form of $\mathbf{A}$ is as follows

$$\begin{pmatrix} B_1 & 0 & \cdots & 0 & 0 \\ 0 & B_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & B_t & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix},$$

where square matrices $\mathbf{B}_i : q_i \times q_i (q_i$ is the algebra multiplicity of $\lambda_i), i = 1, 2, \cdots, t$, are Jordan blocks with the above form.

Note that the elements of $\mathbf{B}_i^k$ are $0, \lambda_i^k, C_k^1\lambda_i^{k-1}, C_k^2\lambda_i^{k-2}, \cdots, C_k^{q_i-1}\lambda_i^{k-q_i+1}$. It is easy to check that $||\mathbf{B}_i^k|| \to 0 (i = 1, \cdots, m-1)$ as $k \to \infty$. Moreover, for fixed q_i, when k is big enough, $C_k^{q_i-1}|\lambda_i|^{k-q_i+1}$ is the biggest elements among $\{0, |\lambda_i|^k, C_k^1|\lambda_i|^{k-1}, C_k^2|\lambda_i|^{k-2}, \cdots, C_k^{q_i-1}|\lambda_i|^{k-q_i+1}\}$; And, for fixed $q_i \leq m$, when k is big enough, $C_k^{q_i-1} \leq C_k^m$. In addition, there exists an invertible matrix $\mathbf{T}$ such that

$$\mathbf{A} = \mathbf{T}^{-1} \times \begin{pmatrix} \mathbf{B}_1 & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{B}_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mathbf{B}_t & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \times \mathbf{T}.$$

If we write

$$\mathbf{B}^* = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

and let $\Pi = \mathbf{T}^{-1}\mathbf{B}^*\mathbf{T}$, then

$$||\mathbf{A}^k - \Pi|| \leq ||\mathbf{T}^{-1}|| \cdot ||\mathbf{T}|| \cdot C_k^m (r(\mathbf{A}))^{k-m+1} \leq c \cdot k^m (r(\mathbf{A}))^k \to 0 (k \to \infty). \quad (1)$$

Note that, for any given $0 < \varepsilon < 1$, $k^m (r(\mathbf{A}))^{\varepsilon k} \to 0 (k \to \infty)$. Hence, for the fixed m and $r(\mathbf{A})$, we have $k^m (r(A))^{\varepsilon k} \leq 1$ as $k \to \infty$. By (1), when k is big enough, we have

$$||\mathbf{A}^k - \Pi|| \leq c \cdot (r(\mathbf{A}))^{(1-\varepsilon)k}, \quad (2)$$

which means that $\mathbf{A}^n$ has geometric convergence rate. □

Proposition 2. *Suppose that square matrix* $\mathbf{A} : m \times m$ *has* m *linear independent eigenvectors and its eigenvalues except* 1 *have absolute values less than* 1*. Then* $\lim_{n \to \infty} \mathbf{A}^n$ *exists and has geometric convergence rate determined by* $r(\mathbf{A})$*.*

Proof. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_q (q < m)$ be eigenvalues of $\mathbf{A}$ not equal 1. Then, we have from the assumption of Proposition 2 that

$$|\lambda_i| < 1, \forall i = 1, \cdots, q.$$

By matrix theory, there exists an invertible matrix $\mathbf{T}$ and the following diagonal matrix

$$\mathbf{B} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & \cdots\cdots\cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots\cdots\cdots\cdots & & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \lambda_q & 0 & \cdots\cdots\cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 \cdots\cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 \cdots 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots\cdots\cdots & & & 0 & 1 \end{pmatrix}$$

such that $\mathbf{A} = \mathbf{T}^{-1}\mathbf{B}\mathbf{T}$. Therefore, we have $\mathbf{A}^k = \mathbf{T}^{-1}\mathbf{B}^k\mathbf{T}$. Write

$$\mathbf{B}^* = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

and let $\Pi = \mathbf{T}^{-1}\mathbf{B}^*\mathbf{T}$. Then

$$\begin{aligned} ||\mathbf{A}^k - \Pi|| &= \mathbf{T}^{-1}(\mathbf{B}^k - \mathbf{B}^*)\mathbf{T} \\ &\leq ||\mathbf{T}^{-1}|| \cdot \max\{|\lambda_k| : k = 1, \cdots, q\} \cdot ||\mathbf{T}|| \\ &= c \cdot r(\mathbf{A})^k \to 0 (k \to \infty). \end{aligned}$$

□

3 Homogeneous Markov Chains with Finite States

Since the limit behaviors of Markov chains depend on the structure of their transition matrixes, the properties of transition matrixes are very useful to describe the limit behaviors of Markov chains. In this section, we will introduce some indexes and definitions at first. Then, we will pay our attention on homogenous Markov chains with finite states space.

Let $\mathbf{P}$ be the transition matrix associated with Markov Chain $\{X_n; n \geq 0\}$ defined on a finite state space $S = \{s_1, s_2, \cdots, s_m\}$. We will also classify the state space in the following.

Definition 3. *(1) a vector:* $v = (v_1, \cdots, v_m)$ *is called a probability vector if* $v_i \geq 0$ *and* $\sum_{i=1}^{m} v_i = 1$*,*

(2) a probability vector v *is called an invariant probability measure(stationary distribution) of transition matrix* $\mathbf{P}$*: if* $v\mathbf{P} = v$*.*

The following notations are usually needed to classify the states of Markov chains.

$f_{ij}^n \doteq P\{X_0 = i, X_1 \neq j, \cdots, X_{n-1} \neq j, X_n = j\}$, is the probability that Markov chain starts at state s_i and reaches state s_j at time n for the first time;

$f_{ij}^* \doteq \sum_{n=1}^{\infty} f_{ij}^n$, is the probability that Markov chain starts at s_i and reaches s_j after finite steps;

$m_{ii} \doteq \infty$, if $f_{ii}^* < 1$; otherwise $m_{ii} \doteq \sum_{n=1}^{\infty} n f_{ii}^n$;

$d_i \doteq$ the biggest common divisor of $\{n : p_{ii}^n > 0\}$, is called the period of state s_i

Definition 4. *The state* s_j *is called a*

(1) transient state, if $f_{jj}^* < 1$*;*
(2) recurrent state, if $f_{jj}^* = 1$*;*
(3)positive recurrent, if $m_{jj} < \infty$*;*
(4)zero recurrent, if s_j *is not a positive recurrent;*
(5)aperiodic, if $d_i = 1$*.*

In the following, we will further describe the states classification of Markov chains. Let $N \subset S$ be the collection of all transient states of S, R^+ be the collection of all positive recurrent states, and R^0 be the collection of all zero recurrent states of S. Then $S = N \bigcup R^0 \bigcup R^+$. Furthermore, R^0 and R^+ can be divided into some irreducible sub-classes, that is, $R^0 = R_1^0 + \cdots + R_i^0$ and $R^+ = R_1^+ + \cdots + R_j^+$.

For Markov chain with finite states, it is well-known that

$$\lim_{k \to \infty} \frac{1}{k} \sum_{l=1}^{k} P_{ij}^l = \Pi_{ij}, \forall i, j \in S. \tag{3}$$

Researchers can refer to relative limit theorems, such as Proposition 3.3.1 in [14]. Moreover, since $\mathbf{P}$ is finite dimensional, hence the limit distribution Π is also a transition matrix on S.

Definition 5. *The subset* $E \subset S$ *is closed if* $i \in E, j \notin E$*, which implies that* $p_{ij} = 0$*, i.e., if* $i \in E$ *then* $\sum_{j \in E} P_{ij} = 1$*. The state space* S *is called reducible, if* S *have no-empty closed subset; otherwise,* S *is irreducible.*

In fact, S is reducible(irreducible) $\Leftrightarrow$ transition matrix $\mathbf{P}$ on state space S is reducible(irreducible).

We have another important fact that if every positive state of $\mathbf{P}$ is aperiodic, then $\lim\limits_{k\to\infty} \mathbf{P}^k$ exists. Combining Proposition 1 and Proposition 2 as well as Theorem 16.0.1 and Theorem 16.0.2 in [14], we can get the following conclusion immediately.

Proposition 3. *Give a Markov chain with transition matrix* $\mathbf{P}: m\times m$ *on finite state space, for the following statements*

(1) $\mathbf{P}$ *is aperiodic,*

(2) $\mathbf{P}^k$ *has geometric convergence rate,*

(3) 1 *is a simple eigenvalue and all other eigenvalues have absolute values less than* 1*,*

(4) P *has* m *linearly independent eigenvectors and and the eigenvalues except* 1 *have absolute values less than* 1*,*

then the relations among them are that

$$(1) \Leftrightarrow (2); \quad (3) \Rightarrow (2); \quad (4) \Rightarrow (2).$$

For a reducible stochastic matrix, there is a very important convergence theorem given by M. Iosifescu[15], which is

Lemma 3. *Let* $\mathbf{P}$ *be a reducible stochastic matrix, where* $\mathbf{C}$ *is a primitive stochastic matrix and* $\mathbf{R}, \mathbf{T} \neq \mathbf{0}$*. Then*

$$P^{\infty} = \lim_{k\to\infty} P^k = \begin{pmatrix} \mathbf{C}^{\infty} & \mathbf{0} \\ \mathbf{R}_{\infty} & \mathbf{0} \end{pmatrix}$$

is a stable stochastic matrix.

In the following, Π is always defined as in Proposition 1 or Proposition 2. It is obvious that

$$\Pi\mathbf{P} = \mathbf{P}\Pi = \Pi = \Pi^2.$$

Thus, we have $(\mathbf{P}-\Pi)^k = \mathbf{P}^k - \Pi, \forall k \geq 1$. Moreover, by Proposition 1 and 2, $\mathbf{P}$ has geometric convergence rate, hence $\sum\limits_{k=1}^{\infty} ||\mathbf{P}^k - \Pi|| < \infty$. Thus, if let $\mathbf{Z} = \mathbf{I} + \sum\limits_{k\geq 1} (\mathbf{P}-\Pi)^k = \frac{\mathbf{I}}{\mathbf{I}-\mathbf{P}+\Pi}$, then $\mathbf{Z}$ is **well-defined** and $\mathbf{Z} = (\mathbf{I}-\mathbf{P}+\Pi)^{-1}$. We can prove that $\mathbf{Z}$ has the following properties.

Proposition 4. *(1)* $(\mathbf{I}-\mathbf{P})\mathbf{Z} = \mathbf{Z}(\mathbf{I}-\mathbf{P}) = \mathbf{I} - \Pi$*,*

(2)$\Pi\mathbf{Z} = \Pi, \mathbf{Z}\mathbf{1} = \mathbf{1}$*,*

(3) all eigenvectors of $\mathbf{P}$ *are those of* $\mathbf{Z}$*; moreover, if* $r_i (\neq 1)$ *is a eigenvalue of* $\mathbf{P}$*, then* $\frac{1}{1-r_i}$ *is the eigenvalue of* $\mathbf{Z}$*.*

Proof. Because (1) and (2) of Proposition 4 are easy to be checked, we only check (3) of Proposition 4 here.

For a vector ν, notice the fact that

$$\mathbf{P}\nu = \nu \Longrightarrow \Pi\nu = \nu \Longrightarrow \mathbf{Z}\nu = \nu$$

$$\nu\mathbf{P} = \nu \Longrightarrow \nu\Pi = \nu \Longrightarrow \nu\mathbf{Z} = \nu.$$

Hence, 1 is a eigenvalue of $\mathbf{Z}$ and those eigenvectors of $\mathbf{P}$ corresponding to 1 are also those of $\mathbf{Z}$. In addition, for all other eigenvalues $|\lambda_k| < 1$ of $\mathbf{P}$, let ν_k be a right eigenvector of $\mathbf{P}$, that is

$$\mathbf{P}\nu_k = \lambda_k\nu_k.$$

Then, $\Pi\mathbf{P} = \Pi$ implies that $\Pi\nu_k = \Pi\mathbf{P}\nu_k = \lambda_k\Pi\nu_k$. If $\lambda_k \neq 1$, then we have $\Pi\nu_k = 0$. Note that

$$\mathbf{Z}\nu_k = \frac{1}{1-\lambda_k}\nu_k. \tag{4}$$

If $\lambda_k \neq 1$, then (4) means that ν_k is right eigenvector of $\mathbf{Z}$ corresponding to eigenvalue $\frac{1}{1-\lambda_k}$. In addition, we have $\Pi\mathbf{Z} = \Pi$, which means that 1 is eigenvalue of $\mathbf{Z}$ corresponding to eigenvector π. The same process can be applied to check left eigenvectors of $\mathbf{P}$. Therefore, this is the proof of (3) of Proposition 4. □

It is easy to know from Perron- Frobenius theorem that if $\mathbf{P}$ is a transition matrix, then 1 is a eigenvalue of $\mathbf{P}$ and there is no other eigenvalues with absolute values greater than 1. This fact implies that $r(\mathbf{P}) \leq 1$.

4 Asymptotic Behaviors of Evolutionary Algorithms

In this section, we consider the following optimization problem: Given an objective function $f : S \to (-\infty, \infty)$, where $S = \{s_1, s_2, \cdots, s_M\}$ is a finite search space. A maximization problem is to find a $x^* \in S$ such that

$$f(x^*) = \max\{f(x) : x \in S\}. \tag{5}$$

We call x^* an **optimal solution** and write $f_{\max} = f(x^*)$ for convenience. If there are more than one optimal solution, then denote the set of all optimal solutions by S^* and call it an **optimal solution set**. Moreover, **optimal populations** refer to those which include at least an optimal solution and the **optimal population set** consists of all the optimal populations.

An evolutionary algorithm with population size $N(\geq 1)$ for solving the optimization problem (5) can be generally described as follows:

step 1. initialize, either randomly or heuristically, an initial population of N individuals, denoted it by $\xi_0 = (\xi_0(1), \cdots, \xi_0(N))$, where $\xi_0(i) \in S, i = 1, \cdots, N$, and let $k = 0$.

step 2. generate a new (intermediate) population by adopting genetic operators (or any other stochastic operators for generating offsprings), and denote it by $\xi_{k+1/2}$.

step 3. select N individuals from populations $\xi_{k+1/2}$ and ξ_k according to certain select strategy , and obtain the next population ξ_{k+1}, then go to step 2.

For convenience, we write that

$$f(\xi_k) = \max\{f(\xi_k(i)) : 1 \leq i \leq N\}, \forall k = 0, 1, 2, \cdots,$$

which represents the maximum in populations $\xi_k, k = 0, 1, 2, \cdots$.

It is well-known that $\{\xi_k; k \geq 0\}$ is a Markov chain with the **state space** S^N because the states of the $(k+1)-th$ generation only depend on the $k-th$ generation. In this section, we assume that the stochastic process, $\{\xi_k; k \geq 0\}$, associated with an EA, is a **homogeneous Markov chain**, and denote its transition probability matrix by $\mathbf{P}$. It is easy to check the following results.

Remark 1. If the selection strategy in step 3 of the EA can lead to the fact that

$$f(\xi_k) \leq f(\xi_{k+1}), \tag{6}$$

then the corresponding transition matrix $\mathbf{P}$ is reducible.

The selection with the property of equation (6) is the so-called elitist selection, which insures that if the population has reached the optimal solution set, then the next generation population cannot reach any other states except those corresponding to the optimal population set. In practical, a lot of EAs have this kind of property. Hence, we always assume that EAs considered here possess the property of equation(6).

Remark 2. If population size $N = 1$ and the optimization problem has only one optimal solution, then

$$\Pi = \mathbf{P}^{\infty} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

Remark 3. If population size $N \geq 1$ and the optimization problem has only one optimal solution, then

$$\Pi = \mathbf{P}^{\infty} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} & 0 & \ldots & 0 \\ a_{21} & a_{22} & \cdots & a_{2m} & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{q1} & a_{q2} & \cdots & a_{qm} & 0 & \ldots & 0 \end{pmatrix},$$

where $q = M^N$, and the former m elements in matrix $\mathbf{P}$ exactly correspond to the m optimal states.

The remark 2 and 3 can be followed by Lemma 3 immediately.

Remark 4. For any initial distribution v_0, $v_k \doteq v_0 \mathbf{P}^k \to (b_1, b_2, \cdots, b_m, 0, \cdots, 0)$ $(k \to \infty)$, which implies that $P(\lim_{k\to\infty} \xi_k \in S^*) = 1$, that is, EAs converges to optimal solution in probability.

In the following, we will prove the main results in this paper.

Theorem 1. *Suppose the optimization problem has only one optimal solution* x^* *and the population size* $N = 1$. *If* $P\{\xi_1 = x^* | \xi_0 = s_j\} > 0$ *for all* $s_j \neq x^*$, *then*

(1) all states except x^* are transient;
(2)x^* is positive recurrent and aperiodic;
(3) $\mathbf{P}^k$ converges, and if writing the limit by Π, then

$$\Pi = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

Proof. Note that $P(\xi_1 = x^*|\xi_0 = s_j) > 0$ and $P(\xi_1 = x^*|\xi_0 = x^*) = 1$. So, we have $f_{jj}^* < 1$ for all $s_j \neq x^*$, which means that $s_j(\neq x^*)$ is transient. This completes the proof of (1).

Since $\mathbf{P}$ is finite dimensional matrix, the positive recurrent states are not empty. Hence, x^* must be positive recurrent by (1) of this theorem. Combine the above fact and $P(\xi_1 = \xi^*|\xi_0 = \xi^*) = 1$, we get that x^* is aperiodic. This is (2).

By Remark 1, we know that $\lim_{k\to\infty} \mathbf{P}^k$ exists and the limit Π has the given form of (3). □

In order to deal with more complicate cases, such as f is not 1-1 and population size $N \geq 1$, we will introduce the following analytic techniques.

Denote the elements in image space of f by $I_f = \{y_1, \cdots, y_q\}$. For $i = 1, \cdots, q$, the **level sets** of original state space S^N are defined by

$$S_i = \{(x_1, \cdots, x_N) \in S^N : \max\{f(x_1), \cdots, f(x_N)\} = y_i\}.$$

Define new transition matrix $\overline{\mathbf{P}}(k)$ on new state space $\{S_1, S_2, \cdots, S_q\}$ by

$$\overline{p}_{ij}(k) = \frac{\sum\limits_{x\in S_i, z\in S_j} P(\xi_{k+1} = z, \xi_k = x)}{\sum\limits_{x\in S_i} P(\xi_k = x)}, \forall S_i, S_j.$$

We can check that $\overline{p}_{ij}(k) = \overline{p}_{ij}(1) \doteq \overline{p}_{ij}, \forall k \geq 1$, which means that $\overline{\mathbf{P}}(k)$ is homogenous. In particular, let $C^* = \{(s_1, \cdots, s_N) \in S^N : \max\{f(s_1), \cdots, f(s_N)\} = f_{max}\}$ be the optimal population set. Then

$$\overline{p}_{ij} = 0, \quad if \quad S_i = C^*, S_j \neq C^*$$

$$\overline{p}_{ii} = 1, \quad if \quad S_i = C^*.$$

Consider new stochastic process $\{\overline{\xi}_k; k \geq 1\}$ defined on new state space $\overline{S} = \{S_1, \cdots, S_q\}$, the distribution of $\overline{\xi_k}$ is given by $P\{\overline{\xi_k} = S_i\} = P\{\xi_k \in S_i\}$. Obviously, $\{\overline{\xi}_k; k \geq 0\}$ is a homogenous Markov chain with transition matrix $\overline{\mathbf{P}}(k)$. We can get the following general results

Theorem 2. *If $P\{\overline{\xi}_1 = C^*|\overline{\xi}_0 = S_j\} > 0$ for all $S_j \neq C^*$, then transition matrix $\overline{\mathbf{P}}$ has the following properties*

(1) all states in new state space except C^* are transient;
(2)C^* is positive recurrent and aperiodic;
(3) $\lim_{k\to\infty} \overline{\mathbf{P}}^k$ exists, and if writing the limit by $\overline{\Pi}$ then

$$\overline{\Pi} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

The proof of this theorem is similar to Thm 1, so we omit it here.

Theorem 3. *If $P\{\overline{\xi}_1 = C^*|\overline{\xi}_0 = S_j\} > 0$ for all $S_j \neq C^*$, then transition matrix $\overline{\mathbf{P}}^k$ has geometric convergence rate determined by $r(\overline{\mathbf{P}})$.*

Proof. Note that we can find a permutation matrix $\mathbf{B}$ such that $\mathbf{B}\overline{\mathbf{P}}\mathbf{B}^T$ is a upper triangular matrix and its diagonal elements are $P\{\overline{\xi_1} = S_j|\overline{\xi_0} = S_j\}$. By the properties of transition matrices corresponding to the EA, 1 is a simple one in diagonal elements and all other diagonal elements are real and less than 1. Similar to Proposition 1, the transition matrix $\mathbf{B}\overline{\mathbf{P}}\mathbf{B}^T$ has geometric convergence rate. Hence, $\overline{\mathbf{P}}$ has also geometric convergence rate determined by $r(\overline{\mathbf{P}})$. □

5 Conclusions and Discussions

This paper confirms mathematically some results on asymptotic behaviors of evolutionary algorithms. Several important facts of the asymptotic behaviors of evolutionary algorithms, which make us understand evolutionary algorithms better, are proved theoretically. From this paper, we know that the convergence rate of EAs is determined by the spectrum radius of transition matrix, so, if the spectrum radium of the transition matrixes of Markov chain associated with the evolutionary algorithm becomes much smaller, the EA will converge much faster. For the simplest case that the objective function is $1-1$, the spectrum radium $r = \max\{P(\xi_{k+1} = s_j|\xi_k = s_j) : s_j \neq x^*\}$. So, we must make $\max\{P(\xi_{k+1} = s_j|\xi_k = s_j) : s_j \neq x^*\}$ become as small as possible in order to attain a fast convergence speed.

In fact, there are still a number of open problems for the further investigation such as, what effect on asymptotic behaviors will be brought by selection strategy, genetic operators and population size, respectively; the question of non-asymptotic behaviors(when the number of iterations depends in some way of the population size); and others. Probably, one can think of many variants and generalization of the algorithm, but the results we obtained in this paper incite us to go on studying simplified models of evolutionary algorithms in order to improve our understanding of their asymptotic behaviors.

Acknowledgments. This work is supported in part by the National Natural Science Foundation of China(Grant no. 60204001), Chengguang Project of Science and Technology for the Young Scholar in Wuhan City (Grant no. 20025001002) and the Youthful Outstanding Scholars Foundation in Hubei Prov. (Grant no. 2005ABB017).

References

1. Agapie, A.: Theoretical analysis of mutation-adaptive evolutionary algorithms. Evolutionary Computation 9, 127–146 (2001)
2. Cerf, R.: Asympototic Convergence of Genetic Algorithms. Advances in Applied Probablity 30, 521–550 (1998)
3. He, J., Kang, L.: On the convergence rate of genetic algorithms. Theoretical Computer Science 229, 23–39 (1999)
4. Lozano, J.A., et al.: Genetic algorithms: bridging the convergence gap. Theoretical Computer Science 229, 11–22 (1999)
5. Nix, A.E., Vose, D.E.: Modeling genetic algorithms with Markov chains. Annals of Mathematics and Artificial Intelligence 5, 79–88 (1992)
6. Poli, R., Langdon, M.: Schema theory for genetic programming with one-point crossover and point mutation. Evolutionary Computation 6, 231–252 (1998)
7. Poli, R.: Exact schema theory for genetic programming variable-length genetic algorithms with one-point crossover. Genetic Programming and Evolvable Machines 2, 123–163 (2001)
8. Rudolph, G.: Convergence analysis of canonical genetic algorithms. IEEE Transactions on Neural Networks 5, 96–101 (1994)
9. Rudolph, G.: Convergence Properties of Evolutionary Algorithms. Verlag Dr. Kovac, Hamburg (1997)
10. Schmitt, L.M.: Theory of genetic algorithms. Theoretical Computer Science 259, 1–61 (2001)
11. Suzuki, J.: A further result on the Markov chain model of genetic algorithms and its application to a simulated annealing-like strategy. Man and Cybernetics-Part B 28, 95–102 (1998)
12. Vose, D.: The Simple Genetic Algorithms: Foundations and Theory. MIT Press, Cambridge (1999)
13. Steward, G.W.: Introduction to Matrix Computation. Academic Press, New York (1973)
14. Meyn, S.P., Tweedie, R.L.: Markov Chains and Stochastic Stability, 3rd edn. Springer-Verlag, New York (1996)
15. Isoifescu, M.: Finite Markov Processes and Their Applications. Wiley, Chichester (1980)