A New Approach to Estimating the Expected First Hitting Time of Evolutionary Algorithms

Yang Yu and **Zhi-Hua Zhou**
National Laboratory for Novel Software Technology
Nanjing University, Nanjing 210093, China
{yuy, zhouzh}@lamda.nju.edu.cn

Abstract

The *expected first hitting time* is an important issue in theoretical analyses of evolutionary algorithms since it implies the average computational time complexity. In this paper, by exploiting the relationship between the convergence rate and the expected first hitting time, a new approach to estimating the expected first hitting time is proposed. This approach is then applied to four evolutionary algorithms which involve operators of mutation, mutation with population, mutation with recombination, and time-variant mutation, respectively. The results show that the proposed approach is helpful for analyzing a broad range of evolutionary algorithms.

Introduction

Evolutionary algorithms (EAs) are a kind of nonlinear global optimization technique, inspired by the natural evolution process. Despite many different implementations (Bäck 1996), e.g. Genetic Algorithm, Genetic Programming and Evolutionary Strategies, traditional evolutionary algorithms can be summarized below by four steps:

1. Generate an initial population of random solutions;
2. Reproduce new solutions based on the current population;
3. Remove relatively bad solutions in the population;
4. Repeat from Step 2 until the stop criterion is met.

EAs solve problems in quite a straightforward way, need not, e.g. differentiable function or continuous function or inverse of matrix. Thus, EAs have been widely applied to diverse areas (Freitas 2003). However, despite the remarkable successes achieved by EAs, they lack a solid theoretical foundation. Recently, many approaches for theoretically analyzing EAs have been proposed (Beyer, Schwefel, & Wegener 2002; He & Yao 2001; 2003).

The *first hitting time* (FHT) of EAs is the time that EAs find the optimal solution for the first time, and the *expected first hitting time* (expected FHT) is the average time that EAs require to find the optimal solution, which implies the average computational time complexity of EAs. Thus, the expected FHT is one of the most important theoretical issues of EAs.

Many works have been devoted to the analysis of simple EAs, say (1+1)-EA (Rudolph 1997; Droste, Jansen, & Wegener 1998), for specific problems (van Nimwegen, Crutchfield, & Mitchell 1999; Garnier, Kallel, & Schoenauer 1999). For this a survey can be found in (Beyer, Schwefel, & Wegener 2002). Recently, significant advances have been made by He and Yao (2001; 2004), who developed an general approach to analyzing a wide class of EAs based on the *drift analysis* (Hajek 1982). However, this approach requires a distance function which does not naturally exist in EAs, yet it is not known how to design such a distance function in practice.

He and Yao (2003) developed another framework to analyze the complexity and to compare two EAs based on the analytical form of the FHT. However, since the analytical form was derived from homogeneous Markov chain models, only EAs with stationary reproduction operators can be analyzed, although EAs with time-variant operators or adaptive operators are very popular and powerful (Eiben, Hinterding, & Michalewicz 1999).

This paper shows that the expected FHT can be derived from the *convergence rate* of EAs, which is another important theoretical issue of EAs. A pair of general upper/lower bounds of the expected FHT are then achieved. Based on a non-homogeneous Markov chain model, the proposed approach can be applied to analyze a wide class of EAs for a wide class of problems. Moreover, the proposed approach needs no distance functions, and is suitable for analyzing dynamic EAs. To show the proposed approach can be applied to a board range of EAs, the expected FHT of four EAs on a hard problem (solved in exponential time) are analyzed for illustration, which verifies the superiority of the proposed approach since previous approaches were applied to easy problems (solved in polynomial time) (Rudolph 1997; Droste, Jansen, & Wegener 1998; He & Yao 2004). In particular, an EA with time-variant operator is theoretically analyzed for the expected FHT, which is, to the best of our knowledge, for the first time.

The rest of this paper starts with a section which briefly reviews some related works and introduces how to model EAs using Markov chain. Then, the main results of the paper are presented and four EAs are analyzed for illustration, which is followed by the conclusion.

Modelling EAs using Markov Chain

EAs evolve solutions from generation to generation. Each generation stochastically depends on the very pervious one, except the initial generation which is randomly generated. This essential can be modeled by the Markov chain model naturally (Nix & Vose 1992; Suzuki 1995; Rudolph 1998; He & Kang 1999; He & Yao 2001; 2003).

Since combinatorial optimization problems are among the most common problems in practice, in this paper, using EAs to tackle combinatorial optimization problems is studied, i.e. the solutions can be represented by a sequence of symbols. To model this kind of EAs, a Markov chain with discrete state space is constructed. The key of constructing such a Markov chain is to map the populations of EAs to the states of the Markov chain. A popular mapping (Suzuki 1995; He & Kang 1999; He & Yu 2001) lets a state of Markov chain correspond to a possible population of EAs, which is the most exquisite mapping. Suppose an EA encodes a solution into a vector with length L, each component of the vector is drawn from an alphabet set $\mathcal{B}$, and each population contain M solutions. Let X be the population space, there are $|X| = \binom{M+L^{|\mathcal{B}|}-1}{M}$ number of different possible populations (Suzuki 1995). A Markov chain modeling the EA is constructed by taking X as the state space, i.e. building a chain $\{\xi_t\}_{t=0}^{+\infty}$ where $\xi_t \in X$.

A population is called the *optimal population* if it contains at least one optimal solution. Let $X^* \in X$ denotes the set of all optimal populations. The goal of EAs is to reach X^* from an initial population. Thus, the process of an EAs seeking X^* can be analyzed by studying the corresponding Markov chain (He & Kang 1999; He & Yu 2001).

First, given a Markov chain $\{\xi_t\}_{t=0}^{+\infty}(\xi_t \in X)$ and a target subspace $X^* \subset X$, let $\mu_t(t = 0, 1, \dots)$ denotes the probability of ξ_t in X^*, that is,

$$\mu_t = \sum\nolimits_{x \in X^*} P(\xi_t = x) \tag{1}$$

Definition 1 (Convergence) *Given a Markov chain $\{\xi_t\}_{t=0}^{+\infty}$ $(\xi_t \in X)$ and a target subspace $X^* \subset X$, $\{\xi_t\}_{t=0}^{+\infty}$ is said to converge to X^* if*

$$\lim_{t \to +\infty} \mu_t = 1 \tag{2}$$

In (He & Yu 2001), *convergence rate* is measured by $1 - \mu_t$ at step t, which is equivalent to the measure used in (Suzuki 1995). Therefore, $1 - \mu_t$ is also used as the measure of convergence rate in this paper.

Definition 2 (Absorbing Markov Chain) *Given a Markov chain $\{\xi_t\}_{t=0}^{+\infty}(\xi_t \in X)$ and a target subspace $X^* \subset X$, $\{\xi_t\}_{t=0}^{+\infty}$ is said to be an absorbing chain, if*

$$P(\xi_{t+1} \notin X^* | \xi_t \in X^*) = 0 \ (\forall t = 0, 1, \dots) \tag{3}$$

An EA can be modeled by an absorbing Markov chain, if it never loses the optimal solution once been found. Many EAs for real problems can be modeled by an absorbing Markov chain. This is because if the optimal solution to the concerned problem can be identified, then an EA will stop when it finds the optimal solution; while if the optimal solution to the concerned problem can not be identified, then an EA will keep the best-so-far solution in each generation.

Definition 3 (Expected first hitting time) *Given a Markov chain $\{\xi_t\}_{t=0}^{+\infty}(\xi_t \in X)$ and a target subspace $X^* \subset X$, let τ be a random variable that denotes the events:*

$$\begin{aligned}
&\tau = 0 : \xi_0 \in X^* \\
&\tau = 1 : \xi_1 \in X^* \wedge \xi_i \notin X^* \ (\forall i = 0) \\
&\tau = 2 : \xi_2 \in X^* \wedge \xi_i \notin X^* \ (\forall i = 0, 1) \\
&\dots
\end{aligned}$$

then the mathematical expectation of τ, $\mathbb{E}[\tau]$, is called the expected first hitting time (expected FHT) of the Markov chain.

Note that this definition of expected FHT is equivalent to those used in (He & Yao 2001; 2003). The expected FHT is the average time that EAs find the optimal solution, which implies the average computational time complexity of EAs.

The Markov chain models the essential of EA process, thus the convergence, convergence rate and expected FHT of EAs can be obtained through analyzing the corresponding Markov chains. So, the convergence, convergence rate and expected FHT of EAs or of corresponding Markov chains won't be distinguished for convenience.

Deriving Expected First Hitting Time from Convergence Rate

The convergence rate has been studied for years (Suzuki 1995; He & Kang 1999) and recently a general bounds of convergence rate have been derived in (He & Yu 2001). Since this paper focuses on using EAs to solve combinatorial optimization problems, a discrete space version of the Theorem 4 in (He & Yu 2001) is proven as below:

Lemma 1 *Given an absorbing Markov chain $\{\xi_t\}_{t=0}^{+\infty}(\xi_t \in X)$ and a target subspace $X^* \subset X$, if two sequences $\{\alpha_t\}_{t=0}^{+\infty}$ and $\{\beta_t\}_{t=0}^{+\infty}$ satisfy*

$$\prod\nolimits_{i=0}^{t-1}(1 - \alpha_t) = 0 \tag{4}$$

and

$$\sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \geq \alpha_t (1 - \mu_t) \tag{5}$$

and

$$\sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \leq \beta_t (1 - \mu_t), \tag{6}$$

then the chain converges to X^* and the converge rate is bounded by

$$(1 - \mu_0) \prod_{i=0}^{t-1} (1 - \alpha_i) \geq 1 - \mu_t \geq (1 - \mu_0) \prod_{i=0}^{t-1} (1 - \beta_i) \tag{7}$$

Proof. According to Eqs.1 and 3, it can be obtained that

$$\mu_t - \mu_{t-1} = \sum_{x \notin X^*} P(\xi_t \in X^* | \xi_{t-1} = x) P(\xi_t = x),$$

and through applying Eqs.5 and 6:

$$\begin{gathered}(1-\mu_{t-1})\alpha_{t-1} \le \mu_t - \mu_{t-1} \le (1-\mu_{t-1})\beta_{t-1} \\ (1-\mu_{t-1})(1-\alpha_{t-1}) \ge 1-\mu_t \ge (1-\mu_{t-1})(1-\beta_{t-1}) \\ (1-\mu_0)\prod_{i=0}^{t-1}(1-\alpha_i) \ge 1-\mu_t \ge (1-\mu_0)\prod_{i=0}^{t-1}(1-\beta_i)\end{gathered}$$

□

Lemma 1 implies that as long as the probability that an EA 'jumps' into the set of optimal solutions can be estimated for each step, the bounds of its convergence rate can be obtained. The only condition is that the EA can be modeled by an absorbing Markov chain, i.e. satisfying Eq.3. As mentioned before, many EAs for real problems meet this requirement.

In Definition 3, the expected FHT is the mathematical expectation of the random variable τ. Meanwhile, the probability distribution of τ is the probability that an optimal solution is found before step t ($t = 0, 1, \ldots$). Thus, as long as the EA can be modeled by an absorbing Markov chain, it holds that

$$\begin{aligned}\mu_{t+1} - \mu_t &= \sum_{x \in X^*} P(\xi_{t+1} = x) - \sum_{x \in X^*} P(\xi_t = x) \\ &= P(\tau = t+1).\end{aligned}$$

This implies that the probability distribution of τ is equal to μ_t, which is 1 minus the convergence rate. Therefore, the convergence rate and the expected FHT is just the two sides of a coin.

Note that the bounds of the probability distribution and bounds of the expectation of the same random variable have a relationship shown in Lemma 2.

Lemma 2 *Suppose U and V are two discrete random variable on nonnegative integer, let $D_u(\cdot)$ and $D_v(\cdot)$ be their distribution functions respectively. If the distributions satisfy*

$$D_u(t) \ge D_v(t) \ (\forall t = 0, 1, \ldots)$$

then the expectations of the random variables satisfy

$$\mathbb{E}[U] \le \mathbb{E}[V] \tag{8}$$

Proof. Since D_u is the distribution of U,

$$\begin{aligned}\mathbb{E}[U] &= 0 \cdot D_u(0) + \sum_{t=1}^{+\infty} t\big(D_u(t) - D_u(t-1)\big) \\ &= \sum_{i=1}^{+\infty}\sum_{t=i}^{+\infty}\big(D_u(t) - D_u(t-1)\big) \\ &= \sum_{i=0}^{+\infty}\big(\lim_{t\to+\infty} D_u(t) - D_u(i)\big) = \sum_{i=0}^{+\infty}\big(1 - D_u(i)\big),\end{aligned}$$

and same for V. Thus,

$$\begin{aligned}\mathbb{E}[U] - \mathbb{E}[V] &= \sum_{i=0}^{+\infty}\big(1 - D_u(i)\big) - \sum_{i=0}^{+\infty}\big(1 - D_v(i)\big) \\ &= \sum_{i=0}^{+\infty}\big(D_v(i) - D_u(i)\big) \le 0.\end{aligned}$$

□

Since 1 minus the convergence rate is the probability distribution of τ, and the expected FHT is the expectation of τ, Lemma 2 reveals that the lower/upper bounds of the expected FHT can be derived from the lower/upper bounds of the convergence rate.

Thus, based on Lemmas 1 and 2, a pair of general bounds of the expected FHT in Theorem 1 can be obtained.

Theorem 1 *Given an absorbing Markov chain $\{\xi_t\}_{t=0}^{+\infty}(\xi_t \in X)$ and a target subspace $X^* \subset X$, if two sequences $\{\alpha_t\}_{t=0}^{+\infty}$ and $\{\beta_t\}_{t=0}^{+\infty}$ satisfy*

$$\prod\nolimits_{i=0}^{t-1}(1-\alpha_t) = 0 \tag{9}$$

and

$$\sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \ge \alpha_t (1 - \mu_t) \tag{10}$$

and

$$\sum_{x \notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \le \beta_t (1 - \mu_t), \tag{11}$$

then the chain converges and, starting from non-optimal solutions, the expected FHT is bounded by

$$\mathbb{E}[\tau] \le \alpha_0 + \sum_{t=2}^{+\infty} t\alpha_{t-1}\prod_{i=0}^{t-2}(1-\alpha_i) \tag{12}$$

and

$$\mathbb{E}[\tau] \ge \beta_0 + \sum_{t=2}^{+\infty} t\beta_{t-1}\prod_{i=0}^{t-2}(1-\beta_i) \tag{13}$$

Proof. Applying Lemma 1 with Eqs.10 and 11, get

$$1 - \mu_t \le (1-\mu_0)\prod\nolimits_{i=0}^{t-1}(1-\alpha_i).$$

Considering that μ_t expresses the distribution of τ, i.e. $\mu_t = D_\tau(t)$, the lower bound of $D_\tau(t)$ can be obtained:

$$D_\tau(t) \ge \begin{cases}\mu_0 & t = 0 \\ 1 - (1-\mu_0)\prod_{i=0}^{t-1}(1-\alpha_i) & t = 1, 2, \ldots\end{cases}$$

Imagine a virtual random variable η whose distribution equals to the lower bound of D_τ. The expectation of η is:

$$\begin{aligned}\mathbb{E}[\eta] &= 0 \cdot \mu_0 + 1 \cdot \Big[1 - (1-\alpha_0)(1-\mu_0) - \mu_0\Big] \\ &+ \sum_{t=2}^{+\infty} t \cdot \Big[(1-\mu_0)\prod_{i=0}^{t-2}(1-\alpha_i) - (1-\mu_0)\prod_{i=0}^{t-1}(1-\alpha_i)\Big] \\ &= \Big[\alpha_0 + \sum_{t=2}^{+\infty} t\alpha_{t-1}\prod_{i=0}^{t-2}(1-\alpha_i)\Big](1-\mu_0)\end{aligned}$$

Since $D_\tau(t) \geq D_\eta(t)$, according to Lemma 2, $\mathbb{E}[\tau] \leq \mathbb{E}[\eta]$. Thus, the upper bound of the expected FHT:

$$\mathbb{E}[\tau] \leq \Big[\alpha_0 + \sum_{t=2}^{+\infty} t\alpha_{t-1} \prod_{i=0}^{t-2}(1-\alpha_i)\Big](1-\mu_0)$$

Note that the EA is assumed to start from non-optimal solutions, i.e. $\mu_0 = 0$.

The lower bound of the expected FHT is derived similarly. □

The bounds of expected FHT, i.e. Eqs. 12 and 13, have an intuitive explanation. The part $\alpha_{t-1}\prod_{i=0}^{t-2}(1-\alpha_i)$ (or replace α_t by β_t) indicates the probability of the event that the EA finds the optimal solution at the t-th step, but does not find at $0, 1, \ldots, t-1$ step.

Theorem 1 discloses that as long as the probability that an EA 'jumps' into the optimal population at each step is estimated, the bounds of the expected FHT are obtained. The more exact the probability is estimated, the more tight the bounds are.

Case Study

Four EAs solving the *Subset Sum Problem* are analyzed in this section, using the results obtained in the previous section. The Subset Sum Problem is described as follows, which has been proven to be an *NP-Complete* problem (Baase & van Gelder 2000).

Definition 4 (Subset Sum Problem) *Given a set of n positive integers $W = \{w_i\}_{i=1}^n$ and a constant c, among subsets of W with elements summed at most c, find the subset $W^* \subseteq W$ whose elements are summed to be the largest.*

Specific subset sum problems can be obtained by specifying W and c, such as

Problem 1: $w_i \geq 2$ $(\forall i = 1, 2, \ldots, n-1)$, $w_n = (\sum_{i=1}^{n-1} w_i) - 1$ and $c = w_n$.

The optimal solution of Problem 1 is $x^* = (000\ldots01)$. Problem 1 was used in (He & Yao 2001) and shown to be solved by an EA in exponential time. In the following, it will be proven that Problem 1 is also solved by other four EAs in exponential time, based on the results of the previous section.

The EAs are configured as below, where the *Reproduction* will be implemented by concrete operators later.

- *Encoding:* Each solution is encoded by a string with n binary bits, where the i-th bit is 1 if w_i is included and 0 otherwise.
- *Initial:* Randomly generate a population of M solutions encoded by binary strings.
- *Reproduction:* Generate M new solutions based on the current population.
- *Selection:* Select the best M solutions from the current population and the reproduced solutions to form the population of the next generation.
- *Fitness:* The fitness of a solution $x = (x_1x_2\ldots x_n)$ is defined as

  $$fitness(x) = c - \theta \sum\nolimits_{i=1}^{n} w_i x_i$$

  where $\theta = 1$ when x is a feasible solution, i.e. $\sum_{i=1}^n w_i x_i \leq c$, and $\theta = 0$ otherwise.
- *Stop Criterion:* If the lowest fitness in population is zero, stop and output the solution with zero fitness.

The first EA using *Mutation 1* to implement the Reproduction, and the population size is 1.

Mutation 1: A new solution is generated by randomly selecting a solution in population and independently flipping each component of the solution with probability $p_m \in (0, 0.5]$.

For this EA, Corollary 1 gives the bounds of its expected FHT.

Corollary 1 *Solving Problem 1 by the EA with Reproduction implemented by Mutation 1 and population size 1, assuming starting from non-optimal populations, the expected FHT is bounded by*

$$\Big(\frac{1}{p_m}\Big)^n \geq \mathbb{E}[\tau] \geq \frac{1}{p_m}\Big(\frac{1}{1-p_m}\Big)^{n-1} \tag{14}$$

where n is the problem size.

Proof. Applying Mutation 1, the probability that a solution $x^{(1)}$ mutates to another solution $x^{(2)}$ is $p_m^k(1-p_m)^{n-k}$, where $x^{(1)}$ and $x^{(2)}$ are different in k bits.

Considering Problem 1 where the optimal solution is $x^* = (000\ldots01)$, it is easy to find that, since $p_m \in (0, 0.5]$, the minimum probability that any non-optimal solution mutates to the optimal solution is p_m^n, and the maximum is $p_m(1-p_m)^{n-1}$.

Population size is 1 means a population is equal to a solution. Then, for $t = 0, 1, \ldots$

$$p_m(1-p_m)^{n-1} \geq P(\xi_{t+1} \in X^* | \xi_t = x) \geq p_m^n.$$

Thus,

$$\begin{aligned} p_m(1-p_m)^{n-1} \sum\nolimits_{x\notin X^*} P(\xi_t = x) \geq \\ \sum\nolimits_{x\notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \\ \geq p_m^n \sum\nolimits_{x\notin X^*} P(\xi_t = x). \end{aligned}$$

Applying Eq.1, it can be obtained that

$$\begin{aligned} p_m(1-p_m)^{n-1}(1-\mu_t) \geq \\ \sum\nolimits_{x\notin X^*} P(\xi_{t+1} \in X^* | \xi_t = x) P(\xi_t = x) \\ \geq p_m^n(1-\mu_t). \end{aligned}$$

Therefore, $\alpha_t = p_m^n$ and $\beta_t = p_m(1-p_m)^{n-1}$ satisfy Eqs.10 and 11, which makes Eq.12 holds.

From Eq.12, let $a = \alpha_t = p_m^n$, the upper bound is:

$$\mathbb{E}(\tau) \leq \alpha_0 + \sum_{t=2}^{+\infty} t\alpha_{t-1} \prod_{i=0}^{t-2}(1-\alpha_i)$$

$$= a + \sum_{t=2}^{+\infty} ta \prod_{i=0}^{t-2}(1-a) = \sum_{t=1}^{+\infty} ta(1-a)^{t-1}$$
$$= a\Big(\sum_{t=0}^{+\infty} t(1-a)^t + \sum_{t=0}^{+\infty}(1-a)^t\Big) = a\Big(\frac{1-a}{a^2} + \frac{1}{a}\Big)$$
$$= a^{-1} = p_m^{-n}$$

The lower bound can be derived similarly. □

Note that $1/(1-p_m) > 1$ in Corollary 1, the lower bound of the expected FHT shows that Problem 1 is solved by the EA with Mutation 1 in exponential time.

The second EA also uses the Mutation 1, but the population size can be larger than 1. Corollary 2 gives the bounds of its expected FHT.

Corollary 2 *Solving Problem 1 by the EA with Reproduction implemented by Mutation 1 and population size M, assuming starting from non-optimal populations, the expected FHT is bounded by*

$$\frac{1}{M}\Big(\frac{1}{p_m}\Big)^n \geq \mathbb{E}[\tau] \geq \frac{1}{M}\frac{1}{p_m}\Big(\frac{1}{1-p_m}\Big)^{n-1} \tag{15}$$

where n is the problem size that is large enough.

Proof. From the proof of Corollary 1, it is known that the minimum probability that any non-optimal solution mutates to the optimal solution is p_m^n, and the maximum is $p_m(1-p_m)^{n-1}$.
When a population contains all the solutions that have the minimum probability to mutate to an optimal solution, it is the population that has the minimum probability to mutate to an optimal population. So, the probability that the population mutates to the optimal population is that for at least one solution mutates to the optimal solution, i.e. $1-(1-p_m^n)^M$. For the same reason, the maximum probability that a population mutates to an optimal population is $1-(1-p_m(1-p_m)^{n-1})^M$.
The sequences $\alpha_t = 1-(1-p_m^n)^M$ and $\beta_t = 1-(1-p_m(1-p_m)^{n-1})^M$ satisfies Eqs.10 and 11. Then, after the same derivation of the proof of Corollary 1, it is easy to get

$$\frac{1}{1-(1-p_m^n)^M} \geq \mathbb{E}[\tau] \geq \frac{1}{1-(1-p_m(1-p_m)^{n-1})^M}$$

Note that when n is large enough, both p_m^n and $p_m(1-p_m)^{n-1}$ approach to 0. Applying the asymptotic equation $(1-x)^n \sim 1-nx$ when $x \to 0$, bounds are

$$\frac{1}{M}\frac{1}{p_m^n} \geq \mathbb{E}[\tau] \geq \frac{1}{M}\frac{1}{p_m(1-p_m)^{n-1}}$$

□

From Corollary 2, it can be observed that increasing the population can help reduce the FHT. However, as long as the population size is polynomial of n, which is common in practice, the expected FHT is still exponential of n.

The third EA using Reproduction implemented by Mutation 1 and Recombination, and the population size is 2.

Recombination: Randomly select two solutions in population, exchange the first σ bits to generate two new solutions, where σ is randomly picked from $\{1, 2, \ldots, n-1\}$.

For this EA, Corollary 3 says the lower bounds of its expected FHT is still exponential in n.

Corollary 3 *Solving Problem 1 by the EA with Reproduction implemented by Mutation 1 and Recombination and population size 2, assuming starting from non-optimal population, the expected FHT is lower bounded by*

$$\mathbb{E}[\tau] \geq \Theta\big((1-p_m)^{-n}\big) \tag{16}$$

where n is the problem size.

Proof. According to Corollary 1 in (He & Yu 2001), the set

$$X_c = \{x | x \in X \wedge P(\xi_{t+1} \in X^* | \xi_t = x) > 0\}$$

is to be found. Suppose the two solutions in the population are $x^{(1)} = (x_1^{(1)} x_2^{(1)} \ldots x_n^{(1)})$ and $x^{(2)} = (x_1^{(2)} x_2^{(2)} \ldots x_n^{(2)})$, and the optimal solution is $x^* = (x_1^* x_2^* \ldots x_n^*)$. The two solution can be exchanged by the first σ bits to be the optimal solution, only when

$$x_i^{(1)} = x_i^*,\ x_j^{(2)} = x_j^*\ (\forall i = 1 \ldots \sigma, \forall j = \sigma+1 \ldots n).$$

So, the probability of this kind of population occurs, i.e. $P(X_c)$, at the initialization is

$$\sum_{\sigma=1}^{n-1}\Big(\sum_{i=\sigma}^{n-1}\frac{1}{2^i}\Big)\Big(\sum_{j=1}^{\sigma}\frac{1}{2^{n-j}}\Big) < \frac{3(n-1)}{2^{n+1}}.$$

Note that for Problem 1, the optimal solution is (000…01). According to the fitness function, given two solutions $x' = (x_1' x_2' \ldots x_n')$ and $x'' = (x_1'' x_2'' \ldots x_n'')$, x' will have lower fitness than x'' or even infeasible if

$$x_j' = x_j^*,\ x_j'' \neq x_j^*, \text{and } x_i' = x_i'' (\forall i = 1, 2 \ldots n,\ i \neq j)$$

This means $P(X_c)$ will reduce as step increases.

Simply treat the probability that a population is recombined to be the optimal as 1. According to the Corollary 1 in (He & Yu 2001), β_t in Theorem 1 can be calculated:

$$\begin{aligned}\beta_t &= 3(n-1)2^{-n-1} + 2p_m(1-p_m)^{n-1} \\ &\quad - 3(n-1)2^{-n-1}p_m(1-p_m)^{n-1} \\ &\in \Theta((1-p_m)^n)\end{aligned}$$

With a similar proof of Corollary 1, the lower bound $\Theta\big((1-p_m)^{-n}\big)$ can be obtained. □

It can be found through the proof that the Recombination helps little firstly because there is few population where Recombination can be useful.

The fourth EA implements Reproduction by a time-variant mutation, *Mutation 2*.

Mutation 2: A new solution is generated by randomly selecting a solution in population and independently flipping each component of the solution with probability $(0.5-d)e^{-t}+d$, where $d \in (0, 0.5)$ and $t = 0, 1, \ldots$ is the index of steps.

At the beginning the probability of mutation is 0.5, the probability reduces as time going, and ultimately the probability is d. For this EA, Corollary 4 shows it still solve the problem in exponential time, since $(1-d) \in (0, 0.5)$.

Corollary 4 *Solving Problem 1 by the EA with Reproduction implemented by Mutation 2 and population size 1, assuming starting from non-optimal population, the expected FHT is lower bounded by*

$$\mathbb{E}[\tau] \geq 2(1-d)^{-n+1} \tag{17}$$

where n is the problem size.

Proof. Similar to the proof of Corollary 1, the maximum probability that a non-optimal solution mutates to be optimal at step t is $[(0.5-d)e^{-t}+d]\big(1-(0.5-d)e^{-t}-d\big)^{n-1}$, which is smaller than $0.5(1-d)^{n-1}$.

Let $\beta_t = 0.5(1-d)^{n-1}$, the sequence $\{\beta_t\}_{t=0}^{+\infty}$ satisfies Eq.11. Then, from a similar derivation of proof of Corollary 1, the lower bound $2(1-d)^{-n+1}$ can be obtained. □

Conclusion

In this paper, a bridge between the convergence rate and the expected first hitting time (expected FHT) of evolutionary algorithms (EAs) is established, which are two of the most important theoretical issues of EAs. Based on the bridge, this paper proposes a new approach of analyzing the expected FHT of EAs. The proposed approach utilizes non-homogeneous Markov chain, thus it is suitable for analyzing a broad range of EAs. To illustrate the helpfulness of the proposed approach, a problem is proven to be hard for four EAs for instance, which involves operators of mutation, mutation with population, mutation with recombination and time-variant mutation, respectively.

It is noteworthy that the proposed method can be used to analyze EAs with dynamic operators. Although EAs using dynamic operators such as time-variant operators are popular, they can hardly be analyzed before this work. A longer version of the paper will show an empirical validation of the expected FHT bounds derived by the proposed method.

Acknowledgments

This work was supported by the National Science Fund for Distinguished Young Scholars of China under Grant No. 60325207.

References

Baase, S., and van Gelder, A. 2000. *Computer Algorithms: Introduction to Design and Analysis, 3rd ed.* Reading, MA: Addison-Wesley.

Bäck, T. 1996. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford, UK: Oxford University Press.

Beyer, H.-G.; Schwefel, H.-P.; and Wegener, I. 2002. How to analyse evolutionary algorithms. *Theoretical Computer Science* 287(1):101–130.

Droste, S.; Jansen, T.; and Wegener, I. 1998. A rigorous complexity analysis of the (1 + 1) evolutionary algorithm for linear functions with boolean inputs. *Evolutionary Computation* 6(2):185–196.

Eiben, A. E.; Hinterding, R.; and Michalewicz, Z. 1999. Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutionary Computation* 3(2):124–141.

Freitas, A. A. 2003. A survey of evolutionary algorithms for data mining and knowledge discovery. In Ghosh, A., and Tsutsui, S., eds., *Advances in Evolutionary Computing: Theory and Applications*. New York, NY: Springer-Verlag. 819–845.

Garnier, J.; Kallel, L.; and Schoenauer, M. 1999. Rigorous hitting times for binary mutations. *Evolutionary Computation* 7(2):173–203.

Hajek, B. 1982. Hitting time and occupation time bounds implied by drift analysis with applications. *Advances in Applied Probability* 14:502–525.

He, J., and Kang, L. 1999. On the convergence rates of genetic algorithms. *Theoretical Computer Science* 229(1-2):23–39.

He, J., and Yao, X. 2001. Drift analysis and average time complexity of evolutionary algorithms. *Artifical Intelligence* 127(1):57–85.

He, J., and Yao, X. 2003. Towards an analytic framework for analysing the computation time of evolutionary algorithms. *Artificial Intelligence* 145(1-2):59–97.

He, J., and Yao, X. 2004. A study of drift analysis for estimating computation time of evolutionary algorithms. *Natural Computing* 3(1):21–35.

He, J., and Yu, X. 2001. Conditions for the convergence of evolutionary algorithms. *Journal of Systems Architecture* 47(7):601–612.

Nix, A. E., and Vose, M. D. 1992. Modeling genetic algorithms with markov chains. *Annals of Mathematics and Artificial Intelligence* 5(1):77–88.

Rudolph, G. 1997. *Convergence Properties of Evolutionary Algorithms*. Hamburg, Germany: Verlag Dr. Kovač.

Rudolph, G. 1998. Finite markov chain results in evolutionary computation: A tour dhorizon. *Fundamenta Informaticae* 35(1-4):67–89.

Suzuki, J. 1995. A markov chain analysis on simple genetic algorithms. *IEEE Transactions on Systems, Man and Cybernetics* 25(4):655–659.

van Nimwegen, E.; Crutchfield, J.; and Mitchell, M. 1999. Statistical dynamics of the royal road genetic algorithm. *Theoretical Computer Science* 229(1-2):41–102.