A New Approach for Analyzing Average Time Complexity of Population-Based Evolutionary Algorithms on Unimodal Problems

Tianshi Chen, *Student Member, IEEE*, Jun He, *Member, IEEE*, Guangzhong Sun, Guoliang Chen, and Xin Yao, *Fellow, IEEE*

***Abstract*—In the past decades, many theoretical results related to the time complexity of evolutionary algorithms (EAs) on different problems are obtained. However, there is not any general and easy-to-apply approach designed particularly for population-based EAs on unimodal problems. In this paper, we first generalize the concept of the takeover time to EAs with mutation, then we utilize the generalized takeover time to obtain the mean first hitting time of EAs and, thus, propose a general approach for analyzing EAs on unimodal problems. As examples, we consider the so-called $(N+N)$ EAs and we show that, on two well-known unimodal problems, LEADINGONES and ONEMAX, the EAs with the bitwise mutation and two commonly used selection schemes both need $O(n \ln n + n^2/N)$ and $O(n \ln \ln n + n \ln n/N)$ generations to find the global optimum, respectively. Except for the new results above, our approach can also be applied directly for obtaining results for some population-based EAs on some other unimodal problems. Moreover, we also discuss when the general approach is valid to provide us tight bounds of the mean first hitting times and when our approach should be combined with problem-specific knowledge to get the tight bounds. It is the first time a general idea for analyzing population-based EAs on unimodal problems is discussed theoretically.**

***Index Terms*—Drift analysis, evolutionary algorithms (EAs), first hitting time, takeover time, unimodal problems.**

I. Introduction

EVOLUTIONARY Algorithms (EAs) are adaptive search algorithms. There have been lots of examples that successfully used EAs in solving various kinds of problems in many fields [3], [21], [23], [36], [39]. Many scholars also concerned the related theoretical study of EAs. One of the important topics in theoretical study is the time-complexity analysis of EAs [2], [6], [26], [32]. During the last decade, there are many theoretical results related to the time complexity of EAs, and among which, the study of EAs on unimodal problems [4], [5], [9], [10], [13], [16], [19], [20], [30], [38] plays an important role. Unimodal problems can be considered as the simplest kind of problems for EAs, and the related analysis reveals how EAs solve optimization problems. In practice, an EA often uses a number of individuals, i.e., the population. There are a few papers that discussed how population-based EAs optimize unimodal problems [13], [16], [20], [38]; however, these results mainly restrict to case studies, and the corresponding results are obtained by different approaches, e.g., drift analysis [13], analytical Markov chain [15], [30], Chernoff bounds [24], and so on. None of the study discusses how population-based EAs solve unimodal problems or provides a general approach to analyze population-based EAs on these unimodal problems.

In this paper, we develop a new general approach to analyze the time complexity of the population-based EAs on unimodal problems, which is based on some well-known concept and models of EAs, such as the takeover time [11], the Markov chain model [9], [15], [33], [34], and the supermartingale model [12], [13], [16], [29]. First, we generalize the original concept of takeover time introduced by Goldberg and Deb [11]: Rather than for a selection operator alone, the takeover time is generalized for EAs with mutation operators in this paper. After that, by some techniques developed on the Markov chain and supermartingale models of EAs, we combine the generalized takeover times together and obtain the overall first hitting time of population-based EAs on some unimodal problems.

As examples, we employ the so-called $(N+N)$ EAs with the bitwise mutation and two commonly used selection schemes, the truncation and two-tournament selection, to show our general approach. These EAs belong to a simplified version of the commonly used $(\mu+\lambda)$ EA [1], but the offspring size is set to be equal to the parent size. We discuss the $(N+N)$ EAs so that we can avoid the discussion of parent and offspring sizes, which is not related to the idea of our general approach. Nevertheless, it is worth noting that our idea is not limited to the $(N+N)$ EAs only but also valid to the $(\mu+\lambda)$ EA.

Manuscript received June 21, 2008; revised October 15, 2008. First published March 24, 2009; current version published September 16, 2009. This work was supported in part by the National Natural Science Foundation of China under Grants 60533020 and U0835002, by the Fund for Foreign Scholars in University Research and Teaching Programs under Grant B07033, by the Fund for International Joint Research Program of Anhui Science and Technology Department under Grant 08080703016, and by an Engineering and Physical Science Research Council Grant in U.K. under Grant EP/C520696/1. This paper was recommended by Associate Editor H. Takagi.

T. Chen, G. Sun, and G. Chen are with the Nature Inspired Computation and Applications Laboratory, Department of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China.

J. He was with the Centre of Excellence for Research in Computational Intelligence and Applications, School of Computer Science, University of Birmingham, B15 2TT Birmingham, U.K. He is now with the Univesity of Wales, SY23 2AX Aberystwyth, U.K.

X. Yao is with the Nature Inspired Computation and Applications Laboratory, Department of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China, and also with the Centre of Excellence for Research in Computational Intelligence and Applications, School of Computer Science, University of Birmingham, B15 2TT Birmingham, U.K. (e-mail: x.yao@cs.bham.ac.uk).

Digital Object Identifier 10.1109/TSMCB.2008.2012167

1083-4419/$25.00 © 2009 IEEE

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

On the other hand, we utilize two well-known unimodal problems, the LEADINGONES and ONEMAX problems in case studies. These problems have received intensive studies in the field of time complexity of EAs [5], [9], [13], [16], [20], [30], [38], and there are already some theoretical results related to some population-based EAs on these problems, such as the $(N+N)$ EAs [13], [16], the $(1+\lambda)$ EA [20], [28], and the $(\mu+1)$ EA [8], [38]. However, the EAs with large parent and offspring sizes have not received enough study, e.g., the mean first hitting time of the $(N+N)$ EAs on LEADINGONES is still unknown. As case studies, in Section IV, we apply our approach to $(N+N)$ EAs on the two problems, and we obtain some new results. From the analysis, it is shown that our general approach cannot only deal with the EAs with large parent and offspring sizes (in our analysis, $\mu=\lambda=N\gg 1$) but also be capable of obtaining some of the upper bound results related to some other kinds of population-based EAs on the two unimodal problems mentioned in the previous literatures.

It is worth noting that the focus of this paper is not on developing new EAs to solve unimodal problems more efficiently. Instead, this paper describes and models the behaviors of different population-based EAs from a theoretical perspective and proposes a novel approach to analyzing the time complexity of population-based EAs on unimodal problems by combining drift analysis and the concept of takeover time together. The new approach leads to a number of new results that are hard to achieve using either idea independently. As a consequence, we are able to gain insight toward the relationship between some algorithmic features and problem characteristics and, thus, move closer toward the ultimate goal of guiding the design of new EAs in practice using theoretical insight. However, further studies are needed in analyzing population-based EAs on more complex problems (with more complex-problem characteristics). The study presented in this paper serves as a solid and necessary foundation for such future work.

The rest of this paper is organized as follows. Section II contains preliminaries of this paper. Section III generalizes the original concept of takeover time and then estimates the overall first hitting time. Section IV applies the propositions obtained in Section III to get new time-complexity results on the LEADINGONES and ONEMAX problems. Section V discuss when our general approach can achieve tighter bounds of computation times of EAs and the potential reasons that our approach cannot reach tight bounds in some cases. In addition, we discuss briefly the possibility of generalizing our approach to more complex problems. Section VI concludes the whole paper.

II. PRELIMINARIES

A. Problems and Algorithms

In this paper, we consider some *maximization* problems $f(x): S\to R$, where $S\subseteq \mathbb{B}^n$ and $x=(s_1,\ldots,s_n)$ is a binary string with length n ($n\in\mathbb{N}$ is called the problem size). The problems all contain a unique global optimum. We consider a particular kind of maximization problems: the unimodal problems.

In order to provide the formal definition of the problems, we should give the definition of Hamming distance first.

Definition 1 (Hamming Distance): The Hamming distance between two binary strings $x=(s_1,\ldots,s_n)$ and $y=(s'_1,\ldots,s'_n)$ is given by $H(x,y)=\sum_{i=1}^n |s_i-s'_i|$.

Given a maximization problem f, a binary string $x^*=(s_1^*,\ldots,s_n^*)$ is called an optimum if, and only if, $\forall x$ that satisfy $H(x,x^*)=1$, $f(x^*)>f(x)$ holds. Then, we can give the formal definition of a unimodal problem.

Definition 2 (Unimodal Problem): A maximization problem $f: S\to R$ is said to be unimodal if, and only if, there is a unique optimum x^* and $\forall x\neq x^*$, $\exists y$ satisfies $H(x,y)=1$ and $f(y)\geq f(x)$.

The LEADINGONES and ONEMAX problems are two well-known examples of unimodal problems, and they are formally defined as follows.

Definition 3 (LEADINGONES [31]): $\text{LEADINGONES}(x)=\sum_{i=1}^n\prod_{j=1}^i s_j$, where $x=(s_1,\ldots,s_n)$.

Definition 4 (ONEMAX [25]): $\text{ONEMAX}(x)=\sum_{i=1}^n s_i$, where $x=(s_1,\ldots,s_n)$. EAs can be used to solve the unimodal problems. Given the initial population ξ_0, the framework of EAs at the tth generation is described as follows.

1) **Recombination**: Individuals in population ξ_t are recombined, and the population $\xi_t^{(C)}$ is obtained.
2) **Mutation**: Individuals in population $\xi_t^{(C)}$ are mutated. An offspring population $\xi_t^{(M)}$ is then obtained.
3) **Selection**: Individuals in the population $\xi_t^{(M)}$ and ξ_t are assigned a survival probability. Then, select some individuals as the next-generation ξ_{t+1} based on the probability.

These procedures are repeated unless the stopping criteria are met. The investigations in which the EAs use some recombination operators (e.g., crossover [22]) will leave to our future work, and we only consider some EAs with mutation and selection in this paper.

The EAs analyzed in this paper are known as the $(N+N)$ EAs. In every generation, the $(N+N)$ EAs considered in this paper use N individuals, and each individual generates its offspring by the following bitwise mutation.

- **Bitwise Mutation**: All bits of an individual in ξ_t will be flipped with a uniform probability p_m, where p_m is called the mutation rate. In this paper, we consider a commonly used value $p_m=1/n$ [7].

After mutation, all individuals, includes the parents and offsprings, will be chosen by the selection scheme. The selection scheme will decide the population of the next generation. In this paper, the EAs all adopt the following elitist strategy in selection.

- **Elitist Selection Strategy (ESS)**: Retain the best individuals in ξ_t and $\xi_t^{(M)}$ directly, the rest of the individuals of the population are selected by some selection schemes.

The following two selection schemes which adopt ESS will be considered in this paper.

1) **Selection I (Truncation Selection)**: Sort $2N$ individuals in ξ_t and $\xi_t^{(M)}$ by their fitness in descending order. Then, select N best individuals as the next generation ξ_{t+1}.
2) **Selection II (Two-Tournament Selection + ESS)**: Retain the best individual in the population directly. Select two individuals randomly from the pool of $2N$

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

individuals (before and after mutation), retain the better one. Repeat this step until N offsprings have been generated for the next-generation ξ_{t+1}.

Concretely, we will consider two $(N+N)$ EAs in this paper: The first one uses the bitwise mutation and Selection I and the second one uses the bitwise mutation and Selection II.

B. Mathematical Concepts and Tools

We introduce the measure of the performances for EAs. It is well known that EAs can be measured by Markov chain [15], [33], [34]. Consequently, the following concept of first hitting time defined for Markov chains can be used to measure the time complexity of EAs.

Definition 5 (First Hitting Time): For a given Markov chain $L:\{L_t, t=0,1,\ldots\}(L_t \in M)$ and a subspace H we are interested in, the first hitting time to H is defined by

$$\tau = \min\{t \geq 0; L_t \in H\}. \tag{1}$$

A corresponding definition of the first hitting time of EAs is given by

$$\tau = \min\{t \geq 0; x^* \in \xi_t\} \tag{2}$$

where ξ_t is the population of the tth generation. The corresponding number of function evaluations of an $(N+N)$ EA to find the global optimum for the first time, denoted by T, is given by $T = N\tau$.

Moreover, if we only concern the first hitting time of an EA, the EA can be further modeled by absorbing Markov chain [14], [15]. The definition of absorbing Markov chain is as follows.

Definition 6 (Absorbing Markov Chain): A Markov chain $L:\{L_t, t=0,1,\ldots\}(L_t \in M)$ is absorbing if there is a subspace $H \subset M$ such that

$$\mathbb{P}(L_{t+1} = L_t | L_t \in H) = 1. \tag{3}$$

If no self-adaptive strategy is used in an EA, then the Markov chain model of the EA is homogeneous. Denote $T = M - H$. The transition matrix $\mathbf{P}$ of the Markov chain associated with the EA is as follows:

$$\mathbf{P} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R} & \mathbf{T} \end{pmatrix} \tag{4}$$

?

where the submatrix $\mathbf{I}$ records the transition probabilities among the absorbing states in H, which is an identity matrix, the submatrix R records the transition probabilities from the states in the subspace T to the absorbing states, and the submatrix $\mathbf{T}$ records the transition probabilities among the states in the subspace T. The following lemma shows an analytical method of Markov chain [15], [18].

Lemma 1: Let τ_i be the first hitting time of Markov Chain $L:\{L_t, t=0,1,\ldots\}$ when starting from state i and denote $m_i = \mathbb{E}[\tau_i]$, $\mathbf{m} = [m_i]_{i \in T}$, then

$$\mathbf{m} = (\mathbf{I} - \mathbf{T})^{-1}\mathbf{1} \tag{5}$$

where $\mathbf{I}$ is an identity matrix and $\mathbf{1}$ denotes the vector $(1,\ldots,1)^{\mathrm{t}}$.

Most often, we apply this lemma when we can estimate the transition matrix of the Markov chain easily. There is another lemma related to the mean first hitting time of the absorbing Markov chain [16], [35]; in this paper, we cite it without proof.

Lemma 2: Assume that $L:\{L_t, t=0,1,\ldots\}(L_t \in M)$ is a homogeneous absorbing Markov chain with the target subspace $H \subset M$. Then, the following equations hold:

$$\begin{cases} \mathbb{E}[\tau|L_0{=}X]{=}0, & X \in H \\ \mathbb{E}[\tau|L_0{=}X]{-}\sum_{Y\in M}\mathbb{P}(X,Y)\mathbb{E}[\tau|L_0{=}Y]{=}1, & X \notin H \end{cases}$$

where τ is defined in (1) and $\mathbb{E}[\tau|L_0 = X]$ is the expected first hitting time of the absorbing Markov chain L, given the condition that the initial state of L is X, i.e., $L_0 = X$. The value of the conditional expectation $\mathbb{E}[\tau|L_0 = X]$ is given by

$$\mathbb{E}[\tau|L_0 = X] = \sum_{k=0}^{+\infty} k\mathbb{P}(\tau = k|L_0 = X).$$

The concepts and tools introduced so far are proposed for the Markov chain model of EAs. An alternative model of EAs is the supermartingale model [13], [16]. The well-known technique, drift analysis [12], [13], [15], [16], [18], [27], [29], was developed on this model, and it is powerful for estimating the mean first hitting time of the EAs. Drift analysis requires distance functions for both individuals and populations to measure their distances to the optimum. Let x^* be the unique optimum of the objective function; let $V(X)$ be the distance between the population X and the optimum x^*; then, the one-step mean drift $\Delta(X,t)$ is given by

$$\begin{aligned} \Delta(X,t) &= \mathbb{E}\left[V(\xi_t) - V(\xi_{t+1})|\xi_t = X\right] \\ &= \sum_{y\in E}\left(V(X) - V(Y)\right)\mathbb{P}(X,Y;t) \end{aligned}$$

where E is the population set (the set that contains all the populations). If the EA does not adopt any self-adaptive strategies, then the transition probability is not time-dependent

$$\mathbb{P}(X,Y;t) = \mathbb{P}(X,Y).$$

In this case, the notation $\Delta(X)$ is used to represent the one-step mean drift for short. If

$$\Delta(X) = \mathbb{E}\left[V(\xi_t) - V(\xi_{t+1})|\xi_t = X\right] \geq 0$$

holds for $t = 0,1,\ldots$, then $\{V(\xi_t): t=0,1,\ldots\}$ is called a supermartingale. Once we can estimate the lower bound of the one-step mean drift, then we can get the upper bound of the mean first hitting time of the EA through the following lemma.

Lemma 3: Let $\{V(\xi_t): t=0,1,\ldots\}$ be a supermartingale describing an EA, if for any time $t = 1,2,\ldots$, if $V(\xi_t) > 0$ and

$$\mathbb{E}\left[V(\xi_t) - V(\xi_{t+1})|\xi_t\right] \geq c_l > 0$$

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

then the mean first hitting time satisfies

$$\mathbb{E}[\tau|\xi_0] \leq \frac{V(\xi_0)}{c_l}$$

where ξ_0 is the initial population of the EA.

The proof of this lemma can be found in [13].[1] To apply this lemma, it is crucial to find a suitable distance function $V(\cdot)$ such that $\{V(\xi_t) : t = 0, 1, \ldots\}$ is a supermartingale. Fortunately, since the condition

$$\mathbb{E}\left[V(\xi_t) - V(\xi_{t+1})|\xi_t\right] \geq c_l > 0$$

has already implied that $\{V(\xi_t) : t = 0, 1, \ldots\}$ is a supermartingale, to apply Lemma 3, we only need to find a distance function which satisfies the earlier inequality. He and Yao [16] utilized a number of different distance functions to study the time complexity of an EA. However, it does not mean that a distance function has to be chosen from these specific functions every time. One can choose a suitable distance function (which satisfies the conditions mentioned earlier) according to the need of analysis. An appropriate choice can lead to a tighter bound for the first hitting time of an EA.

Moreover, Chernoff bounds [24] are powerful tools for investigating stochastic algorithms [5], [20]. In this paper, we will utilize this tool to deal with the stochastic behaviors of the selection operator for population-based EAs. A common form of Chernoff bounds is as follows.

Lemma 4 (Chernoff Bound): Let $b_1, b_2, \ldots, b_k \in \{0, 1\}$ be k-independent random variables with the same distribution

$$\forall i \neq j : \mathbb{P}(b_i = 1) = \mathbb{P}(b_j = 1) \quad \mathbb{P}(b_i = 0) = \mathbb{P}(b_j = 0)$$

where $i, j \in \{1, \ldots, k\}$. Let $b = \sum_{i=1}^{k} b_i$, then $\forall 0 < \delta < 1$

$$\mathbb{P}\left(X < (1-\delta)\mathbb{E}[b]\right) < \mathrm{e}^{-\mathbb{E}[b]\delta^2/2}.$$

In this paper, we use the following notations to compare the asymptotic order of functions [20].

1) $f = O(g)$, iff $\exists n_0 \in \mathbb{N}$, $c \in \mathbb{R}^+ : \forall n > n_0$, $f(n) \leq cg(n)$.
2) $f = \Omega(g)$, iff $g = O(f)$.
3) $f = \Theta(g)$, iff $f = O(g)$ and $f = \Omega(g)$ both hold.
4) $f = o(g)$, iff $\lim_{n\to\infty} f(n)/g(n) = 0$.
5) $f = \omega(g)$, iff $g = o(f)$.

In the notations, n is the problem size.

III. From the Takeover Times to the First Hitting Time

In this section, we will generalize the concept of the takeover time first. After that, we will use the generalized takeover time to bound the mean first hitting time of the population-based EAs on unimodal problems.

A. Generalize the Takeover Time

The takeover time of a selection scheme is the number of generations needed for the best individuals to fill a population under pure selection without other operators. Its formal definition is given as follows.

Definition 7 (Takeover Time): Let $\mathcal{A}$ be an iterated selection process,[2] where $\mathcal{A}$ is initialized by a selection operator and a population of individuals. At each generation, the selection operator selects a number of individuals from the current population to form the population of the next generation without changing the population size. Let N_t be the number of individuals with the best fitness in the population at the tth generation ($t \in \mathbb{N}_0$); the takeover time (denoted by η) of the iterated selection process $\mathcal{A}$ is defined by

$$\eta = \min\{t; N_t = N\} \tag{6}$$

where N is the population size.

The takeover time was proposed by Goldberg and Deb [11], and it is a useful concept to study and compare the selection schemes, but the previous study is restricted to pure selection process only.

Since the EAs with both mutation and selection also contain similar steps of accumulating promising individuals by the help of selection operators, it would be helpful if we generalize the concept of aforementioned takeover time further to those EAs. Once the generalization can be done, we are able to know more about how population-based EAs solve optimization problems and, consequently, provide some time-complexity results on some problems.

To illustrate our idea, we take the EAs with ESS on the LEADINGONES problem as examples. We show how the takeover time is generalized for the population-based EAs with both mutation and selection through the examples. Next, we introduce the metric for the later decomposition of the whole population space E, where E contains all the populations.

1) For an individual $x = (s_1, \ldots, s_n)$, we use the Hamming distance: $H(x, \mathbf{1}) = \sum_{i=1}^{n} |s_i - 1|$.
2) For a population X, define

$$d(X) = \max\left\{H(y, \mathbf{1}); f(y) = \max\left\{f(z); z \in X\right\}\right\}.$$

Based on the metric $d(X)$, the whole population set E can be divided into $(n+1)$ subsets, E_k $(k = 0, 1, 2, \ldots, n)$, where

$$E_k = \{X|d(X) = k\}, \qquad k = 0, \ldots, n$$

and furthermore, we let E_k^- be $\bigcup_{i=0}^{k-1} E_i$. Moreover, for any population $X \in E_k$, we define a subset of X

$$G = \{g|f(g) = \max\{f(z); z \in X\}, g \in X\}$$

[1]It is noteworthy that c_l is restricted to value one in [16] to show the existence of distance function. However, there is no such restriction for c_l in our case, since the lower bound one can always be obtained by updating the distance function $V(\cdot)$ through multiplying an additional factor $1/c_l$ (given the condition that there does exist a positive number c_l such that $\mathbb{E}[V(\xi_t) - V(\xi_{t+1})|\xi_t] \geq c_l$ holds).

[2]In other words, the process is equivalent to the EA with selection operator only.

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

where f represents the fitness function, and in this paper, it is refereed to the LEADINGONES function. The elements of G are only the current best individuals in X (if $d(X) = 0$, these elements are the global best). Given different numbers of elements in G, the population subsets E_k $(k = 0, \ldots, n)$ can further be decomposed into several subsets. To introduce the decompositions in detail, we need the following definition.

Definition 8 (LOI): Given a population set E_k and a population $X \in E_k$, we call an individual $g \in G$ the local optimal individual (LOI) of X on E_k (LOI for short, if we restrict the discussion on a given k and X), if and only if $g \in G$; we call an individual g' the advanced LOI for $\forall X \in E_k$ (advanced LOI for short, if we restrict the discussion on a given k), if and only if g' is the LOI of Y, where Y is a population that satisfies $Y \in \bigcup_{i=0}^{k-1} E_i = E_k^-$.

The search of a population-based EA on LEADINGONES can be divided into $n+1$ different stages $(E_n, \ldots, E_1, E_0)$. In E_k, the LOIs of a population keep on accumulating; Pessimistically, we can consider that the advanced LOIs will not be produced until the LOIs has filled the whole population or reach a certain proportion of the population. Once the advanced LOIs have been produced, the evolution in E_k finishes (never return to E_k if the EA adopts an elitist strategy), and a new similar process begins in $E_i (0 < i < k)$. In this paper, the takeover time is mainly used to measure the time of the LOIs' accumulation, which is similar to the original definition of the takeover time. However, the definition of the takeover time should still be generalized to make us available to divide the entire evolution "path" of the EAs on LEADINGONES into some small "steps" and model the behaviors of EAs on each step.

Let $l = |G|$, then the population set E_k can be further divided into some subsets $E_{k,l}$, where $1 \le l \le N$. If a population X belongs to the subset $E_{k,i} (1 \le i < N)$, then at the next generation, its offspring population X', may belong to $N+1$ different population sets $E_{k,1}, E_{k,2}, \ldots, E_{k,N}$ and E_k^-. If $X' \in E_{k,N}$ or $X' \in E_k^-$, then we say "X' *strictly* takes over." However, in practice, the population will not always be taken over by the LOIs, "many" is more realistic then "all": even if the LOIs has not filled the population, the advanced LOIs can take a very large probability to be produced. Hence, we make a compromise or relaxation, allowing the so-called "(k, ϵ)-takeover" ($0 < \epsilon \le 1$, ϵ is a constant).

Definition 9 ((k, ϵ)-Takeover): A population $X \in E_k$ is said to be (k, ϵ)-takeover if, and only if, its number of the LOIs has reached ϵN[3] or X has already contained some advanced LOIs on E_k.

If a population is not (k, ϵ)-takeover, we pessimistically consider that the advanced LOIs cannot be generated. Once a population is (k, ϵ)-taken over, it will concentrate on producing advanced LOIs on E_k^-. When the offspring population has reached E_k^-, if the selection operator preserves the best individual in every generation, then the population will begin a new trip in E_k^- and never return to E_k. Now, we formally introduce the definition of the (k, ϵ)-takeover time. First, for the population ξ_{t_k} in the t_kth generation, we define its (k, ϵ) first hitting time as

$$\eta_\epsilon^{(k)}(\xi_{t_k}) = \min\left\{ t_k' - t_k; \xi_{t_k'} \in \bigcup_{i=\epsilon N}^{N} E_{k,i} \cup E_k^- \middle| \xi_{t_k} \right\} \quad (7)$$

where $\eta_\epsilon^{(k)}$ can be considered as the first hitting time to the population subset $E_{k,\epsilon N}$. Its expectation is called the (k, ϵ) *takeover time*, which is given by

$$\bar{\eta}_\epsilon^{(k)}(X) = \mathbb{E}\left[\eta_\epsilon^{(k)}(\xi_{t_k}) | \xi_{t_k} = X \right].$$

Obviously, if $X \in \bigcup_{i=\epsilon N}^{N} E_{k,i} \cup E_k^-$, then $\bar{\eta}_\epsilon^{(k)}(X) = 0$. Then, we define the maximal (k, ϵ) takeover time as

$$\bar{\eta}_{\max,\epsilon}^{(k)} = \max\left\{ \bar{\eta}_\epsilon^{(k)}(X); X \in E_k \right\}$$

and the maximal ϵ takeover time is defined by

$$\bar{\eta}_{\max,\epsilon} = \max\left\{ \bar{\eta}_{\max,\epsilon}^{(k)}; 1 \le k \le n \right\}.$$

On the other hand, when the population of the EA has already (k, ϵ) taken over, i.e., the number of the LOIs has reached a proportion of ϵ on E_k, the probability of generating at least one advanced LOI in one generation (the procedure is called *upgrade*), namely, the (k, ϵ) *upgrade probability* (denoted by $p_{u,\epsilon}^{(k)}$), is given by

$$p_{u,\epsilon}^{(k)} = \min\left\{ \sum_{Y \in E_k^-} \mathbb{P}(X, Y); X \in E_{k,\epsilon N} \right\}$$

where $\mathbb{P}(X, Y)$ is the transition probability from X to Y. We call the reciprocal of the upgrade probability the *upgrade time*. For the EA with ESS, the upgrade time measures the time of the EA spent in generating at least one advanced LOI after the population has (k, ϵ) taken over.

B. From the Takeover Times to the First Hitting Time

In this section, we use the homogeneous absorbing Markov chains to model population-based EAs and estimate the upper bound of the mean first hitting time by calculating the generalized takeover times. The result will be proven by drift analysis, and the proof will use the definitions given in the last section.

We still take the $(N+N)$ EA with the ESS on LEADINGONES problem as an example. Without losing any generality, we choose $\epsilon = 1/2$. In order to bound the overall first hitting time of EAs, we should bound the $(1/2, k)$ takeover time by restricting the Markov chain on E_k. We carry out some pessimistic analysis: If the number of the LOIs of a population Z has not reached ϵN, we ignore all potential emergences of advanced LOIs. Concretely, once one or more advanced LOIs are generated when the number of the LOIs of a population Z has not reached ϵN, we can compulsively transform those

[3]For convenience, we assume that ϵN is an integer in this paper.

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

advanced LOIs to LOIs by some fixed rule (e.g., let every advanced LOI be replaced by some LOI that is randomly selected from the LOIs of the current population). As the result of the compulsive transformation, the population Z is transformed to $Y(Y \in E_{k,i})$ first. Afterward, according to the values of i (number of LOIs), we need to consider two cases: $(N/2) \leq i \leq N$ and $1 \leq i < (N/2)$. The following mapping "$\rightarrow_k$" from E_k^- to E_k represents the second pessimistic compulsive transformation.

1) If $(N/2) \leq i \leq N$, then $Z \rightarrow_k Y$ (an advanced LOI is transformed to a LOI).
2) If $1 \leq i < (N/2)$, then $Z \rightarrow_k Y'$, where Y' is obtained by directly replacing the best $(N/2)$ individuals of Y with $(N/2)$ randomly selected LOIs, $Y' \in E_{k,(N/2)}$ (we ignore the advanced LOIs but consider that the population has $(1/2, k)$ taken over).

After the transformation, the population is always restricted on E_k. Consequently, we can introduce an auxiliary homogeneous absorbing Markov chain $(\zeta_t^{(k)}, \mathrm{t} = 0, 1, \ldots)$ that is strictly restricted on E_k. The corresponding transition probabilities of $\zeta^{(k)}$ are given in (8), shown at the bottom of this page. In (8), $X, Y \in E_k$.

It is easy to see that

$$\sum_{Y \in E_k} \bar{\mathbb{P}}(X, Y) = 1.$$

Obviously, $\bigcup_{i=(N/2)}^{N} E_{k,i} \cup E_k^-$ is absorbing in $\zeta^{(k)}$. The other subspaces are $E_{k,1}, \ldots, E_{k,(N/2)-1}$. According to Lemma 2, we know that, for $X \in \bigcup_{i=1}^{(N/2)-1} E_{k,i}$

$$\sum_{Y \in E_k} \left(\bar{\eta}_{\frac{1}{2}}^{(k)}(X) - \bar{\eta}_{\frac{1}{2}}^{(k)}(Y)\right) \bar{\mathbb{P}}(X, Y) = 1. \tag{9}$$

Now, we give the first result of this paper as follows.

Proposition 1: For a given selection strategy, if $\forall X \in \bigcup_{i=(N/2)}^{N} E_{k,i}$

$$\sum_{Y \in \bigcup_{i=1}^{\frac{N}{2}-1} E_{k,i}} \bar{\eta}_{\frac{1}{2}}^{(k)}(Y) \mathbb{P}(X, Y) \rightarrow 0 \tag{10}$$

holds when $n \rightarrow \infty$ and $N(n) = \omega(1)$,[4] then the mean first hitting time of the $(N+N)$ EA with this selection strategy (and also preserves the best individual in every generation) and the bitwise mutation on the LEADINGONES problem satisfies

$$\mathbb{E}[\tau] = O\left(\sum_{k=1}^{n} \left(\bar{\eta}_{\max,1/2} + 1/p_{u,1/2}^{(k)}\right)\right)$$

where $\bar{\eta}_{\max}$ is dependent on the selection strategy.

Proof: According to Definition 5, starting from the initial population ξ_0, the first hitting time to the subspace E_0 is given by

$$\tau = \min\{t \geq 0; \xi_t \in E_0 | \xi_0\}.$$

Let ξ_{t_k} be the population of the t_kth generation. For $\xi_{t_k} \in E_k$, we define its first hitting time to the population set E_k^-, starting from ξ_{t_k}

$$\tau_k = \min\left\{t_k' - t_k \geq 0; \xi_{t_k'} \in E_k^- \middle| \xi_{t_k} \in E_k\right\}.$$

The conditional mean first hitting time (to the set E_k^-) of EAs started from the population $\xi_{t_k} = X$, denoted by τ_k, is given by

$$\mathbb{E}[\tau_k | \xi_{t_k} = X] = \sum_{t=0}^{\infty} t \mathbb{P}(\tau_k = t | \xi_{t_k} = X).$$

Let μ_{t_k} be the distribution of the population in the t_kth generation (note that we have assumed that $\xi_{t_k} \in E_k$), then

$$\mathbb{E}[\tau_k] = \sum_{Y \in E_k} \mu_{t_k}(Y) \mathbb{E}[\tau_k | \xi_{t_k} = Y]$$

is called the mean first hitting time to population set E_k^-.

We now carry out drift analysis to bound $\mathbb{E}[\tau_k]$. First, define a distance function $V^{(k)}(X)$ for $X \in E_k \cup E_k^-$, it is as follows:

$$V^{(k)}(X) = \begin{cases} 0, & \forall X \in E_k^- \\ 1/p_{u,1/2}^{(k)} + \bar{\eta}_{\frac{1}{2}}^{(k)}(X), & \forall X \in E_k. \end{cases}$$

For each population set E_k $(k = 1, \ldots, n)$, we show that the one-step mean drift of the populations in it is no less than some positive constant when the problem size $n \rightarrow \infty$. Assume that $X \in E_k$, let $\Delta(X)$ be its one-step mean drift, then there will be two cases (11) and (12), shown at the bottom of the next page. Given the results obtained in (11) and (12), we have shown that the one-step mean drift of X is always no less than some positive constant. According to Lemma 3, we further have

$$\mathbb{E}\left[\tau_k | \xi_{t_k} = X\right] = O\left(1/p_{u,1/2}^{(k)} + \bar{\eta}_{\max,1/2}\right).$$

[4]Equation (10) holds for many selection schemes, such as two-tournament or linear ranking selection, which can be proven by Chernoff bounds [24]. We will discuss it in Lemma 6.

$$\bar{\mathbb{P}}(X, Y) = \begin{cases} \mathbb{P}(X, Y), & X \neq Y,\ X \notin \bigcup_{i=\frac{N}{2}}^{N} E_{k,i},\ Y \notin \bigcup_{i=\frac{N}{2}}^{N} E_{k,i} \\ \mathbb{P}(X, Y) + \sum_{Z \in E_k^-, Z \rightarrow_k Y} \mathbb{P}(X, Z), & X \neq Y,\ X \notin \bigcup_{i=\frac{N}{2}}^{N} E_{k,i},\ Y \in \bigcup_{i=\frac{N}{2}}^{N} E_{k,i} \\ 0, & X \neq Y,\ X \in \bigcup_{i=\frac{N}{2}}^{N} E_{k,i},\ Y \in \bigcup_{i=\frac{N}{2}}^{N} E_{k,i} \\ 0, & X \neq Y,\ X \in \bigcup_{i=\frac{N}{2}}^{N} E_{k,i},\ Y \notin \bigcup_{i=\frac{N}{2}}^{N} E_{k,i} \\ 1 - \sum_{Y \neq X} \bar{\mathbb{P}}(X, Y), & X = Y \end{cases} \tag{8}$$

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

Then, the first hitting time from E_k to E_k^- satisfies

$$\begin{aligned}\mathbb{E}[\tau_k] &= \sum_{X\in E_k} \mu_{t_k}(X)\mathbb{E}\left[\tau_k|\xi_{t_k}=X\right]\\ &= O\left(1/p_{u,1/2}^{(k)} + \bar{\eta}_{\max,1/2}\right)\sum_{X\in E_k}\mu_{t_k}(X)\\ &= O\left(1/p_{u,1/2}^{(k)} + \bar{\eta}_{\max,1/2}\right).\end{aligned}$$

For $X \in E_{k'}(k' \neq k)$, by similar way, we can obtain the same lower bound of the one-step mean drift. On the other hand, the selection operator of the EAs considered here preserves the best individual in every generation (ESS); thus, once the population of the EA has reached E_k^-, it will never return to E_k again. Hence, combining the takeover times directly, we can bound the overall first hitting time for the $(N+N)$ EAs

$$\mathbb{E}[\tau] = O\left(\sum_{k=1}^{n}\left(\bar{\eta}_{\max,1/2} + 1/p_{u,1/2}^{(k)}\right)\right).$$

■

Proposition 1 provides us with an approach to bound the first hitting time of the population-based EAs. Note that the condition described in (10) is necessary to our result, and we will investigate it specifically then.

On ONEMAX, similar result can be obtained immediately by the same technique.

Proposition 2: For a given selection strategy, if $\forall X \in \bigcup_{i=N/2}^{N} E_{k,i}$

$$\sum_{Y\in\bigcup_{i=1}^{\frac{N}{2}-1}E_{k,i}} \bar{\eta}_{\frac{1}{2}}^{(k)}(Y)\mathbb{P}(X,Y) \to 0 \tag{13}$$

holds when $n \to \infty$ and $N(n) = \omega(1)$, then the mean first hitting time of the $(N+N)$ EA with this selection strategy (and also preserves the best individual in every generation) and the bitwise mutation on the ONEMAX problem satisfies

$$\mathbb{E}[\tau] = O\left(\sum_{k=1}^{n}\left(\bar{\eta}_{\max,1/2} + 1/p_{u,1/2}^{(k)}\right)\right)$$

where $\bar{\eta}_{\max}$ is dependent on the selection strategy.

IV. Analysis of the $(N+N)$ EAs: Case Study

In this section, we study the time complexity of the $(N+N)$ EAs with Selections I and II by applying the propositions proven in the last section. The most crucial step of applying the propositions given in the last section is to find a way to estimate the maximal ϵ takeover time. A natural idea is that we

---

For $X \in \bigcup_{i=\frac{N}{2}}^{N} E_{k,i}, Y \in E_k \cup E_k^-$

$$\begin{aligned}\Delta(X) &= \sum_{Y\in E_k^-}\left(V^{(k)}(X) - V^{(k)}(Y)\right)\mathbb{P}(X,Y) + \sum_{Y\in E_k}\left(V^{(k)}(X) - V^{(k)}(Y)\right)\mathbb{P}(X,Y)\\ &\geq 1 - \sum_{Y\in\bigcup_{i=1}^{\frac{N}{2}-1}E_{k,i}} \bar{\eta}_{\frac{1}{2}}^{(k)}(Y)\mathbb{P}(X,Y) > c, \qquad \%\text{The condition given in (10)}\end{aligned} \tag{11}$$

where $c \in (0,1)$ is some constant

For $X \in \bigcup_{i=1}^{\frac{N}{2}-1} E_{k,i}, Y \in E_k \cup E_k^-$

$$\begin{aligned}\Delta(X) &= \sum_{Y\in E_k^-}\left(V^{(k)}(X) - V^{(k)}(Y)\right)\mathbb{P}(X,Y) + \sum_{Y\in E_k}\left(V^{(k)}(X) - V^{(k)}(Y)\right)\mathbb{P}(X,Y)\\ &\geq \left(\sum_{Y\in E_k^-} + \sum_{Y\in\bigcup_{i=\frac{N}{2}}^{N}E_{k,i}}\right)\bar{\eta}_{\frac{1}{2}}^{(k)}(X)\mathbb{P}(X,Y) + \sum_{Y\in\bigcup_{i=1}^{\frac{N}{2}-1}E_{k,i}}\left(\bar{\eta}_{\frac{1}{2}}^{(k)}(X) - \bar{\eta}_{\frac{1}{2}}^{(k)}(Y)\right)\mathbb{P}(X,Y)\\ &= \sum_{Y\in\bigcup_{i=\frac{N}{2}}^{N}E_{k,i}} \bar{\eta}_{\frac{1}{2}}^{(k)}(X)\bar{\mathbb{P}}(X,Y) + \sum_{Y\in\bigcup_{i=1}^{\frac{N}{2}-1}E_{k,i}}\left(\bar{\eta}_{\frac{1}{2}}^{(k)}(X) - \bar{\eta}_{\frac{1}{2}}^{(k)}(Y)\right)\mathbb{P}(X,Y)\\ &= \sum_{Y\in E_k}\left(\bar{\eta}_{\frac{1}{2}}^{(k)}(X) - \bar{\eta}_{\frac{1}{2}}^{(k)}(Y)\right)\bar{\mathbb{P}}(X,Y) = 1. \qquad \%\text{The result of (9)}\end{aligned} \tag{12}$$

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

can calculate the takeover time by means of Markov chains. According to the number of the LOIs, the populations have ϵN different potential states. By estimating the transition matrix $\mathbf{P}$ of the Markov chain and applying Lemma 1, we can calculate the exact maximal (k, ϵ) takeover time. However, the obtained result will be in a very complex form, and it would be hard for us to compare these results with former results. An alternative choice to estimate the takeover time is the approximate method by Goldberg and Deb [11]. This method assumes that the performance of a selection operator is in its *expected case*: In one generation, the proportion of a typical individual x after selection will be the expectation

$$R^{(s)}(x) = \mathbb{E}\left[Sel\left(R(x)\right)|\xi_t, \xi_t^{(M)}\right] \tag{14}$$

where $R(x)$ and $R^{(s)}(x)$ are proportions of x before and after selection in the tth generation and Sel stands for the random-selection operator. Under this assumption, the takeover time is estimated by deterministic difference or differential equations. They also verified their approximations by experiments. In this paper, in order to get more explicit expressions of the time complexity, we will adopt this approximate method.

Now, we start with the analysis of the $(N+N)$ EA with the bitwise mutation and Selection I, the truncation selection. Truncation selection assigns only the best S individuals a survival probability greater than zero, where S is called a "threshold." In this section, the truncation selection we consider retains the best N individuals in the $2N$ individuals. The mean first hitting time of the $(n+n)$ EA (where $N = n$, n is the problem size) with truncation selection on the ONEMAX problem has been studied in [16], and an upper bound of $O(n)$ has been achieved (i.e., $O(n^2)$ in the sense of function evaluations). In this paper, we will analyze both LEADINGONES and ONEMAX problems.

The first proposition obtained in this section is about the LEADINGONES problem.

Proposition 3: The first hitting time of the $(N+N)$ EA with the bitwise mutation and Selection I on LEADINGONES satisfies

$$\mathbb{E}[\tau] = O\left(n \ln N + \frac{n^2}{N}\right).$$

Proof—Step 1: We estimate the maximal 1/2 takeover time $\bar{\eta}_{\max,1/2}$ of the EA with the bitwise mutation and Selection I. We only consider the behaviors of the LOIs. Here, we should also make some compromises: First, if a population is not $(k, 1/2)$ takeover, we ignore the upgrade of all LOIs to advanced LOIs. Second, non-LOIs may generate LOIs, but here we ignore it.

In this section, we introduce some notations that are utilized in the proof. Let the probability for a LOI to copy itself without any change be p_1, the probability for a LOI to generate a worse individual (than itself) be p_2, and the probability for a LOI to generate a better individual (than itself) be p_3. These three probabilities satisfy

$$p_1 + p_2 + p_3 = 1 \quad p_1 = \left(1 - \frac{1}{n}\right)^n \quad p_3 \geq \frac{1}{n}\left(1 - \frac{1}{n}\right)^{n-1}.$$

Let r be the expected proportion of LOIs of the parent population, and R to be the expected proportion of LOIs of the offspring population, then we have

$$\mathbb{P}\left(R = r + \frac{i}{N}\right) = \binom{rN}{i}(1-p_2)^i p_2^{rN-i}$$

where $0 \leq i \leq rN$. The above equation shows the transition probability from set $E_{k,rN}$ to $E_{k,rN+i}$. As a consequence, we can estimate R by r

$$\begin{aligned} R &= r + \sum_{i=0}^{rN} \frac{i}{N}\binom{rN}{i}(1-p_2)^i p_2^{rN-i} \\ &= r + (1-p_2)r = (2p_1 + p_2 + 2p_3)r \\ &\geq \left(1 + \left(1 - \frac{1}{n}\right)^n\right) r \geq \frac{5}{4} r. \end{aligned} \tag{15}$$

Then, deducted by the above equation from the initial value $r = (1/N)$ to the terminal value $R = (1/2)$, we can achieve the following equation describing an upper bound of the takeover time:

$$\frac{1}{N}(2p_1 + p_2 + 2p_3)^{\bar{\eta}_{\max,1/2}} = \frac{1}{2}.$$

By solving this equation, we can obtain the maximal 1/2 takeover time of the EA with truncation selection

$$\bar{\eta}_{\max,1/2} = O(\ln N).$$

Moreover, let us consider the condition of Proposition 1. It is obvious that the number of the LOIs will not decrease, since the selection operator always preserve the best N individuals among the total $2N$ individuals. Hence, if $X \in \bigcup_{i=(N/2)}^{N} E_{k,i}$, then for $\forall Y \in \bigcup_{i=1}^{(N/2)-1} E_{k,i}$, we have

$$\mathbb{P}(X, Y) = 0$$

which implies

$$\sum_{Y \in \bigcup_{i=1}^{\frac{N}{2}-1} E_{k,i}} \bar{\eta}_{\frac{1}{2}}^{(k)}(Y)\mathbb{P}(X, Y) = 0.$$

Step 2: Once we have $X \in E_{k,(N/2)} \cup E_{k,(N/2)+1} \cup \cdots \cup E_{k,N}$, $p_{u,1/2}^{(k)}$ can be calculated as follows:

$$\begin{aligned} p_{u,1/2}^{(k)} &\geq 1 - \left(1 - \frac{1}{n}\left(1 - \frac{1}{n}\right)^{\text{LEADINGONES(LOI)}}\right)^{\frac{N}{2}} \\ &\geq 1 - \left(1 - \frac{1}{en}\right)^{\frac{N}{2}} \\ &= 1 - \left(\left(1 - \frac{1}{en}\right)^{en}\right)^{\frac{N}{2en}} \end{aligned}$$

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

where LEADINGONES (LOI) is the fitness (objective function) value of the LOI. Hence, we have

$$1/p_{u,1/2}^{(k)} \leq \frac{1}{1-\left(\left(1-\frac{1}{en}\right)^{en}\right)^{\frac{N}{2en}}} \leq \frac{e^{\frac{N}{2en}}}{e^{\frac{N}{2en}}-1} \leq 1+\frac{2en}{N}.$$

Finally, according to Proposition 1, we have

$$\mathbb{E}[\tau] = O\left(n\ln N + \frac{n^2}{N}\right).$$

■

It is worth noting that if $N = O(n/\ln n)$, the proposition becomes

$$\mathbb{E}\left[\tau \middle| N = O\left(\frac{n}{\ln n}\right)\right] = O\left(\frac{n^2}{N}\right)$$
$$\mathbb{E}\left[T \middle| N = O\left(\frac{n}{\ln n}\right)\right] = O(n^2)$$

where $T = N\tau$ is the number of function evaluations of the EA.

On the other hand, we can also obtain the result on ONEMAX problem.

Proposition 4: The first hitting time of the $(N+N)$ EA with the bitwise mutation and Selection I on the ONEMAX problem satisfies

$$\mathbb{E}[\tau] = O\left(n\ln N + \frac{n\ln n}{N}\right).$$

The corresponding proof is given in the Appendix. Similar to the case of LEADINGONES, if $N = O((\ln n)/(\ln\ln n))$, we obtain

$$\mathbb{E}\left[\tau \middle| N = O\left(\frac{\ln n}{\ln\ln n}\right)\right] = O\left(\frac{n\ln n}{N}\right)$$
$$\mathbb{E}\left[T \middle| N = O\left(\frac{\ln n}{\ln\ln n}\right)\right] = O(n\ln n).$$

According to the proof of the proposition earlier, on the two unimodal problems, the takeover times of the EA with both the bitwise mutation and truncation selection can be obtained by Goldberg and Deb's approximation. Afterward, the upper bound of the mean first hitting time or the mean number of function evaluations of the corresponding $(N+N)$ EAs is obtained. Similarly, for EA with the two-tournament selection, we can estimate the mean first hitting times and the mean numbers of function evaluations on the two unimodal problems. However, we will prove two lemmas first. The first one is related to the takeover time of the $(N+N)$ EA; the other one discussed the conditions (10) and (13) of Propositions 1 and 2, respectively, when the two-tournament selection is adopted by the EA.

Lemma 5: The maximal 1/2 takeover time of the $(N+N)$ EA with the bitwise mutation and Selection II on LEADINGONES and ONEMAX satisfies $\bar{\eta}_{\max,1/2} = O(\ln N)$.

By Chernoff bounds [24], the following result which serves as the conditions in Propositions 1 and 2 can be proven for the $(N+N)$ EAs with Selection II on LEADINGONES and ONEMAX.

Lemma 6: For the $(N+N)$ EA $(N = N(n))$ with Selection II on LEADINGONES and ONEMAX, $\forall X \in \cup_{i=(N/2)}^{N} E_{k,i}, \forall k = 1,\ldots,n$

$$\sum_{Y \in \bigcup_{i=1}^{\frac{N}{2}-1} E_{k,i}} \bar{\eta}_{\frac{1}{2}}^{(k)}(Y)\mathbb{P}(X,Y) \to 0 \tag{16}$$

holds when $n \to \infty$ and $N(n) \to \infty$.

The proofs of both lemmas are provided in the Appendix. By the lemmas, we obtain the following propositions.

Proposition 5: The first hitting time of the $(N+N)$ EA with the bitwise mutation and Selection II on LEADINGONES satisfies

$$\mathbb{E}[\tau] = O\left(n\ln N + \frac{n^2}{N}\right).$$

Proposition 6: The first hitting time of the $(N+N)$ EA with the bitwise mutation and Selection II on ONEMAX satisfies

$$\mathbb{E}[\tau] = O\left(n\ln N + \frac{n\ln n}{N}\right).$$

We will not give the proof of these propositions, since they are direct corollaries of Lemmas 5 and 6 and Propositions 1 and 2.

On the basis of the earlier propositions for the first hitting time τ, we can compare the performances of the two $(N+N)$ EAs through their minimal numbers of the function evaluations $(T = N\tau)$. The results of the two $(N+N)$ EAs on the two unimodal problems are summarized in the Table I. Combining our results with the lower bounds results proven by Droste *et al.* [5], we know that tight bounds have been achieved on ONEMAX and LEADINGONES. On the two problems, given well-chosen population sizes, the two EAs have similar performances (in the sense of asymptotic order).

Moreover, in comparison with the $(1+1)$ EA, the $(N+N)$ EA may not be worse on the unimodal problems if we choose suitable population sizes. Probably, there are two reasons: First, the $(N+N)$ EAs preserve the diversity of individuals, which may shorten the evolution path, leads to the global optimum. Second, if a small population size is used, the takeover time will be small, which recovers the gap between the $(1+1)$ EA and the $(N+N)$ EA.

V. Conjectures and Discussions

In this section, we will discuss the results we have obtained. The discussions contain two parts: First, we provide some conjectures that there are tighter bounds than our previous results on LEADINGONES and ONEMAX. After that, we focus on the general approach for analyzing the population-based EAs on unimodal problems.

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COMPARISON AMONG THE MINIMAL MEAN FUNCTION EVALUATIONS OF EAs WITH DIFFERENT POPULATION SIZES OR SELECTION SCHEMES (THE FOLLOWING LOWER BOUNDS RESULTS ARE GIVEN BY DROSTE *et al.* [5])

Algorithm Type	ONEMAX MMFEs	Well-Chosen Population Size	LEADINGONES MMFEs	Well-Chosen Population Size
$(1+1)M1$ Elitist [5]	$\Theta(n \ln n)$	1	$\Theta(n^2)$	1
$(N+N)M1-S1$ Truncation	$\Theta(n \ln n)$	$O(\frac{\ln n}{\ln \ln n})$	$O(n^2)$	$O(\frac{n}{\ln n})$
$(N+N)M1-S2$ 2-tournament	$\Theta(n \ln n)$	$O(\frac{\ln n}{\ln \ln n})$	$O(n^2)$	$O(\frac{n}{\ln n})$

A. Tighter Upper Bounds

A large population size may benefit the EAs on unimodal problems, *in the sense of first hitting time*. For example, if the initialization of the EAs use the uniform distribution, then the EA with a larger population size will take higher probability of generating some individuals near the global optimum in the initialization.

Moreover, since the probability of generating the advanced LOIs will increase if the number of LOIs increases, a larger population size may accelerate the replication of the LOIs sometimes, or at least, it will not decelerate the replication in expectation.

Concretely, we consider the case of the EA with the bitwise mutation and two-tournament selection (the case of truncation selection is similar). Similar to (22), in the Appendix, we estimate the proportion of the LOIs after mutation and then write the equation that describes the expected proportion of the LOIs obtained by replication at the tth generation (denoted by R). Given the proportion of the LOIs at the $(t-1)$th generation (denoted by r), we have

$$1-\left(1-\frac{5}{8}r\right)^2 < R < 1-(1-r)^2 \tag{17}$$

where the lower bound of R is as (22). In the upper bound of R [right-hand side of (17)], we assume that the LOIs are not destroyed by mutation operator at all. We see that R is determined by the proportion of the LOIs after mutation at the $(t-1)$th generation, and we assume that this proportion is δr, where δ is a ratio. Since (17) holds, we can assume that R has the following form:

$$R = 1-(1-\delta r)^2 = 2\delta r - \delta^2 r^2$$

where $(5/8) < \delta < 1$. Let a be the number of the LOIs of the $(N+N)$ EA at the $(t-1)$th generation, then $r = (a/N)$. The expected number of the LOIs at the tth generation, denoted by b, is given by

$$b = 2\delta a - \frac{\delta^2 a^2}{N}.$$

In average, δ only depends on the mutation operator, thus a larger population size can increase the expected number of the LOIs by reducing the item $\delta^2 a^2/N$. Hence, the advanced LOIs are very likely to be generated earlier if a larger population is employed, which may lead to a smaller first hitting time. Now, we provide our conjecture formally.

Conjecture 1: For the $(N+N)$ EAs with the bitwise mutation and Selection I (or Selection II) on unimodal problems, assume that $N_1 = N_1(n)$ and $N_2 = N_2(n)$ are population sizes satisfying $N_1(n) = \omega(N_2(n))$, then we have

$$\mathbb{E}[\tau_{(N_1)}] = O\left(\mathbb{E}\left[\tau_{(N_2)}\right]\right)$$

where $\tau_{(N_1)}$ and $\tau_{(N_2)}$ are the first hitting times of the (N_1+N_1) EA and the (N_2+N_2) EA, respectively.

If Conjecture 1 holds, then we can further improve our previous results for LEADINGONES. According to the earlier conjecture, the mean first hitting time of the case $N = O(n/\ln n)$ can bound the mean first hitting time of the case $N = \omega(n/\ln n)$ from earlier. Hence, we improve the results obtained in the last section.

Conjecture 2: The first hitting time of the $(N+N)$ EA with the bitwise mutation and Selection I (or Selection II) on LEADINGONES satisfies

$$\mathbb{E}[\tau] = O\left(n \ln n + \frac{n^2}{N}\right).$$

Similar improvement may also hold for the ONEMAX problem. We use the upper bound of the mean first hitting time of the case $N = O((\ln n)/(\ln \ln n))$ to bound the case $N = \omega((\ln n)/(\ln \ln n))$.

Conjecture 3: The first hitting time of the $(N+N)$ EA with the bitwise mutation and Selection I (or Selection II) on ONEMAX satisfies

$$\mathbb{E}[\tau] = O\left(n \ln \ln n + \frac{n \ln n}{N}\right).$$

B. General Approach for Analyzing Population-Based EAs on Unimodal Problems

The approach we developed in this paper can be generalized to analyze EAs on some other unimodal problems, such as the long-path problem [4], [10], [17], [30], SPI problem [19] and so on. How does the general idea work? We will discuss this topic in this section.

In our analysis of the $(N+N)$ EAs on ONEMAX and LEADINGONES, we utilize implicitly a simple fact that, for every individual, there is always a path leading to the global optimum with increasing fitness, and the path is similar to that of the $(1+1)$ EA. This idea can be generalized, and a general approach for analyzing population-based EAs with the most commonly used bitwise mutation on other unimodal problems will be obtained.

First, considering the $(1+1)$ EA on unimodal problems, we introduce the following concepts

Definition 10 (Evolution Path and Path Gaps): The evolution path $\mathcal{L}$ of an $(1+1)$ EA with the elitist selection that

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

accepts individual with equal fitness is a directed path that is made up of different individuals with nondecreasing fitness. Let $\mathcal{B}(t)$ be the individual at the tth generation, then $\mathcal{L}$ is constructed as follows.

$\mathcal{L}(0) = \mathcal{B}(0), i = 1, t = 1;$
REPEAT
  `if` $f(\mathcal{B}(t)) \geq f(\mathcal{L}(i))$ `and` $\mathcal{B}(t) \neq \mathcal{L}(i)$
  $\mathcal{L}(i) = \mathcal{B}(t);$
    $i = i + 1;$
    $t = t + 1;$
  `else`
    $t = t + 1;$
UNTIL $\mathcal{A}$ FINDS THE GLOBAL OPTIMUM.

The path gap, denoted by $\Delta_g(i)(i \geq 0)$, records the Hamming distance between $\mathcal{L}(i-1)$ and $\mathcal{L}(i)$

$$\Delta_g(i) = H\left(\mathcal{L}(i+1), \mathcal{L}(i)\right).$$

If $\Delta_g(i) = h$, then $\Delta_g(i)$ is called an h-path gap, where $h = 0, 1, \ldots, n$. $\|\mathcal{L}\|$ is defined to be the length of $\mathcal{L}$.

The definition of the evolution path is deterministic. However, in practice, a stochastic algorithm is very unlikely to generate such a deterministic behavior in different runs. Thus, the concept of evolution path here can only demonstrate some "trend" of the random behaviors of the algorithm. The trend of an evolution path $\mathcal{L}$ can be represented by functions (of the problem size n) of the amounts of the h gap $(h = 0, 1, \ldots, n)$. Nevertheless, the constant parameters of the functions can be ignored. For two paths $\mathcal{L}_1$ and $\mathcal{L}_2$, if $\forall h$, the amount functions of the h gap, are with the same order, then we can say that the two paths are *the same*.

Now, we consider the behaviors of individuals in population-based EAs with the bitwise mutation. Here, we ignore the effect of initialization and consider the following condition.

Condition 1: The best individuals in the $(N+N)$ EAs walk along the same path $\mathcal{L}$ to the unique individual in the $(1+1)$ EA.

Assume that Condition 1 holds, then the first hitting times and the numbers of the function evaluations of $(N+N)$ EAs can be represented in the following simple (probably not rigorous) form:

$$\tau_{N+N} = \sum_{i:i\leq\|\mathcal{L}\|, f(\mathcal{L}(i))<f(\mathcal{L}(i+1))} \left(\eta_{g(N),\mathcal{L}(i)} + u_{g(N),\mathcal{L}(i)}\right) + \sum_{i:i\leq\|\mathcal{L}\|, f(\mathcal{L}(i))=f(\mathcal{L}(i+1))} C_{N,\mathcal{L}(i)} \tag{18}$$

$$T_{N+N} = N \sum_{i:i\leq\|\mathcal{L}\|, f(\mathcal{L}(i))<f(\mathcal{L}(i+1))} \left(\eta_{g(N),\mathcal{L}(i)} + u_{g(N),\mathcal{L}(i)}\right) + N \sum_{i:i\leq\|\mathcal{L}\|, f(\mathcal{L}(i))=f(\mathcal{L}(i+1))} C_{N,\mathcal{L}(i)} \tag{19}$$

where $\eta_{N,\mathcal{L}(i)}$ is the takeover time when facing the gap $(\mathcal{L}(i), \mathcal{L}(i+1))$, $u_{N,\mathcal{L}(i)}$ is the upgrade time for traveling through this gap, and $g(N) \leq N$ is called *the active population*.[5] When $f(\mathcal{L}(i)) = f(\mathcal{L}(i+1))$, the gap $(\mathcal{L}(i), \mathcal{L}(i+1))$ is called a plateau [19]; hence, we have $\eta_{N,\mathcal{L}(i)} = 0$ and $u_{N,\mathcal{L}(i)} = 0$. In addition, in order to deal with the costs on plateaus in $\mathcal{L}$, in the equation earlier, there is another item $\sum_{i:i\leq\|\mathcal{L}\|, f(\mathcal{L}(i))=f(\mathcal{L}(i+1))} C_{\mathcal{L}(i)}$ describing the costs of the random walks on plateaus, which is denoted by $C_{N,\text{plateaus}}$ in the following parts of this paper.

If the asymptotic order (of the problem size n) of the takeover time is no higher than that of the upgrade time, then the whole population is said to be *active*, i.e., the population is very likely to be taken over by the LOIs before one or more LOIs upgrades, and thus, a number of LOIs can join the attempts of generating advanced LOIs. Moreover, the maximal size of the so-called active population should promise that the takeover time can be bounded from above by the upgrade time, or at most, with the same order (of the problem size n) to the upgrade time. Formally, we have

$$\eta_{\bar{N},\mathcal{L}(i)} = \Theta\left(u_{\bar{N},\mathcal{L}(i)}\right)$$

where $\bar{N}$ is called the maximal active population size. If $N = O(\bar{N}(n))$, we say that *the population size is well chosen*. In this case, the upper bounds shows in (18) and (19) are very likely to be tight

$$\begin{aligned} \tau_{N+N} &= k_1 \sum_{i:i\leq\|\mathcal{L}\|, f(\mathcal{L}(i))<f(\mathcal{L}(i+1))} u_{N,\mathcal{L}(i)} + C_{N,\text{plateaus}} \\ T_{N+N} &= k_1 N \sum_{i:i\leq\|\mathcal{L}\|, f(\mathcal{L}(i))<f(\mathcal{L}(i+1))} u_{N,\mathcal{L}(i)} + N C_{N,\text{plateaus}} \end{aligned} \tag{20}$$

where $k_1 > 1$ is a constant. It is noteworthy that, when we set $N = 1$, the equations earlier presents the case of them $(1+1)$ EA

$$\tau_{1+1} = T_{1+1} = \sum_{i:i\leq\|\mathcal{L}\|, f(\mathcal{L}(i))<f(\mathcal{L}(i+1))} u_{1,\mathcal{L}(i)} + C_{1,\text{plateaus}}. \tag{21}$$

Since Condition 1 holds, by simple calculation of upgrading probability, we can obtain that the first items of (20) and (21) are in the same order of the problem size n

$$\sum_{i:i\leq\|\mathcal{L}\|, f(\mathcal{L}(i))<f(\mathcal{L}(i+1))} u_{1,\mathcal{L}(i)} = \Theta\left(N \sum_{i:i\leq\|\mathcal{L}\|, f(\mathcal{L}(i))<f(\mathcal{L}(i+1))} u_{N,\mathcal{L}(i)}\right).$$

In other words, when Condition 1 holds and the population size of the $(N+N)$ EA is well chosen, the large difference between the performances of the $(1+1)$ and $(N+N)$ EAs, if there is any, mainly comes from the costs on plateaus. In this case, when the affect of the plateaus can be neglected, the

[5]If the population is too large, it is very likely for the advanced LOIs to be produced before LOIs take over the population. In other words, only part of the population has been taken over, we call this part "the active population."

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

$(N+N)$ EA can perform as good as the $(1+1)$ EA, and a tight bound of the mean first hitting time can be achieved.

However, Condition 1 will not hold sometimes. Since the population-based EAs take the advantage of diversity, it is possible that population-based EAs can find some shortcuts on the landscape to reduce the overall number of function evaluations for solving some particular problems (if there are), while the $(1+1)$ EA is very likely to miss those shortcuts. In this case, the upper bound achieved by our general approach will not be tight, particularly when the population size of the population-based EA is large (consequently, the probability of finding shortcuts becomes large). In order to achieve tighter bounds, we should use some problem-specific knowledge in the analysis. The discussion of such kind of problems will leave to our future work.

Our approach might also be valid for analyzing different types of population-based EAs on unimodal problems, since the "takeover-upgrade" process still exists in the general $(\mu+\lambda)$ EA framework. Once the takeover and upgrade times of the EAs can be estimated, then it is not difficult to achieve the bounds of the mean first hitting times. However, it is still noteworthy that the discussions here are not rigorous enough, and one can use similar method (e.g., drift analysis, Chernoff bounds) to the previous theoretical investigations of this paper to make the analysis more rigorous.

Finally, we discuss briefly the possibility of applying our approach to some deceptive multimodal problems which contain more than one optima. Let us consider a special deceptive problem with a global optimum and a local optimum: The global optimum of the problem is with no basin of attraction,[6] while the basin of attraction of the local optimum is whole search space (except the global optimum). The global optimum is very far away from the local optimum (measured by Hamming distance). In this case, it is not hard to use our approach directly to obtain the mean first hitting time of the local optimum. By considering the probabilities of jumping to the global optimum from the individuals on the evolution path leads to the local optimum, it is not hard to prove that the mean first hitting time of finding the global optimum is very large. However, our current approach is still not capable of analyzing population-based EAs on more complex multimodal problems, since these problems may contain a number of basins of attraction, and the interactions of these basins will interrupt the so-called "takeover-upgrade" processes, leading to a specific attractor (optimum). To analyze population-based EAs on multimodal problems in the future, it is essential to find some analytical tools to model the selection operators of EAs such that the interactions among different basins (leading to different optima) can be quantitatively analyzed.

VI. Conclusion

In this paper, we propose a new approach for analyzing the computational time complexity of population-based EAs on unimodal problems by combining drift analysis and the concept of takeover time together. It is shown that the $(N+N)$ EA with the bitwise mutation and truncation selection (or two-tournament selection) needs $O(n\ln n+n^2/N)$ and $O(n\ln\ln n+n\ln n/N)$ generations to find the global optimum of the LEADINGONES and ONEMAX problems, respectively. We further discuss how our ideas can be extended to a more general approach and when the obtained bounds will be tight. It is the first time that a general approach to analyzing *population-based* EAs on unimodal problems is proposed.

The results obtained in this paper show that, on LEADINGONES and ONEMAX, the $(N+N)$ EAs are not worse than the $(1+1)$ EA (in the sense of asymptotic order of the problem size n) when the population sizes are well chosen. On the other hand, when the population sizes N are small, since the evolution paths of the two kinds of EAs are similar, and the takeover times of the $(N+N)$ EAs are small (hence, compared with the upgrade time, the takeover time can be omitted), the general conclusion on all unimodal problems that the population-based EAs are not worse than the $(1+1)$ EA probably holds.

However, when the population size is large, exact comparison between the two kinds of EAs become hard. It is possible that population-based EAs perform significantly better, since the two kinds of EAs may choose different evolution path sometimes: The population-based EAs can choose every step very carefully while the $(1+1)$ may lose some shortcuts. It is also possible that population-based EAs take no advantage, e.g., when the problems contain lots of plateaus [19]. In this case, in order to achieve tighter time-complexity bounds in theoretical analysis, our approach should be combined with some problem-specific characteristics.

Our future work will focus on generalizing the approach described in this paper to more complex population-based EAs on more complex multimodal problems. We will also try to find some unimodal examples that the $(1+1)$ and population-based EAs perform significantly different. In particular, the role of recombination and the interactions among different operators need to be studied in depth.

Appendix

Proof of Proposition 4: Step 1 will be the same with the proof of the Proposition 3.

Step 2: Once we have $X\in E_{k,N/2}\cup E_{k,N/2+1}\cup\cdots\cup E_{k,N}$, $p_{u,1/2}^{(k)}$ is bounded from below by

$$\begin{aligned}p_{u,1/2}^{(k)}&\geq 1-\left(1-\frac{k}{n}\left(1-\frac{1}{n}\right)^{n-1}\right)^{\frac{N}{2}}\\&\geq 1-\left(1-\frac{k}{en}\right)^{\frac{N}{2}}\geq 1-e^{-\frac{kN}{2en}}.\end{aligned}$$

Hence, we have

$$\begin{aligned}1/p_{u,1/2}^{(k)}&\leq\frac{1}{1-e^{-\frac{kN}{2en}}}\leq 1+\frac{2en}{kN}\\\sum_{i=1}^{n}1/p_{u,1/2}^{(k)}&\leq n+\sum_{i=1}^{n}\frac{2en}{kN}=n+\frac{2en\ln n}{N}.\end{aligned}$$

[6]Basin of attraction [37]: "The set of points in the space of system variables such that initial conditions chosen in this set dynamically evolve to a particular attractor."

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

Moreover, as we have shown in the proof of Proposition 3, the condition of Proposition 2 also holds. Hence, according to Proposition 2, we have

$$\mathbb{E}[\tau] = O\left(n \ln N + \frac{n \ln n}{N}\right).$$

■

Proof of Lemma 5: We estimate the maximal $(k, (1/2))$ takeover time of the EA with Selection II. We only consider the behaviors of the best individuals in a population, i.e., the behaviors of LOIs. Our analysis is in a pessimistic style: First, if the population is not 1/2 takeover, we ignore the upgrade of all LOIs to advanced LOIs. Second, non-LOIs may generate LOIs, but here we ignore it.

Here, we borrow the notations p_1, p_2, and p_3 from the proof of Proposition 3. It is easy to see that

$$p_1 + p_2 + p_3 = 1 \quad p_1 = \left(1 - \frac{1}{n}\right)^n \quad p_3 \geq \frac{1}{n}\left(1 - \frac{1}{n}\right)^{n-1}.$$

Let r be the expected proportion of LOIs in the parent population, R be the expected proportion of LOIs in the offspring population, and K be the proportion of LOIs in the pool for selection. Given $2N$ individuals, the lower bound of K can be calculated as follows:

$$K \geq \frac{(p_1 + p_3)rN + rN}{2N} = \frac{1}{2}(p_1 + p_3)r + \frac{1}{2}r.$$

We can denote p as the probability for EAs to select a LOI in one turn of the tournament selection (N turns in all), and we can estimate its lower bound in the following way:

$$\begin{aligned} p &= 1 - (1-K)^2 \geq 1 - \left(1 - \frac{1}{2}(p_1 + p_3)r - \frac{1}{2}r\right)^2 \\ &\geq 1 - \left(1 - \frac{1}{2}\left(1 - \frac{1}{n}\right)^n r - \frac{1}{2}r\right)^2 \geq 1 - \left(1 - \frac{5}{8}r\right)^2 \end{aligned}$$

where p is the probability for LOIs to be selected as a LOI in the offspring population. We know that the two-tournament selection strategy can be considered as an action repeating N times. Hence, we have

$$\begin{aligned} R &= \sum_{i=0}^{n} \frac{i}{N}\binom{N}{i} p^i (1-p)^{N-i} = \frac{Np}{N} = p \\ &\geq 1 - \left(1 - \frac{5}{8}r\right)^2 = \frac{5}{4}r - \frac{25}{64}r^2. \end{aligned} \tag{22}$$

Assume that the current population is in the $(t-1)$th generation restricted in Markov chain $\zeta^{(k)}$, we add the generation number to the earlier inequality

$$R_t - R_{t-1} \geq \frac{1}{4}R_{t-1} - \frac{25}{64}R_{t-1}^2.$$

Consider the upper case of the running time t by setting "$\geq$" to be "$=$" in the earlier inequality. Approximating this finite difference by its derivative in one step, we get the following differential equation:

$$\frac{dR(t)}{dt} = \frac{1}{4}R(t) - \frac{25}{64}R^2(t).$$

By setting $R(0) = (1/N)$, we can get

$$R(t) = \frac{16e^{\frac{t}{4}}}{25e^{\frac{t}{4}} + 16N - 25}.$$

Now, it is easy for us to prove this lemma by solving $R(\bar{\eta}_{\max,1/2}) = (1/2)$

$$\bar{\eta}_{\max,1/2} \leq 4 \ln\left(\frac{1}{7}(16N - 25)\right).$$

■

Proof of Lemma 6: For the $(N+N)$ EA with Selection II, we prove the fact by Chernoff bounds [24]. Assume that $X \in E_{k,i}$, where $(N/2) \leq i \leq N$. Denote the proportion of the LOIs of X as r, obviously, $r = (i/N)$. The expected number of the LOIs at the next generation, denoted by R, satisfies (see the Appendix, proof of Lemma 5)

$$R > \frac{5}{4}r - \frac{25}{64}r^2 = \frac{5i}{4N} - \frac{25i^2}{64N^2} > \frac{135}{256}.$$

Assume that $Y \in E_{k,j}$, where $1 \leq j < (N/2)$. We have the expected number of the LOIs in hand, then we can use Lemma 4 (Chernoff bounds [24]) to investigate the actual number of LOIs after selection: Since the two-tournament selection for the $(N+N)$ EA can be regarded as an N-independent Bernoulli trials with the same parameter, according to Lemma 4, we obtain

$$\sum_{Y \in E_{k,j}, j < \frac{N}{2}} \mathbb{P}(X, Y) < e^{-NR\delta^2/2}$$

where RN is the expected number of the number of LOIs after selection, j is the actual number of LOIs after selection, and δ satisfies

$$\delta = \frac{RN - j}{RN} > 1 - \frac{1}{2R} = \Theta(1)$$

since $j < (N/2)$. Hence, we have

$$\sum_{Y \in E_{k,j}, j < \frac{N}{2}} \mathbb{P}(X, Y) < e^{-\Theta(N)}.$$

According to Lemma 5, for the two-tournament selection, we have

$$\bar{\eta}^{(k)}_{\max,\frac{1}{2}} = O(\ln N).$$

Hence, given $n \to \infty$

$$\sum_{Y \in \bigcup_{j=1}^{\frac{N}{2}-1} E_{k,j}} \bar{\eta}^{(k)}_{\frac{1}{2}}(Y)\mathbb{P}(X, Y) < O(\ln N)e^{-\Theta(N)} \to 0$$

holds, since $N = \omega(1)$. ■

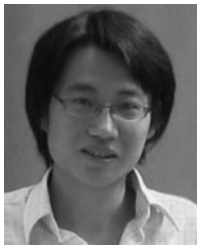
Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.

ACKNOWLEDGMENT

The authors would like to thank Dr. K. Tang and Dr. P. K. Lehre and also the anonymous reviewers for their valuable suggestions.

REFERENCES

[1] H.-G. Beyer, *The Theory of Evolution Strategies*, ser. Natural Computing Series. Heidelberg, Germany: Springer-Verlag, 2001.
[2] H.-G. Beyer, H.-P. Schwefel, and I. Wegener, "How to analyse evolutionary algorithms," *Theor. Comput. Sci.*, vol. 287, no. 1, pp. 101–130, Sep. 2002.
[3] X.-B. Cao, H. Qiao, and J. Keane, "A low-cost pedestrian-detection system with a single optical camera," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 58–67, Mar. 2008.
[4] S. Droste, T. Jansen, and I. Wegener, "On the optimization of unimodal functions with the $(1+1)$ evolutionary algorithm," in *Proc. PPSN V*, Amsterdam, The Netherlands, 1998, pp. 13–22.
[5] S. Droste, T. Jansen, and I. Wegener, "On the analysis of the $(1+1)$ evolutionary algorithm," *Theor. Comput. Sci.*, vol. 276, no. 1/2, pp. 51–81, Apr. 2002.
[6] A. E. Eiben and G. Rudolph, "Theory of evolutionary algorithms: A bird eye view," *Theor. Comput. Sci.*, vol. 229, no. 1/2, pp. 3–9, Nov. 1999.
[7] A. E. Eiben, R. Hinterding, and Z. Michalewicz, "Parameter control in evolutionary algorithms," *IEEE Trans. Evol. Comput.*, vol. 3, no. 2, pp. 124–141, Jul. 1999.
[8] T. Friedrich, P. S. Oliveto, D. Sudholt, and C. Witt, "Theoretical analysis of diversity mechanisms for global exploration," in *Proc. GECCO*, Atlanta, GA, 2008, pp. 945–952.
[9] J. Garnier, L. Kallel, and M. Schoenauer, "Rigorous hitting times for binary mutations," *Evol. Comput.*, vol. 7, no. 2, pp. 167–203, 1999.
[10] J. Garnier and L. Kallel, "Statistical distribution of the convergence time of evolutionary algorithms for long path problems," *IEEE Trans. Evol. Comput.*, vol. 4, no. 1, pp. 16–30, Apr. 2000.
[11] D. E. Goldberg and K. Deb, "A comparative analysis of selection scheme used in genetic algorithms," in *Foundations of Genetic Algorithms*. San Mateo, CA: Morgan Kaufmann, 1991, pp. 69–93.
[12] B. Hajek, "Hitting time and occupation time bounds implied by drift analysis with applications," *Adv. Appl. Probab.*, vol. 14, no. 3, pp. 502–525, 1982.
[13] J. He and X. Yao, "Drift analysis and average time complexity of evolutionary algorithms," *Artif. Intell.*, vol. 127, no. 1, pp. 57–85, Mar. 2001.
[14] 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. Evol. Comput.*, vol. 6, no. 5, pp. 495–511, Oct. 2002.
[15] J. He and X. Yao, "Towards an analytic framework for analysing the computation time of evolutionary algorithms," *Artif. Intell.*, vol. 145, no. 1/2, pp. 59–97, Apr. 2003.
[16] J. He and X. Yao, "A study of drift analysis for estimating computation time of evolutionary algorithms," *Natural Comput.*, vol. 3, no. 1, pp. 21–35, Mar. 2004.
[17] J. Horn, D. E. Goldberg, and K. Deb, "Long path problems," in *Proc. PPSN III*, Jerusalem, Israel, 1994, pp. 149–158.
[18] M. Iosifescu, *Finite Markov Processes and Their Applications*. Chichester, U.K.: Wiley, 1980.
[19] T. Jansen and I. Wegener, "Evolutionary algorithms—How to cope with plateaus of constant fitness and when to reject strings of the same fitness," *IEEE Trans. Evol. Comput.*, vol. 5, no. 6, pp. 589–599, Dec. 2001.
[20] T. Jansen, K. A. D. Jong, and I. Wegener, "On the choice of the offspring population size in evolutionary algorithms," *Evol. Comput.*, vol. 13, no. 4, pp. 413–440, 2005.
[21] S. Khuri, T. Bäck, and J. Heitkotter, "An evolutionary approach to combinatorial optimization problems," in *Proc. 22nd Annu. ACM Comput. Sci. Conf.*, 1994, pp. 66–73.
[22] P. K. Lehre and X. Yao, "Crossover can be constructive when computing unique input output sequences," in *Proc. 7th Int. Conf. SEAL*, Melbourne, Australia, 2008, pp. 595–604.
[23] K.-H. Liang, X. Yao, C. S. Newton, and D. Hoffman, "A new evolutionary approach to cutting stock problems with and without contiguity," *Comput. Oper. Res.*, vol. 29, no. 12, pp. 1641–1659, Oct. 2002.
[24] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 1995.
[25] H. Mühlenbein and D. Schlierkamp-Voosen, "Predictive models for the breeder genetic algorithm, I.: Continuous parameter optimization," *Evol. Comput.*, vol. 1, no. 1, pp. 25–49, 1993.
[26] P. S. Oliveto, J. He, and X. Yao, "Time complexity of evolutionary algorithms for combinatorial optimization: A decade of results," *Int. J. Autom. Comput.*, vol. 4, no. 3, pp. 281–293, Jul. 2007.
[27] P. S. Oliveto, J. He, and X. Yao, "Evolutionary algorithms and the vertex cover problem," in *Proc. IEEE CEC*, Singapore, 2007, pp. 1870–1877.
[28] P. S. Oliveto, J. He, and X. Yao, "Analysis of population-based evolutionary algorithms for the vertex cover problem," in *Proc. IEEE WCCI*, Hong Kong, 2008, pp. 1563–1570.
[29] P. S. Oliveto and C. Witt, "Simplified drift analysis for proving lower bounds in evolutionary computation," in *Proc. PPSN X*, Dortmund, Germany, 2008, pp. 82–91.
[30] G. Rudolph, "How mutation and selection solve long path problems in polynomial expected time," *Evol. Comput.*, vol. 4, no. 2, pp. 195–205, 1996.
[31] G. Rudolph, "Convergence properties of evolutionary algorithms," Ph.D. dissertation, Verlag Dr. Kovač, Hamburg, Germany, 1997.
[32] G. Rudolph, "Finite Markov chain results in evolutionary computation: A tour d'horizon," *Fundam. Inform.*, vol. 35, no. 1–4, pp. 67–89, Aug. 1998.
[33] J. Suzuki, "A Markov chain analysis on simple genetic algorithms," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 4, pp. 655–659, Apr. 1995.
[34] J. Suzuki, "A further result on the Markov chain model of genetic algorithms and its application to a simulated annealing-like strategy," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 28, no. 1, pp. 95–102, Feb. 1998.
[35] R. Syski, *Passage Times for Markov Chains*. Amsterdam, The Netherlands: IOS Press, 1992.
[36] M. Tang and X. Yao, "A memetic algorithm for VLSI floorplanning," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 1, pp. 62–69, Feb. 2007.
[37] E. W. Weisstein, "Basin of attraction," *MathWorld—A Wolfram Web Resource*. [Online]. Available: http://mathworld.wolfram.com/BasinofAttraction.html
[38] C. Witt, "An analysis of the $(\mu+1)$ EA on simple Pseudo-Boolean functions," in *Proc. GECCO*, Seattle, WA, 2004, pp. 761–773.
[39] J. X. Yu, X. Yao, C.-H. Choi, and G. Gou, "Materialized view selection as constrained evolutionary optimization," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 33, no. 4, pp. 458–467, Nov. 2003.



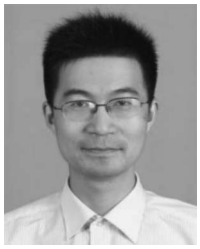
Tianshi Chen (S'07) received the B.Sc. degree in mathematics from the Special Class for the Gifted Young, University of Science and Technology of China, Hefei, China, in 2005, where he is currently working toward the Ph.D. degree in the Nature Inspired Computation and Applications Laboratory, Department of Computer Science and Technology.

His research interests include computational time complexity of evolutionary algorithms and various real-world applications of evolutionary algorithms.

Jun He (M'06) received the Ph.D. degree in computer science from Wuhan University, Wuhan, China in 1995.

From 2001 to 2007, he was with the School of Computer Science, University of Birmingham, Birmingham, U.K. He is currently a Researcher with the Department of Computer Science, University of Wales, Aberystwyth, U.K. His research interests include evolutionary computation, data mining, network security, and parallel computing. He has published over 50 refereed papers in the these areas.

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.



Guangzhong Sun received the B.Eng. and Ph.D. degrees from the University of Science and Technology of China (USTC), Hefei, China, in 2000 and 2005, respectively.

He is currently a Lecturer with the Nature Inspired Computation and Applications Laboratory, Department of Computer Science and Technology, USTC. His main research interests include high-performance computing and theoretical analysis of algorithms.

Guoliang Chen received the B.Sc degree from Xi'an Jiaotong University, Xi'an, China, in 1961.

Since 1973, he has been with the University of Science and Technology of China, Hefei, China, where he is currently the Academic Committee Chair of the Nature Inspired Computation and Applications Laboratory, a Professor with the Department of Computer Science and Technology, and the Director of the School of Software Engineering. From 1981 to 1983, he was a Visiting Scholar at Purdue University, IN. He is currently also the Director of the National High Performance Computing Center at Hefei, Hefei. His research interests include parallel algorithm, computer architecture, computer network, and computational intelligence. He has published 9 books and more than 200 research papers.

Prof. Chen is an Academician of Chinese Academy of Sciences. He was the recipient of the National Excellent Teaching Award of China in 2003.

Xin Yao (M'91–SM'96–F'03) received the B.Sc. degree from the University of Science and Technology of China (USTC), Hefei, China, in 1982, the M.Sc. degree from the North China Institute of Computing Technology, Beijing, China, in 1985, and the Ph.D. degree from USTC in 1990.

From 1985 to 1990, he was an Associate Lecturer and Lecturer with USTC, while working toward the Ph.D. degree in simulated annealing and evolutionary algorithms. In 1990, he was a Postdoctoral Fellow with the Computer Sciences Laboratory, Australian National University (ANU), Canberra, Australia, where he continued his work on simulated annealing and evolutionary algorithms. In 1991, he was with the Knowledge-Based Systems Group, CSIRO (Commonwealth Scientific and Industrial Research Organisation) Division of Building, Construction and Engineering, Melbourne, Australia, where he worked primarily on an industrial project on automatic inspection of sewage pipes. In 1992, he returned to Canberra to take up a lectureship in the School of Computer Science, University College, University of New South Wales, Australian Defence Force Academy, Sydney, NSW, Australia, where he was later promoted to a Senior Lecturer and Associate Professor. Attracted by the English weather, he moved to the University of Birmingham, Birmingham, U.K., where he was a Professor (Chair) of computer science on April 1, 1999. He is currently the Director of the Centre of Excellence for Research in Computational Intelligence and Applications, School of Computer Science, University of Birmingham. He is currently also a Changjiang (Visiting) Chair Professor (Cheung Kong Scholar) with the Nature Inspired Computation and Applications Laboratory, Department of Computer Science and Technology, University of Science and Technology of China. He has given more than 50 invited keynote and plenary speeches at conferences and workshops worldwide. He has more than 300 refereed publications. His major research interests include evolutionary artificial neural networks, automatic modularization of machine-learning systems, evolutionary optimization, constraint-handling techniques, computational time complexity of evolutionary algorithms, coevolution, iterated prisoner's dilemma, data mining, and real-world applications.

Dr. Yao was the Editor-in-Chief of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION from 2003 to 2008, an Associate Editor or editorial board member of 12 other journals, and the Editor of the *World Scientific Book Series on Advances in Natural Computation*. He was the recipient of the President's Award for Outstanding Thesis by the Chinese Academy of Sciences for his Ph.D. work on simulated annealing and evolutionary algorithms in 1989. He was the recipient of the 2001 IEEE Donald G. Fink Prize Paper Award for his work on evolutionary artificial neural networks.

Authorized licensed use limited to: Southern University of Science and Technology. Downloaded on January 21,2024 at 02:39:25 UTC from IEEE Xplore. Restrictions apply.