

Some Theoretical Results About the Computation Time of Evolutionary Algorithms

Lixin Ding
State Key Lab of Software Engineering
Wuhan University
Wuhan, 430072, P. R. China
lx_ding@263.net

Jinghu Yu
Wuhan Institute of Physics and Mathematics
Chinese Academy of Sciences
Wuhan, 430071, P. R. China
yujh@wipm.ac.cn

ABSTRACT

This paper focuses on the computation time of evolutionary algorithms. First, some exact expressions of the mean first hitting times of general evolutionary algorithms in finite search spaces are obtained theoretically by using the properties of Markov chain. Then, by introducing drift analysis and applying Dynkin's Formula, the general upper and lower bounds of the mean first hitting times of evolutionary algorithms are given rigorously under some mild conditions. These results obtained in this paper, and the analytic methods used in this paper, are widely valid for analyzing the computation time of evolutionary algorithms in any search space (finite or infinite) as long as some simple technique processes are introduced.

Categories and Subject Descriptors

F.2 [Analysis of Algorithms and Problem Complexity]: Miscellaneous; I.2.8 [Problem Solving, Control Methods, and Search]: Complexity—*computation time*

General Terms

Algorithms

Keywords

evolutionary algorithms, computation time, drift analysis, Dynkin's formula

1. INTRODUCTION

The research for the computation time of evolutionary algorithms (EAs for brevity) used to solve optimization problems is an important topic in the foundations and theory of EAs, which reveals the number of expected generations needed to reach an optimal solution [1, 2]. In the last over ten years, some progresses have been made towards this direction: Bäck [3] and Mühlenbein [4] studied the time com-

plexity of EAs based on the simple *ONE-MAX problem*. Rudolph [5] gave a comprehensive survey of the theoretical works up to 1997 and provided an $O(n \log n)$ upper bound for the $(1 + 1)$ -EA using the 1-bit-flip mutation for *ONE-MAX problem*. Garnier et al [6] compared two different mutations in $(1 + 1)$ -EAs when they are applied to the *ONE-MAX problem*, and obtained the different bounds on the EA's average computation time, respectively. Droste et al [7, 8] improved these results and generalized them to any linear binary functions for the $(1 + 1)$ -EA. Some *long path problems* in unimodal functions have also proved to be solvable in polynomial time [9, 10]. He and Yao, who have done a series of works about the computation time and the time complexity for several kinds of EAs and different problems [11 – 16], are specially deserved to be mentioned.

Markov chain models have been used widely in the theoretical analysis of EAs [17 – 19]. Although drift analysis introduced from stochastic process is a very useful technique in estimating computation time and time complexity for the stochastic algorithms [20, 11, 13 – 16], most of results obtained till now almost focus on some simple evolutionary algorithms and optimization problems, such as $(1 + 1)$ -EAs, $(N + N)$ -EAs, *ONE-MAX problem*, linear functions, etc. It is important for us to develop other mathematical methods and tools to analyze rigorously more complex EAs based on more complex problems so that insights can be gained into them.

As we know that the first hitting time of a Markov chain plays an important role in analyzing the computation time and time complexity of EAs. And some initial successes have been made in this topic. However, few rigorous results are available to analyze the time complexity and the computation time of EAs. He and Yao [13] built a general framework for analyzing the average first hitting time of EAs based on their absorbing Markov chain models. In their framework, the time complexity for some typical EAs based on some model problems are traversed. He and Yao's works [11 – 16] provided some good and valuable ideas to make further researches on the computation time and time complexity of EAs. But, it is still necessary to do further researches for obtaining some thorough and profound theoretical results, and furthermore, for improving on the former works.

In this paper we consider a Markov chain associated with a general EA based on a finite search space. By introducing the definitions of the first hitting times of EAs, some exact expressions of the mean first hitting times of EAs are

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO'05, June 25–29, 2005, Washington, DC, USA.
Copyright 2005 ACM 1-59593-010-8/05/0006 ...\$5.00.

obtained according to the definitions. The other results in this paper concern with the upper and lower bounds of the mean first hitting times of EAs by applying Dynkin's Formula and some advanced analytic techniques. In fact, it is a tough task for us to estimate the upper and lower bounds of the mean first hitting times of EAs by using our limited analytic techniques and available Markov chain theory [21].

The remaining parts of this paper are organized as follows. In section 2, we describe the formalization models of EAs. In section 3, we obtain some exact expressions of the mean first hitting times of EAs. In section 4, we give the upper and lower bounds of the mean first hitting times of EAs under some mild conditions by using different analytic techniques, respectively. In the final section, we conclude the paper with a short discussion and moreover, present some key open problems which are necessary to be solved urgently in the field of the time complexity of EAs in the future.

2. DESCRIPTION OF THE MODEL

In this paper, we consider the following optimization problem: Given an objective function $f : S \rightarrow R$, where S is a finite search space and R is the real line, a maximization problem is to find a $x^* \in S$ such that

$$f(x^*) = \max\{f(x) : x \in S\}. \quad (1)$$

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**.

The formalization model of evolutionary algorithms with the population size N for solving the optimization problem (1) can be generally described as follows:

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

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

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

In the above algorithm, we write $f(\xi_k) = \max\{f(\xi_k(i)) : 1 \leq i \leq N\}, \forall k = 0, 1, 2, \dots$.

It is well known that $\{\xi_k; k \geq 0\}$ is a Markov chain with **the state space** S^N because the state of the $(k+1)$ -th generation often depends only on the k -th generation [1]. The concept of first hitting time of a Markov chain has been widely used in various areas stretching from search problems. In this paper, we will consider the mean first hitting time of general EAs described as above. That is to say, we will research the number of generations for the EA to find an optimal solution for the first time under the context of expectation.

Let $d'(\cdot)$ be a given non-negative **test function defined on** S . Usually, d' is taken as the distance between the individual and the optimal solution(or optimal solution set). For example, we can define it by $f_{\max} - f(\cdot)$. For a population

$\xi = (\xi(1), \dots, \xi(N)) \in S^N$, define

$$d(\xi) = \min\{d'(\xi(i)) : i = 1, \dots, N\}. \quad (2)$$

Then d is also a non-negative **test function defined on** S^N and it is used to measure the distance between the population and the **optimal population (or optimal population set)**. In this paper, the optimal populations refer to those which include at least an optimal solution, while the optimal population set consists of all the optimal populations. The optimal population set associated with d is defined by

$$(S_d^N)^* = \{\xi \in S^N : d(\xi) = 0\}. \quad (3)$$

For convenience, we write $C^* = (S_d^N)^*$.

The one-step drift of stochastic sequence $\{\xi_k; k \geq 0\}$ at time k is generally defined by

$$\Delta(d(\xi_k)) = d(\xi_{k+1}) - d(\xi_k). \quad (4)$$

The one-step drift describes the local performance of EAs [22]. By using some techniques of the stochastic process, we can analyze the mean first hitting time on C^* under some mild conditions imposed on the one-step drift.

Let $N \geq 1$ be a fixed integer, which represents the population size of EAs. Write $Z^+ = \{1, 2, 3, \dots\}$. Throughout this paper, we always assume that the stochastic process, $\{\xi_k; k \geq 0\}$, associated with an EA, is a **finite homogeneous Markov chain**. Let \mathbf{E} denote the expectation operator and $I_A(\cdot)$ an index function on the set A .

In the following section, we will give some exact expressions of the mean first hitting times of EAs.

3. SOME EXACT EXPRESSIONS OF THE MEAN FIRST HITTING TIMES

Let $\{\xi_k; k \geq 0\}$ be a homogeneous Markov chain from probability space (Ω, \mathcal{F}, P) , which can support all randomization used in this paper, to state space S^N associated with an EA described in section 2. Suppose that there are m (Usually, $m = 2^n$, where n is the length of binary bit string) feasible solutions in search space S , thus we can order all states in S^N by s_1, s_2, \dots, s_{m^N} . Let $P_{m^N \times m^N} = (p_{ij})_{m^N \times m^N}$ (where p_{ij} is the transition probability from state s_i to state $s_j, i, j = 1, \dots, m^N$) be **the transition probability matrix** and $q = (q_1, \dots, q_{m^N})$ be **the starting distribution**, that is, $P\{\xi_0 = s_j\} = q_j, j = 1, 2, \dots, m^N$.

In the beginning, we recall the definition of the optimal population. A population, $\xi^* = (\xi^*(1), \dots, \xi^*(N))$, is called an optimal population in S^N , if $\xi^*(j) \in S^*$ for at least one $j (j \in \{1, \dots, N\})$. **The first hitting time on** ξ^* can be defined by

$$\tau(\xi^*) = \min\{k \geq 0 : \xi_k = \xi^*\}. \quad (5)$$

For any given optimal population ξ^* , there exists $i (i \in \{1, \dots, m^N\})$ such that $s_i = \xi^*$. Let us write P_{ξ^*} the $(m^N - 1) \times (m^N - 1)$ matrix obtained from $P_{m^N \times m^N}$ by deleting those elements of its the i -th column and the i -th row. Also let $q_{\xi^*} = (q_1, \dots, q_{i-1}, q_{i+1}, \dots, q_{m^N})$.

Let \mathbf{I} denote the $(m^N - 1) \times (m^N - 1)$ identity matrix and let $\mathbf{1} = (1, 1, \dots, 1)'$ be the $(m^N - 1)$ -dimension vector. Then we have

THEOREM 1. *Let $\tau(\xi^*)$ be the number of generations for the EA to find the optimal population ξ^* for the first time.*

For the optimal population ξ^* , if $\mathbf{I} - P_{\xi^*}$ is invertible, then

$$\mathbf{E}[\tau(\xi^*)] = q_{\xi^*}(\mathbf{I} - P_{\xi^*})^{-1}\mathbf{1}. \quad (6)$$

PROOF. By Markov property of $\{\xi_k; k \geq 0\}$, for any $l \geq 1$, one has

$$\begin{aligned} P\{\tau(\xi^*) \geq l\} &= P\{\xi_0 \neq \xi^*, \xi_1 \neq \xi^*, \dots, \xi_{l-1} \neq \xi^*\} \\ &= \sum_{y_0 \neq \xi^*, y_1 \neq \xi^*, \dots, y_{l-1} \neq \xi^*} P\{\xi_0 = y_0, \dots, \xi_{l-1} = y_{l-1}\} \\ &= \sum_{y_0 \neq \xi^*, y_1 \neq \xi^*, \dots, y_{l-1} \neq \xi^*} P\{\xi_0 = y_0\} \times P\{\xi_1 = y_1 | \xi_0 = y_0\} \\ &\times P\{\xi_2 = y_2 | \xi_1 = y_1\} \times \dots \times P\{\xi_{l-1} = y_{l-1} | \xi_{l-2} = y_{l-2}\} \\ &= q_{\xi^*} P_{\xi^*}^{l-1} \mathbf{1}. \end{aligned}$$

Hence, we have

$$\begin{aligned} \mathbf{E}[\tau(\xi^*)] &= \sum_{k \geq 0} k \times P\{\tau(\xi^*) = k\} \\ &= \sum_{l \geq 1} P\{\tau(\xi^*) \geq l\} \\ &= q_{\xi^*} \cdot \sum_{l \geq 1} P_{\xi^*}^{l-1} \cdot \mathbf{1} \\ &= q_{\xi^*}(\mathbf{I} - P_{\xi^*})^{-1}\mathbf{1}. \end{aligned}$$

This is our assertion. \square

More generally, suppose that $C^* = \{s_{i_1}, \dots, s_{i_r}\} \subset S^N$. We can define the **first hitting time on C^*** by

$$\tau(C^*) = \min\{k \geq 0 : \xi_k \in C^*\}. \quad (7)$$

Similarly, let us denote P_{C^*} the $(m^N - r) \times (m^N - r)$ matrix obtained from $P_{m^N \times m^N}$ by deleting those elements of its the i_1 -th, \dots , the i_r -th columns and the i_1 -th, \dots , the i_r -th rows, and $q_{C^*} = (q_1, \dots, q_{i_1-1}, q_{i_1+1}, \dots, q_{i_r-1}, q_{i_r+1}, \dots, q_{m^N})$. Then we also have

THEOREM 2. Let $\tau(C^*)$ be the number of generations for the populations of the EA to enter the optimal population set C^* for the first time. If $\mathbf{I} - P_{C^*}$ is invertible, then

$$\mathbf{E}[\tau(C^*)] = q_{C^*}(\mathbf{I} - P_{C^*})^{-1}\mathbf{1}, \quad (8)$$

where \mathbf{I} is a $(m^N - r) \times (m^N - r)$ identity matrix and $\mathbf{1} = (1, 1, \dots, 1)'$ is a $(m^N - r)$ -dimension vector.

The proof of Theorem 2 is almost similar to Theorem 1. We omit it here.

Remark 1 In fact, for any set $A \subset S^N$, we can define the first hitting times on A , and at this time Theorem 2 still holds. In addition, although each optimal solution corresponds to many optimal populations which contain this optimal solution, Theorem 1 is still without losing meaning in theory and practice. Usually, the definition (5) is more suitable for $(1+1)$ EAs than the definition (7), while the latter is often used in the case of the population size $N > 1$.

In the above Theorems, we only consider the unconditional expectations of random variables $\tau(\xi^*)$ and $\tau(C^*)$, which can be regarded as the expressions of the mean first hitting times for EAs to find an optimal population under any initialization. By using the same method as the above,

we can obtain the expressions of the conditional expectations $\mathbf{E}[\tau(\xi^*) | \xi_0 = X]$ and $\mathbf{E}[\tau(C^*) | \xi_0 = X]$, for some $X \in S^N$, respectively.

For any optimal population ξ^* and $X \in S^N (X \neq \xi^*)$, there exist i and $j (j \in \{1, \dots, m^N\})$ such that $\xi^* = s_i$ and $X = s_j$. Let v_{X, ξ^*} be the $(m^N - 1)$ -dimension vector obtained from the j -th row of $P_{m^N \times m^N}$ by deleting the i -th element of this row. P_{ξ^*} , \mathbf{I} and $\mathbf{1}$ are the same as Theorem 1. Then we have the following theorem.

THEOREM 3. Let $\tau(\xi^*)$ be the number of generations for the EA to find the optimal population ξ^* for the first time. For the optimal population ξ^* , if P_{ξ^*} and $\mathbf{I} - P_{\xi^*}$ are invertible, then

$$\mathbf{E}[\tau(\xi^*) | \xi_0 = X] = \begin{cases} v_{X, \xi^*} (P_{\xi^*})^{-1} (\mathbf{I} - P_{\xi^*})^{-1} \mathbf{1}, & X \neq \xi^* \\ 0, & X = \xi^* \end{cases} \quad (9)$$

PROOF. By Markov property of $\{\xi_k; k \geq 0\}$, for any $l \geq 1$ and $X \neq \xi^*$, we have

$$\begin{aligned} &P\{\tau(\xi^*) \geq l | \xi_0 = X\} \\ &= \frac{P(\xi_0 = X, \xi_1 \neq \xi^*, \dots, \xi_{l-1} \neq \xi^*)}{P(\xi_0 = X)} \\ &= \sum_{y_1 \neq \xi^*, \dots, y_{l-1} \neq \xi^*} \frac{P(\xi_0 = X, \xi_1 = y_1, \dots, \xi_{l-1} = y_{l-1})}{P(\xi_0 = X)} \\ &= \sum_{y_1 \neq \xi^*, \dots, y_{l-1} \neq \xi^*} P(\xi_1 = y_1 | \xi_0 = X) \times P(\xi_2 = y_2 | \xi_1 = y_1) \\ &\times \dots \times P(\xi_{l-1} = y_{l-1} | \xi_{l-2} = y_{l-2}) \\ &= v_{X, \xi^*} P_{\xi^*}^{l-2} \mathbf{1}. \end{aligned}$$

Hence, by using the same technique as Theorem 1, it is easy for us to get

$$\mathbf{E}[\tau(\xi^*) | \xi_0 = X] = \begin{cases} v_{X, \xi^*} (P_{\xi^*})^{-1} (\mathbf{I} - P_{\xi^*})^{-1} \mathbf{1}, & X \neq \xi^* \\ 0, & X = \xi^* \end{cases}$$

\square

Similarly, for $X = s_j$ and $C^* = \{s_{i_1}, s_{i_2}, \dots, s_{i_r}\}$, let v_{X, C^*} be the $(m^N - r)$ -dimension vector obtained from the j -th row of $P_{m^N \times m^N}$ by deleting those the i_1 -th, the i_2 -th, \dots , the i_r -th elements of this row. P_{C^*} , \mathbf{I} and $\mathbf{1}$ are the same as Theorem 2. Similarly, we have the following theorem.

THEOREM 4. Let $\tau(C^*)$ be the number of generations for the populations of the EA to enter the set C^* for the first time. If P_{C^*} and $\mathbf{I} - P_{C^*}$ are invertible, then

$$\mathbf{E}[\tau(C^*) | \xi_0 = X] = \begin{cases} v_{X, C^*} (P_{C^*})^{-1} (\mathbf{I} - P_{C^*})^{-1} \mathbf{1}, & X \notin C^* \\ 0, & X \in C^* \end{cases} \quad (10)$$

The proof of Theorem 4 is similar to Theorem 3. We also omit it here.

Remark 2 For the EA based on general search space S , we also have the expressions similar to the above theorems, in which the operators will substitute the corresponding matrixes, respectively.

Although the above results have great significance in theory, the estimation of the upper and lower bounds is more useful in practice. Therefore, we will further consider the upper and lower bounds of the mean first hitting times on C^* in the following section.

4. THE UPPER AND LOWER BOUNDS OF THE MEAN FIRST HITTING TIMES

Note that the sequence $\{d(\xi_k) : k = 0, 1, 2, \dots\}$ generated by the EA is also a homogeneous Markov chain, where $d(\cdot)$ is defined in (2). By (7), the first hitting times on C^* are defined by

$$\tau(C^*) = \min\{k \geq 0 : \xi_k \in C^*\} = \min\{k \geq 0 : d(\xi_k) = 0\}. \quad (11)$$

We will impose some constraints on the one-step drift $\Delta(d(\xi_k))$ in order to obtain the upper and lower bounds of $\mathbf{E}[\tau(C^*)|\xi_0 = X]$. Some other marks and definitions should be stated beforehand.

Let $\{\mathcal{F}_n^\xi, n \geq 0\}$ be the σ -algebra given by $\xi_0, \xi_1, \dots, \xi_n$. By Proposition 3.4.4 in [21], $\tau(C^*)$ is a **stopping time** with respect to σ -algebra sequence $\{\mathcal{F}_n^\xi : n \geq 0\}$.

For any $C \subset S^N$, define $\sigma_C = \min\{n \geq 1 : \xi_n \in C\}$, which is **the first return time on C** . **Dynkin's Formula** was usually used to study the upper bound of the mean first return time by controlling the one-step average increment. In this paper, we will use it to estimate the upper and lower bounds of $\tau(C^*)$.

For stopping time $\tau(C^*)$ (τ for brevity in the following) defined in (11), we write

$$\tau^n = \min\{\tau, n, \inf\{k \geq 0 : d(\xi_k) \geq n\}\}, \quad \forall n \in \mathbb{Z}^+.$$

Obviously, τ^n is also a stopping time and furthermore we have

LEMMA 1. (*Dynkin's Formula* [21]) For any $X \in S^N$ and $n \in \mathbb{Z}^+$,

$$\begin{aligned} & \mathbf{E}[d(\xi_{\tau^n})|\xi_0 = X] \\ &= d(X) + \mathbf{E}\left[\sum_{i=1}^{\tau^n} (\mathbf{E}[d(\xi_i)|\mathcal{F}_{i-1}^\xi] - d(\xi_{i-1}))|\xi_0 = X\right]. \end{aligned} \quad (12)$$

Remark 3 If d is a test function from $S^N \rightarrow [0, \infty)$, then (12) still holds for the stopping time $\tau_n = \min\{\tau, n\}$ when n is large enough. In fact, the test function $d(\cdot)$ defined in (2) is non-negative bounded when the state space S^N is finite. Otherwise, one little restriction, $\sup_{X \in S^N} d(X) < \infty$, must be imposed on it.

In the following, we must state another related result in [21], which is

LEMMA 2. (*Theorem 11.3.4 in [21]*) Suppose that there exist some constant $b < \infty$ and an extended real-valued func-

tion $d : S^N \rightarrow [0, \infty]$ such that

$$\mathbf{E}[d(\xi_1) - d(\xi_0)|\xi_0 = X] \leq -1 + bI_C(X), \quad X \in S^N$$

for some set $C \subset S^N$. Then

$$\mathbf{E}[\sigma_C|\xi_0 = X] \leq d(X) + bI_C(X).$$

According to Lemma 2, we can get the following conclusion immediately.

THEOREM 5. Let τ be the number of generations for the populations of the EA to enter the optimal population set C^* for the first time. Suppose d satisfies the following condition

$$\mathbf{E}[d(\xi_1) - d(\xi_0)|\xi_0 = X] \leq -a + bI_{C^*}(X), \quad X \in S^N, \quad (\mathbf{C1})$$

for the constants $a > 0$ and $b < \infty$. Then

$$\mathbf{E}[\tau|\xi_0 = X] \begin{cases} \leq d(X)/a, & X \in S^N \setminus C^* \\ = 0, & X \in C^* \end{cases}$$

PROOF. Note that, if $X \in C^*$, then $\tau = 0$; if $X \notin C^*$, then $\sigma_{C^*} = \tau$. Regarding $d(\cdot)/a$ as the function $d(\cdot)$ of lemma 2, we can get the desired result immediately. \square

In the following, we still put our interests on the special set C^* and give the lower bound for the mean first hitting time on C^* . Dynkin's Formula and some conditions on one-step drift are needed. Our conclusion is

THEOREM 6. Let τ be the number of generations for the populations of the EA to enter the optimal population set C^* for the first time. Suppose d satisfies that

$$-a_2 + a_2I_{C^*}(X) \leq \mathbf{E}[d(\xi_1) - d(\xi_0)|\xi_0 = X] \leq -a_1 + a_1I_{C^*}(X) \quad (\mathbf{C2})$$

for any $X \in S^N$ and some positive constants a_1, a_2 . Then

$$\mathbf{E}[\tau|\xi_0 = X] \begin{cases} \geq d(X)/a_2, & X \in S^N \setminus C^* \\ = 0, & X \in C^* \end{cases}$$

PROOF. Since $\{\xi_k; k \geq 0\}$ is homogenous Markov chain, as implies that if $\mathbf{E}[d(\xi_1) - d(\xi_0)|\xi_0 = X]$ satisfies **(C1)** and **(C2)**, then $\mathbf{E}[d(\xi_{k+1}) - d(\xi_k)|\xi_k = x]$ satisfies **(C1)** and **(C2)** for all $k \geq 1$. Note that if $\omega \in \{\xi_k = X\}$, then we have

$$\mathbf{E}[d(\xi_{k+1})|\mathcal{F}_k^\xi](\omega) = \mathbf{E}[d(\xi_{k+1})|\xi_k = X].$$

Write $Q_k = \bigcup_{X \in S^N \setminus C^*} \{\omega : \xi_k = X\}$, then we have

$$\begin{aligned} \mathbf{E}d(\xi_{k+1}) &= \mathbf{E}[\mathbf{E}[d(\xi_{k+1})|\xi_k]] \\ &= \int_{Q_k} + \int_{\Omega \setminus Q_k} \mathbf{E}[d(\xi_{k+1})|\xi_k]dP \\ &\leq \int_{Q_k} (d(\xi_k) - a_1)dP + \int_{\Omega \setminus Q_k} d(\xi_k)dP \\ &= \mathbf{E}d(\xi_k) - a_1P(Q_k). \end{aligned}$$

By induction on k , we have

$$0 \leq \mathbf{E}d(\xi_{k+1}) \leq \mathbf{E}d(\xi_0) - \sum_{i=0}^k a_1P(Q_i), \quad \forall k \geq 1.$$

Hence, we must have

$$P(Q_k) \rightarrow 0 \quad \text{as } k \rightarrow \infty. \quad (13)$$

Since the state space is finite, (13) implies that

$$\mathbf{E}d(\xi_k) \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

So

$$\mathbf{E}[d(\xi_k)|\xi_0 = X] = \frac{\int_{\xi_0=X} d(\xi_k)dP}{P(\xi_0 = X)} \leq \frac{\mathbf{E}d(\xi_k)}{P(\xi_0 = X)} \rightarrow 0 \quad (14)$$

as $k \rightarrow \infty$. By the hypotheses and Dynkin's Formula, we know that if $X \in S^N \setminus C^*$, then we have

$$\begin{aligned} a_2 \mathbf{E}[\tau^n | \xi_0 = X] &\geq d(X) - \mathbf{E}[d(\xi_{\tau^n}) | \xi_0 = X] \\ &\geq d(X) - \mathbf{E}[d(\xi_\tau) | \xi_0 = X] - \mathbf{E}[d(\xi_n) | \xi_0 = X] \\ &= d(X) - \mathbf{E}[d(\xi_n) | \xi_0 = X], \quad \forall n \in \mathbb{Z}^+. \end{aligned}$$

Note that $\tau^n \uparrow \tau$ ($n \rightarrow \infty$). By the Monotone Convergence Theorem and (14), it follows that

$$\mathbf{E}[\tau | \xi_0 = X] \geq d(X)/a_2, \quad X \in S^N \setminus C^*.$$

In addition, it is easy to know that $\mathbf{E}[\tau | \xi_0 = X] = 0$, for $X \in C^*$, from the definition of τ .

This completes the proof. \square

From the proof of Theorem 6, we can get

PROPOSITION 1. *If there exists some set $C \subset S^N$ such that test function d satisfies*

$$-a_2 + b_2 I_C(X) \leq \mathbf{E}[d(\xi_1) - d(\xi_0) | \xi_0 = X] \leq -a_1 + b_1 I_C(X) \quad (\mathbf{C3})$$

for any $X \in S^N$ and some constants $b_1 \geq a_1 > 0, a_2 > 0$ and $b_2 < \infty$. Then

$$P(\Omega \setminus Q_k(C)) \rightarrow a_1/b_1, \quad k \rightarrow \infty,$$

where $Q_k(C) = \bigcup_{X \in S^N \setminus C} \{\omega : \xi_k = X\}$.

Remark 4 We can use a result in [21] to explain condition **C1**. According to [21], if $\mathbf{E}[d(\xi_1) - d(\xi_0) | \xi_0 = X] \geq 0$ for $X \in S^N \setminus C^*$, then the mean first hitting times, $\mathbf{E}[\tau | \xi_0 = X]$, are infinite for $X \in S^N \setminus C^*$. Hence, condition **C1** is necessary for upper bound. Condition **C2** says that if EA reaches the optimal population set at the n th step, then the next step, the $n+1$ step, EA still remains at the optimal set; moreover, in order to get lower bound, the one-step drift must be bounded. Hence, **C2** is reasonable.

Remark 5 Proposition 1 tells us that under the condition **(C3)**, the probability which ξ_k reaches the set C tends to a fixed constant a_1/b_1 as the number of generations $k \rightarrow \infty$. Note that Proposition 1 does not imply the convergence of EAs under the sense of probability if $a_1 \neq b_1$.

In fact, We can also adopt the other method to get the upper and lower bounds under the condition **(C2)**, in which the upper bound can be obtained similar to Theorem 5 and the lower bound is analyzed as follows. As though our analysis is somewhat similar to He & Yao's [11–13]. But, strictly speaking, people hardly have a good choice except using the Dynkin's Formula if they just apply the drift properties to estimate the mean first hitting time of EAs. Otherwise, the analytic process will not be rigorous.

By using the method different from Theorem 6, we estimate the lower bound as follows.

For any $X \in S^N$, let

$$A_k(X) = \{\omega : \xi_k = X\}$$

and

$$T_k = \sum_{X \in C^*} A_k(X) = \Omega \setminus Q_k(C^*).$$

It is easy to check that the condition **(C2)** implies that

$$\mathbf{E}[d(\xi_k) - d(\xi_{k+1}) | \xi_k](\omega) = 0, \quad \omega \in T_k, \quad (15)$$

and

$$a_1 \leq \mathbf{E}[d(\xi_k) - d(\xi_{k+1}) | \xi_k](\omega) \leq a_2, \quad \omega \in \Omega \setminus T_k. \quad (16)$$

By Markov property, we have

$$\begin{aligned} &\mathbf{E}[d(\xi_k) | \xi_0](\omega) \\ &= \mathbf{E}[\mathbf{E}[d(\xi_{k-1}) + (d(\xi_k) - d(\xi_{k-1})) | \xi_{k-1}] | \xi_0](\omega) \\ &= \sum_{X \in S^N} \frac{\int_{A_0(X)} \mathbf{E}[d(\xi_{k-1}) + (d(\xi_k) - d(\xi_{k-1})) | \xi_{k-1}] dP}{P(A_0(X))} \\ &\quad \cdot I_{A_0(X)}(\omega). \end{aligned}$$

Hence, for any $X \notin C^*$, we can check that

$$\begin{aligned} &\mathbf{E}[d(\xi_k) | \xi_0 = X] \\ &= \frac{\int_{A_0(X)} \mathbf{E}[d(\xi_{k-1}) + (d(\xi_k) - d(\xi_{k-1})) | \xi_{k-1}] dP}{P(A_0(X))} \\ &= \frac{\int_{A_0(X) \cap T_{k-1}} \mathbf{E}[d(\xi_{k-1}) + (d(\xi_k) - d(\xi_{k-1})) | \xi_{k-1}] dP}{P(A_0(X))} \\ &\quad + \frac{\int_{A_0(X) \cap Q_{k-1}} \mathbf{E}[d(\xi_{k-1}) + (d(\xi_k) - d(\xi_{k-1})) | \xi_{k-1}] dP}{P(A_0(X))}. \end{aligned}$$

By (15) and (16), for any $X \notin C^*$, we have

$$\begin{aligned} &\frac{\int_{A_0(X) \cap Q_{k-1}} (d(\xi_{k-1}) - a_2) dP}{P(A_0(X))} \\ &\leq \mathbf{E}[d(\xi_k) | \xi_0 = X] \\ &\leq \frac{\int_{A_0(X) \cap Q_{k-1}} (d(\xi_{k-1}) - a_1) dP}{P(A_0(X))}. \end{aligned}$$

But, for any $X \notin C^*$ and the constant $c > 0$

$$\begin{aligned}
& \mathbf{E}[d(\xi_{k-1}) - c | \xi_0 = X] \\
&= \frac{\int_{A_0(X)} (d(\xi_{k-1}) - c) dP}{P(A_0(X))} \\
&= \frac{\int_{A_0(X) \cap T_{k-1}} (d(\xi_{k-1}) - c) dP}{P(A_0(X))} \\
&+ \frac{\int_{A_0(X) \cap Q_{k-1}} (d(\xi_{k-1}) - c) dP}{P(A_0(X))} \\
&= \frac{-c}{P(A_0(X))} \cdot P(A_0(X) \cap T_{k-1}) \\
&+ \frac{\int_{A_0(X) \cap Q_{k-1}} (d(\xi_{k-1}) - c) dP}{P(A_0(X))}.
\end{aligned}$$

Therefore, for any $X \notin C^*$, we have

$$\begin{aligned}
& \mathbf{E}[d(\xi_k) | \xi_0 = X] \\
&\geq \mathbf{E}[d(\xi_{k-1}) - a_2 | \xi_0 = X] + \frac{a_2}{P(A_0(X))} \cdot P(A_0(j) \cap T_{k-1}) \\
&= \mathbf{E}[d(\xi_{k-1}) | \xi_0 = X] - a_2 + \frac{a_2}{P(A_0(X))} \cdot P(A_0(X) \cap T_{k-1}) \\
&= \mathbf{E}[d(\xi_{k-1}) | \xi_0 = X] \\
&- \frac{a_2}{P(A_0(X))} [P(A_0(X)) - P(A_0(X) \cap T_{k-1})] \\
&= \mathbf{E}[d(\xi_{k-1}) | \xi_0 = X] - \frac{a_2}{P(A_0(X))} [P\{\omega : \xi_0 = X, \tau \geq k\}] \\
&= \mathbf{E}[d(\xi_{k-1}) | \xi_0 = X] - a_2 \cdot \mathbf{E}[I_{\tau \geq k}(\omega) | \xi_0 = X].
\end{aligned}$$

By induction on k , for any $X \notin C^*$, we have

$$\mathbf{E}[d(\xi_k) | \xi_0 = X] \geq d(X) - a_2 \sum_{n=1}^k \mathbf{E}[I_{\tau \geq n}(\omega) | \xi_0 = X]. \quad (17)$$

Since $\mathbf{E}[\tau | \xi_0 = X] < \infty$ by Theorem 5, then $\mathbf{E}[\tau | \xi_0 = X] = \sum_{n=1}^{\infty} \mathbf{E}[I_{\tau \geq n}(\omega) | \xi_0 = X]$. Combining with (14), for any $X \notin C^*$, we get from (17)

$$a_2 \cdot \mathbf{E}[\tau | \xi_0 = X] \geq d(x).$$

On the other hand, it is easy to know that

$$\mathbf{E}[\tau | \xi_0 = X] = 0, \forall X \in C^*.$$

Hence

$$\mathbf{E}[\tau | \xi_0 = X] \begin{cases} \geq d(X)/a_2, & X \in S^N \setminus C^* \\ = 0, & X \in C^* \end{cases}$$

This completes the proof. \square

In the above analysis, we only use the condition **(C2)**. In fact, it is easy to find that the condition **(C2)** implies the condition **(C1)** when we take b as a in the condition **(C1)**. In order to obtain the lower bound of $\mathbf{E}[\tau | \xi_0 = X]$, we do not only need to limit the drift speed (i.e. the left inequality of the condition **(C2)**) but also keep up the certain positive drift (i.e. the right inequality of the condition **(C2)**) such that the convergence can be ensured.

Remark 6 In specialty, we claim that the results obtained in section 4 are true for EAs to solve optimization problems based on general search space S provided the test function $d(\cdot)$ defined on S^N is imposed little restriction, i.e. $\sup_{X \in S^N} d(X) < \infty$, where a real-valued objective function f is only claimed bounded from above for the maximization problem or bounded from below for the minimization problem.

5. CONCLUSIONS AND DISCUSSIONS

This paper has given some general results about the time complexity of EAs, which have great importance in theory and practice. More important, some analytic techniques and methods used in this paper, which may supply the researchers in the area of EA-theory the uses of references, are foundational and even essential for investigating the time complexity problems in EAs.

This paper has shown that Markov chain is a convenient model which can be used to describe the EAs and that drift analysis is a practical means which is useful to estimate the computation time of EAs. In the meantime, it has also implied that some more profound results about the computation time of EAs can be derived by using the drift analysis and other tools in stochastic process theory.

As mentioned in [15], drift analysis reduces the behavior of EAs in a higher dimensional population space S^N into a super-martingale on the one-dimensional space by the introduction of a distance function for the population space. This makes the theoretical analysis much simpler than analyzing the original Markov chain associated with the EAs. The key point in applying drift analysis is to define a good test function on the population space S^N .

It can be seen from this paper that the application of Dynkin's Formula is a key technique in order to obtain a rigorous theoretical analysis, which has not been used in the previously related works.

The application of drift analysis to studying computation time and time complexity of EAs is still at its early days. A number of problems are still open: How to describe the relation between the time complexity and the space complexity (which is related to both problem size and population size.)? In a given kind of problems, how to analyze the time complexity of different EAs which are constructed by different algorithmic component parts? How to show the time complexity of a given EA which is used in the different kind of problems? What is the relation between the time complexity and the precision of ε -optimal solution? How to classify definitely both the EA-hard problems and the EA-easy problems? Why is it important to research further the computational dynamics properties and explanations associated with the time complexity of EAs? More essential, whether there is a kind of EAs which can be used to solve (or under the sense of ε -optimum) a NP-problem within the polynomial time theoretically or not? All these problems are well worth being investigated in the field of the time complexity of EAs in the future.

6. 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 Youth Scholars in Wuhan City (Grant no. 20025001002) and the Open

7. 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] T. Bäck. The interaction of mutation rate, selection and self-adaption within a genetic algorithm. In *PPSN-II Conference Proceedings*, pages 85-94. North-Holland, Amsterdam, 1992.
- [4] H. Mühlenbein. How genetic algorithms really works I: Mutation and hill-climbing. In *PPSN-II Conference Proceedings*, pages 15-25. North-Holland, Amsterdam, 1992.
- [5] G. Rudolph. *Convergence Properties of Evolutionary Algorithms*. Ph.D. Thesis, Verlag Dr. Kovač, Hamburg, 1997.
- [6] J. Garnier, L. Kallel, and M. Schoenauer. Rigorous hitting times for binary mutations. *Evolutionary Computation*, 7(2):167-203, 1999.
- [7] S. Droste, T. Jansen, and I. Wegener. A rigorous complexity analysis of the $(1 + 1)$ evolutionary algorithm for linear functions with Boolean inputs. *Evolutionary Computation*, 6(2):185-196, 1998.
- [8] S. Droste, T. Jansen, and I. Wegener. On the analysis of the $(1 + 1)$ evolutionary algorithms. *Theoretical Computer Science*, 276(1-2):51-81, 2002.
- [9] G. Rudolph. How mutation and selection solve long path problems in polynomial expected time. *Evolutionary Computation*, 4(2):194-205, 1996.
- [10] J. Garnier and L. Kallel. Statistical distribution of the convergence time of evolutionary algorithms for long path problems. *IEEE Trans.on Evolutionary Computation*, 4(1):16-30, 2000.
- [11] J. He and X. Yao. Drift analysis and average time complexity of evolutionary algorithms. *Artificial Intelligence*, 127(1):57-85, 2001.
- [12] J. He and X. Yao. From an individual to a population: An analysis of the first hitting time of population-based evolutionary algorithms. *IEEE Trans. on Evolutionary computation*, 6(5):495-511, 2002.
- [13] J. He and X. Yao. Towards an analytic framework for analyzing the computation time of evolutionary algorithms. *Artificial Intelligence*, 145(1-2):59-97, 2003.
- [14] J. He and X. Yao. An analysis of evolutionary algorithms for finding approximation solutions to hard optimisation problems. In *CEC'03 Conference Proceedings*, pages 2004-2010. IEEE Press, 2003.
- [15] J. He and X. Yao. A study of drift analysis for estimating computation time of evolutionary algorithms. *Natural Computing*, 3:21-35, 2004.
- [16] J. He and X. Yao. Time complexity analysis of an evolutionary algorithm for finding nearly maximum cardinality matching. *Journal of Computer science & Technology*, 19(4):450-458, 2004.
- [17] E. A. Nix and M. D. Vose. Modeling genetic algorithms with Markov chain. *Ann. Math. Artificial Intelligence*, 5:79-88, 1992.
- [18] J. Suzuki. A Markov chain analysis on simple genetic algorithms. *IEEE Trans. on Systems Man & Cybernetics*, 25:655-659, 1995.
- [19] M. D. Vose. *The Simple Genetic Algorithms: Foundations and Theory*. MIT Press, Cambridge, 1999.
- [20] G. H. Sasaki and B. Hajek. The time complexity of maximum matching by simulated annealing. *J ACM*, 35(2):387-403, 1988.
- [21] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*, (3rd ed.). Springer-Verlag, New York, 1996.
- [22] H. -G. Beyer. *The Theory of Evolutionary Strategies*. Springer-Verlag, Berlin, 2001.