Contents lists available at SciVerse ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

A large population size can be unhelpful in evolutionary algorithms[☆]

Tianshi Chen [a,b], Ke Tang [b,*], Guoliang Chen [b], Xin Yao [b,c]

[a] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

[b] Nature Inspired Computation and Applications Laboratory (NICAL), School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, China

[c] Centre of Excellence for Research in Computational Intelligence and Applications (CERCIA), School of Computer Science, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

ARTICLE INFO

ABSTRACT

The utilization of populations is one of the most important features of evolutionary algorithms (EAs). There have been many studies analyzing the impact of different population sizes on the performance of EAs. However, most of such studies are based on computational experiments, except for a few cases. The common wisdom so far appears to be that a large population would increase the population diversity and thus help an EA. Indeed, increasing the population size has been a commonly used strategy in tuning an EA when it did not perform as well as expected for a given problem. He and Yao (2002) [8] showed theoretically that for some problem instance classes, a population can help to reduce the runtime of an EA from exponential to polynomial time. This paper analyzes the role of population further in EAs and shows rigorously that large populations may not always be useful. Conditions, under which large populations can be harmful, are discussed in this paper. Although the theoretical analysis was carried out on one multimodal problem using a specific type of EAs, it has much wider implications. The analysis has revealed certain problem characteristics, which can be either the problem considered here or other problems, that lead to the disadvantages of large population sizes. The analytical approach developed in this paper can also be applied to analyzing EAs on other problems.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

As a crucial characteristic of Evolutionary Algorithm (EA), the utilization of population enables explorations to different parts of the search space via a number of individuals. Although, over the past decades, most practical EAs have employed populations, rigorous theoretical investigations on the impact of population on evolutionary algorithms were mainly carried out in the past eight years. Concerning this issue, He and Yao [8] took one of the first attempts via the comparisons of the mean first hitting times of both $(N + N)$ and $(1 + 1)$ EAs on a class of multimodal problems derived from the well-known ONEMAX problem, and the purpose is to demonstrate the impact of population. Later, a number of theoretical investigations have been dedicated to study the first hitting times of EAs with either multiple parents [17,5,14] or offsprings [10,14]. It is expected that the EAs under investigations, which were known as $(\mu + 1)$ and $(1 + \lambda)$ EAs, can establish a bridge from analyzing the $(1 + 1)$ EA to studying $(\mu + \lambda)$ EAs. In the meantime, it has also been reported that the recent investigations on EAs with multiple parents and offsprings (e.g., $(N + N)$ EAs) have eventually brought to the community broader perspectives on understanding the behaviors of population-based EAs. Chen et al. analyzed the time complexity of $(N + N)$ EA on

unimodal problems [2] and a wide-gap problem [1]. Lehre and Yao studied the impact of mutation–selection balance on the performance of $(N + N)$ EA with linear ranking selection on a multimodal problem [11], but the influence of employing different population sizes was not investigated. It is still not clear how different population sizes influence the performance of $(N + N)$ EA on multimodal problems. In this paper, we carry out theoretical investigations on this issue.

To study the effect of population size via theoretical investigation, it is clear that we need a suitable measure of performance for EAs. Most previous investigations adopt the well-known first hitting time measure, which is a random variable demonstrating the number of generations required by the EA to find the global optimum for the first time. Here we denote by τ the first hitting time, and $\mathbb{P}[\tau \leq a]$ the accumulated probability of τ. We further denote by τ_1 and τ_2 the first hitting times of two arbitrary EAs, say, EA-I and EA-II, respectively. If the following conditions holds

- $a_1(n)$ is a polynomial function of the problem size n;
- $a_2(n)$ is a super-polynomial function of the problem size n;
- $\mathbb{P}[\tau_1 \leq a_1(n)]$ is super-polynomially close to 1;
- $\mathbb{P}[\tau_2 \leq a_2(n)]$ is super-polynomially close to 0,

then one can conclude that EA-I is more efficient than EA-II with a probability that is super-polynomially close to 1. A recent example, which successfully utilizes the above methodology in comparison of different EAs, is provided in [3]. However, when considering the following case, direct comparison of the first hitting times would become infeasible:

- $a_1(n)$ is a polynomial function of the problem size n;
- $a_2(n)$ is a super-polynomial function of the problem size n;
- $a_3(n)$ is a super-polynomial function of the problem size n;
- The reciprocal of $\mathbb{P}[\tau_1 \leq a_1(n)]$ is bounded from above by some polynomial function of n;
- The reciprocal of $\mathbb{P}[\tau_1 > a_2(n)]$ is bounded from above by some polynomial function of n;
- $\mathbb{P}[\tau_2 \leq a_3(n)]$ is super-polynomially close to 0.

The reason is that EA-I is likely to perform as inefficient as EA-II. Nevertheless, since EA-I still takes a relatively high probability to perform efficient while EA-II performs inefficiently almost surely, we can still compare the performances of EA-I and EA-II by employing the *solvable rate* as an alternative measure, where the solvable rate is the probability that the EA finds the global optimum of an optimization problem within a polynomial number of generations. The solvable rate can be considered as a generalized measure based on the probability distribution of the traditional first hitting time measure, and it concerns more about the probability with which an EA performs efficiently in general, rather than the detailed computation time for finding the optimum. In fact, in previous investigations (e.g., [13]), the idea of solvable rate has been utilized implicitly in company with the first hitting time results, though it has not been adopted as a measure of performance.

By employing the solvable rate measure in this paper, we carry out theoretical analysis to study the impact of the population size on the performance of an $(N + N)$ EA on a multimodal problem. This multimodal problem, which is called the TRAPZEROS problem, contains a global optimum $(1, \ldots, 1)$ and a local optimum $(0, \ldots, 0)$. The attraction basin of the global optimum only consists of solutions with the leading substring made up of $\ln^2 n + 2$ consecutive 1-bits. To find the global optimum, the EA has to enter its basin of attraction first by resisting the selection pressure which tends to preserve the solutions with leading 0-bits. For the $(N + N)$ EA on the above problem, we consider three cases for the population size N, $N = 1$, $N = O(\ln n)$ and $N = \Omega(n / \ln n)$, where the well-known $(1 + 1)$ EA is considered as a special case of $(N + N)$ EA ($N = 1$). It is discovered that when the population size is relatively small ($N = 1$ or $N = O(\ln n)$), the solvable rate of $(N + N)$ EA is still larger than the reciprocal of some polynomial function of the problem size, which implies that the EA, if running on an appropriate polynomial number of processors simultaneously and independently, can find the global optimum with a polynomial number of generations. However, given a large enough population size $N = \Omega(n / \ln n)$, the solvable rate of the $(N + N)$ EA has dropped to a level that is super-polynomially close to 0, implying that the EA cannot find the global optimum within a polynomial number of generations, unless one can offer a super-polynomial number of processors for the EA to run on.

The rest of the paper is organized as follows: Section 2 introduces the algorithm and problem investigated in this paper. Section 3 presents the mathematical tools utilized in this paper. Section 4 analyzes the $(N + N)$ EA with the population size $N = 1$. Section 5 shows the analytical result of the $(N + N)$ EA with the population size $N = O(\ln n)$. Section 6 concerns the $(N + N)$ EA with the population size $N = \Omega(n / \ln n)$. Section 7 carries out discussions on the results presented in the previous sections. Section 8 concludes the whole paper.

2. Problem and algorithm

In this section, we introduce the concrete optimization problem and EA investigated in this paper.

2.1. Problem

The *maximization* problem we consider in this paper, defined over the domain $\mathbf{x} = (x_1, \ldots, x_n) \in \{0, 1\}^n$, is called TRAPZEROS:
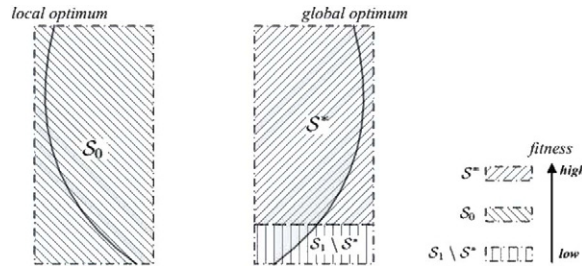
Fig. 1. Illustration of the TRAPZEROS problem.

$$
\text{TRAPZEROS}(\mathbf{x}) \triangleq
\begin{cases}
2n + \sum_{i=1}^{n}\prod_{j=1}^{i}(1 - x_j), & \text{if } (x_1 = 0) \wedge (x_2 = 0); \\[2mm]
3n + \sum_{i=1}^{n}\prod_{j=1}^{i}x_j, & \text{if } (x_1 = 1) \wedge (x_2 = 1) \wedge \left(\prod_{i=3}^{\ln^2 n + 2} x_i = 1 \right); \\[2mm]
n + \sum_{i=1}^{n}\prod_{j=1}^{i}x_j, & \text{if } (x_1 = 1) \wedge (x_2 = 1) \wedge \left(\prod_{i=3}^{\ln^2 n + 2} x_i = 0 \right); \\[2mm]
0, & \text{if } (x_1 = 0) \wedge (x_2 = 1); \\
1, & \text{if } (x_1 = 1) \wedge (x_2 = 0); \\
0, & \text{Other.}
\end{cases}
\tag{1}
$$

The TRAPZEROS problem is a multimodal problem, and its global optimum is $\mathbf{x}^* = (1, \ldots, 1)$. For TRAPZEROS, increasing the leading 0-bits in its solution may eventually lead to the local optimum $(0, \ldots, 0)$ instead of leading to the global optimum $(1, \ldots, 1)$. To reach the attraction basin of the global optimum, an optimization algorithm should first find some solutions with the leading substring consisting of $\ln^2 n + 2$ consecutive 1-bits. Otherwise, the selection pressure of an EA will tend to preserve the solutions with leading 0-bits. To facilitate the later investigations, we define three schemata as follows:

$$
\mathcal{S}_1 = \{(1, 1, *, \ldots, *)\},
$$
$$
\mathcal{S}_0 = \{(0, 0, *, \ldots, *)\},
$$
$$
\mathcal{S}^* = \left\{ (1, 1, \underbrace{1, \ldots, 1}_{\ln^2 n}, *, \ldots, *) \right\},
$$
$$
\mathcal{S}^* \in \mathcal{S}_1,
$$

where "$*$" can represent either 0 or 1, $\mathcal{S}^*$ and $\mathcal{S}_1$ are the schemata containing the global optimum. Fig. 1 illustrates the fitness landscape of TRAPZEROS with respect to the schemata defined above, and it shows that the individuals belonging to $\mathcal{S}^*$ are strictly better than any individual belonging to $\mathcal{S}_0$, while the individuals belonging to $\mathcal{S}_0$ are strictly better than any individual belonging to $\mathcal{S}_1 \setminus \mathcal{S}^*$. Utilizing this property, we will carry out rigorous analysis of an EA on TRAPZEROS later.

2.2. Algorithm

The $(N + N)$ EA studied in this paper is with equal parent and offspring sizes. The detailed algorithm is described as follows:

1. **Initialization:** The N initial individuals are generated uniformly at random, and the initial population ξ_1 is obtained.
2. **Mutation:** At the tth generation ($t \in \mathbb{N}^+$), the N individuals in the parent population ξ_t are mutated, and the offspring population $\xi_t^{(m)}$ is obtained. The mutation of each individual in ξ_t utilizes the *bitwise mutation*, i.e., each bit of the individual is flipped independently with a uniform probability $1/n$, where n is the problem size.
3. **Selection:** After the mutation step at the tth generation ($t \in \mathbb{N}^+$), the best N individuals in the parent and offspring populations $(\xi_t \cup \xi_t^{(m)})$ are selected to form the population ξ_{t+1}, which is the parent population of the $(t+1)^{th}$ generation. Afterward, set $t = t + 1$ and then go to the mutation step.

The execution of the EA will stop if the stopping criterion is met. The above algorithm adopts the truncation selection, and does not employ any recombination operator. The investigation of EAs with recombination operator and other selection operators will be left as our future work.

With respect to the $(N + N)$ EA, the population size N must be a polynomial function of the problem size n, otherwise each generation of the EA would require super-polynomial number of fitness evaluations. When $N = 1$, the above $(N + N)$ EA degenerates to the well-known $(1 + 1)$ EA [4].

2.3. Solvable rate

So far we have introduced the problem and algorithm investigated in this paper. In this subsection, we present the measure of performance for EA. First, let us review the first hitting time τ, which is defined formally as below:

$$\tau = \min\{t; x^* \in \xi_t\},$$

where x^* is the global optimum, and ξ_t is the population at the tth generation. As the example we have presented in Section I, the first hitting time measure may sometimes become invalid for comparing the performance of two EAs. Here we provide an alternative measure, the solvable rate, to deal with the situation shown in the aforementioned example. Denoted by κ, the solvable rate is formally defined by

$$\kappa = \mathbb{P}\left(\tau \prec Poly(n)\right),$$

where the event $\tau \prec Poly(n)$ means that there exists some polynomial function (of the problem size n) $F(n)$ such that $\tau < F(n)$ holds for any $n > n_0$, and n_0 is a positive constant. Generally speaking, to derive appropriate bounds for the solvable rate, we have to concern the first hitting time τ. In the next section, we introduce the mathematical tools for our further investigations.

3. Analytical approaches

In this section, we present the analytical tools utilized in this paper.

3.1. Probability inequalities

First of all, three well-known probabilistic inequalities are necessary to our later analysis. The inequalities are presented as the following lemmas:

Lemma 1 (*Chernoff Bounds [12,4]*)**.** *Let $X_1, X_2, \ldots, X_k \in \{0, 1\}$ be k independent random variables with a same distribution:*

$$\forall i \neq j : \mathbb{P}[X_i = 1] = \mathbb{P}[X_j = 1],$$

where $i, j \in \{1, \ldots, k\}$. Let X be the sum of those random variables, i.e., $X = \sum_{i=1}^{k} X_i$, then we have

- $\forall 0 < \psi < 1$: $\mathbb{P}\left[X < (1 - \psi)\mathbb{E}[X]\right] < e^{-\mathbb{E}[X]\psi^2/2}$.
- $\forall 0 < \psi \leq 2e - 1$: $\mathbb{P}\left[X > (1 + \psi)\mathbb{E}[X]\right] < e^{-\mathbb{E}[X]\psi^2/4}$.
- $\forall \psi > 0$: $\mathbb{P}\left[X > (1 + \psi)\mathbb{E}[X]\right] < \left[\frac{e^\psi}{(1+\psi)^{1+\psi}}\right]^{\mathbb{E}[X]}$.

Lemma 2 (*Chebyshev Inequality [15]*)**.** *Let X be a random variable with expectation $\mathbb{E}[X]$ and finite variance $Var[X]$. Then for any real number $r > 0$,*

$$\mathbb{P}\left[\left|X - \mathbb{E}[X]\right| \geq r \cdot \sqrt{Var[X]}\right] \leq \frac{1}{r^2}. \tag{2}$$

Lemma 3 (*Markov Inequality [12]*)**.** *Let $X \geq 0$ be a random variable with expectation $\mathbb{E}[X]$. Then for $a > 0$, we have*

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}. \tag{3}$$

3.2. Decomposition of population set

In this subsection, we introduce the concrete definitions and approaches for analyzing the EA on TRAPZEROS, which are inherited from our previous investigation [2]. Recall that we have defined the schemata $\mathcal{S}_1$, $\mathcal{S}_0$ and $\mathcal{S}^*$ in Section 2.1. Further, we denote by E the whole population set containing all populations. Based on the aforementioned definitions of the three schemata, we now present the decomposition of the population set E, which is necessary for our analytical approaches:

- We denote by E_0 the population set consisting of population ξ with its best individual belonging to neither $\mathcal{S}_1$ nor $\mathcal{S}_0$.
- For any population ξ with its best individual belonging to $\mathcal{S}_1$, we define the metric $m^{(A)}(\xi)$

$$m^{(A)}(\xi) = \min\left\{g^{(A)}(y); \text{TRAPZEROS}(y) = \max\left\{\text{TRAPZEROS}(z); z \in \xi\right\}, y \in \mathcal{S}_1, y \in \xi\right\},$$

where $g^{(A)}(y) = n - 2 - \sum_{i=3}^{n} |y_i - 1|$ and $y = (y_1, y_2, \ldots, y_n)$ and z are individuals belonging to ξ. Based on the metric $m^{(A)}(\xi)$, we obtain $n - 1$ population sets, $E_\rho^{(A)}$ ($\rho = 0, \ldots, n - 2$), where

$$E_\rho^{(A)} = \{\xi; m^{(A)}(\xi) = \rho\}, \quad \rho = 0, \ldots, n - 2.$$

The above definition, along with the fitness function TRAPZEROS, implies that $E_{n-2}^{(A)}$ is the population set that is made up of all the population containing the global optimum $x^* = (1, 1, \ldots, 1)$.

Moreover, for any population $\xi \in E_\rho^{(A)}$, we define a subset of ξ:

$$G^{(A)} = \{y; \text{TRAPZEROS}(y) = \max\{\text{TRAPZEROS}(z); z \in \xi\}, y \in \mathscr{S}_1, y \in \xi\}.$$

- For any population ξ with its best individual belonging to $\mathscr{S}_0$, we define the metric $m^{(B)}(\xi)$

$$m^{(B)}(\xi) = \min\left\{g^{(B)}(y); \text{TRAPZEROS}(y) = \max\left\{\text{TRAPZEROS}(z); z \in \xi\right\}, y \in \mathscr{S}_0, y \in \xi\right\},$$

where $g^{(B)}(y) = n - 2 - \sum_{i=3}^n y_i$ and $y = (y_1, y_2, \ldots, y_n)$. Based on the metric $m^{(B)}(\xi)$, we obtain $n - 1$ population sets, $E_\rho^{(B)}$ $(\rho = 0, \ldots, n - 2)$, where

$$E_\rho^{(B)} = \{\xi; m^{(B)}(\xi) = \rho\}, \quad \rho = 0, \ldots, n - 2.$$

Moreover, for any population $\xi \in E_\rho^{(B)}$, we define a subset of ξ:

$$G^{(B)} = \{y; \text{TRAPZEROS}(y) = \max\{\text{TRAPZEROS}(z); z \in \xi\}, y \in \mathscr{S}_0, y \in \xi\}.$$

- According to the above definitions, we have

$$E = E_0 \cup \left(\bigcup_{\rho=0}^{n-2} E_\rho^{(A)}\right) \cup \left(\bigcup_{\rho=0}^{n-2} E_\rho^{(B)}\right).$$

On the basis of the decomposition of E, we introduce the definitions of local optimal individual (type A and type B):

Definition 1 (*Local Optimal Individual, type A (LOIA)*)**.** Given a population set $E_\rho^{(A)}$ $(\rho = 0, \ldots, n - 2)$ and a population $\xi \in E_\rho^{(A)}$, we call an individual $x \in G^{(A)}$ the ρ-LOIA of ξ on $E_\rho^{(A)}$ (LOIA for short, if we restrict the discussion on a given $E_\rho^{(A)}$ and ξ), if and only if $x \in G^{(A)}$; We call an individual x' the advanced LOIA for $\forall \xi \in E_\rho^{(A)}$ (advanced LOIA for short, if we restrict the discussion on a given $E_\rho^{(A)}$), if and only if x' is the LOIA of a population ζ satisfying that $\zeta \in \bigcup_{i=\rho+1}^{n-2} E_i^{(A)}$.

Definition 2 (*Local Optimal Individual, type B (LOIB)*)**.** Given a population set $E_\rho^{(B)}$ $(\rho = 0, \ldots, n - 2)$ and a population $\xi \in E_\rho^{(B)}$, we call an individual $x \in G^{(B)}$ the ρ-LOIB of ξ on $E_\rho^{(B)}$ (LOIB for short, if we restrict the discussion on a given $E_\rho^{(B)}$ and ξ), if and only if $x \in G^{(B)}$. We call an individual x' the advanced LOIB for $\forall \xi \in E_\rho^{(B)}$ (advanced LOIB for short, if we restrict the discussion on a given $E_\rho^{(B)}$), if and only if x' is the LOIB of a population ζ satisfying that $\zeta \in \bigcup_{i=\rho+1}^{n-2} E_i^{(B)}$.

So far we have defined LOIA and LOIB so as to characterize the individuals belonging to the attraction basins of the global and local optima respectively, which enables us to define the so-called takeover times for the $(N + N)$ EA. Briefly, the takeover time, proposed by Goldberg and Deb [6], originally measures the number of generations required by the population to accumulate enough promising individuals, where a repeated selection process is concerned. Chen et al. [2] generalized the concept of "takeover process" to characterize the behavior population-based EAs on unimodal problems containing no local optimum. In this paper, on the TRAPZEROS problem, we further define two types of takeover processes for the global and local optima respectively:

Definition 3 (*(A, ρ, ϵ)-takeover*)**.** A population ξ is said to be (A, ρ, ϵ)-takeover $(\rho = 0, \ldots, n - 2,$ and $0 < \epsilon \le 1)$ if and only if its number of the ρ-LOIAs has reached $\lceil \epsilon N \rceil$ (i.e., the proportion of LOIAs in ξ has reached ϵ), where all advanced LOIAs are pessimistically considered as ρ-LOIAs.

Definition 4 (*(B, ρ, ϵ)-takeover*)**.** A population ξ is said to be (B, ρ, ϵ)-takeover $(\rho = 0, \ldots, n - 2,$ and $0 < \epsilon \le 1)$ if and only if its number of the ρ-LOIBs has reached $\lceil \epsilon N \rceil$ (i.e., the proportion of LOIBs in ξ has reached ϵ), where all advanced LOIBs are pessimistically considered as ρ-LOIBs.

As it is assumed pessimistically in [2], throughout the paper we consider that the advanced LOIAs (LOIBs) cannot be generated, if a population is not (A, ρ, ϵ)-taken over $((B, \rho, \epsilon)$-taken over). Once a population has been (A, ρ, ϵ)-taken over $((B, \rho, \epsilon)$-taken over), it will concentrate on producing advanced LOIAs (LOIBs) for $E_\rho^{(A)}$ ($E_\rho^{(B)}$). Next, we formally define the (A, ρ, ϵ)-takeover time and (B, ρ, ϵ)-takeover time, and then characterize the processes of generating advanced LOIAs and LOIBs by the so-called (A, ρ, ϵ) and (B, ρ, ϵ) upgrade times, respectively.

Given the population $\xi_{t_\rho} \in E_\rho^{(A)}$ at the t_ρ^{th} generation, we define its (A, ρ, ϵ)-first hitting time to $E_\rho^{(A)}$ be

$$\eta_{\rho,\epsilon}^{(A)}(\xi_{t_\rho}) = \min\left\{t'_\rho - t_\rho; \xi_{t'_\rho} \in \left(\bigcup_{i=\lceil \epsilon N \rceil}^N E_{\rho,i}^{(A)}\right) \cup \left(\bigcup_{j=\rho+1}^{n-2} E_j^{(A)}\right)\right\}, \tag{4}$$

where $E_{\rho,i}^{(A)}$ is the population set containing the populations containing i ρ-LOIAs $(i = 1, \ldots, N)$. The expectation of $\eta_{\rho,\epsilon}^{(A)}(\xi_{t_\rho})$, restricted to a finite $\eta_{\rho,\epsilon}^{(A)}(\xi_{t_\rho})$ and conditional on the starting population ξ, is defined by

$$\bar{\eta}_{\rho,\epsilon}^{(A)}(\xi) = \mathbb{E}\left[\eta_{\rho,\epsilon}^{(A)}(\xi_{t_\rho}), \eta_{\rho,\epsilon}^{(A)}(\xi_{t_\rho}) < \infty \mid \xi_{t_\rho} = \xi\right].$$

The expectation $\bar{\eta}_{\rho,\epsilon}^{(A)}(\xi)$ is called the (A, ρ, ϵ)-***takeover time*** of population ξ. Afterward, we define the maximal (A, ρ, ϵ)-takeover time as:

$$\bar{\eta}_{\max,\rho,\epsilon}^{(A)} = \max\left\{\bar{\eta}_{\rho,\epsilon}^{(A)}(\xi); \xi \in E_\rho^{(A)}\right\},$$

and the $(A, \cdot, \epsilon)$-takeover time is defined by

$$\bar{\eta}_\epsilon^{(A)} = \max\left\{\bar{\eta}_{\max,\rho,\epsilon}^{(A)}; 0 \le \rho \le n-2\right\}. \tag{5}$$

Similarly, we can define (B, ρ, ϵ)-takeover time $\bar{\eta}_{\rho,\epsilon}^{(B)}(\xi)$, and the $(B, \cdot, \epsilon)$-takeover time $\bar{\eta}_\epsilon^{(B)}$.

According to the notion presented in [2], the evolution of individuals, if restricted in a specific attraction basin, can be characterized by the repeated "takeover-upgrade processes". Concretely, to reach the optimum in the attraction basin, the $(N+N)$ EA, with both mutation and selection, may often take a number of steps. At each step, the EA may need to accumulate enough promising individuals first, while the qualities of the individuals may not be significantly improved. Afterward, when a considerable amount of promising individuals have been accumulated, the population will take a high probability for generating one or more better individuals. Formally, we further define the so-called (A, ρ, ϵ)- and (B, ρ, ϵ)-upgrade times. Given the population $\xi_{t_\rho} \in \left(\bigcup_{i=\epsilon N}^N E_{\rho,i}^{(A)}\right)$ at the t_ρ^{th} generation, we define its (A, ρ, ϵ)-upgrade time be

$$\phi_{\rho,\epsilon}^{(A)}(\xi_{t_\rho}) = \min\left\{t'_\rho - t_\rho; \xi_{t'_\rho} \in \left(\bigcup_{i=\rho+1}^{n-2} E_i^{(A)}\right), t_\rho < \infty\right\}. \tag{6}$$

Similarly, the (B, ρ, ϵ)-upgrade time $\phi_{\rho,\epsilon}^{(B)}(\xi_{t_\rho})$ can be defined. In the meantime, when a population ξ has already (A, ρ, ϵ)-taken over, the probability of generating at least one advanced LOIA in one generation, denoted by $u_{\rho,\epsilon}^{(A)}$, is given by

$$u_{\rho,\epsilon}^{(A)} = \min\left\{\sum_{\zeta \in \bigcup_{i=\rho+1}^{n-2} E_i^{(A)}} \mathbb{P}(\xi, \zeta); \xi \in E_{\rho,\epsilon N}\right\},$$

where $\mathbb{P}(\xi, \zeta)$ is the one-generation transition probability from population ξ to population ζ. The reciprocal of $u_{\rho,\epsilon}^{(A)}$ is an upper bound of the mean (A, ρ, ϵ)-upgrade time, due to the property of geometric distribution [16]. Similarly, we can define the upgrade probability $u_{\rho,\epsilon}^{(B)}$, and its reciprocal bounds the mean (B, ρ, ϵ)-upgrade time from above.

3.3. Bounding first hitting time

As defined above, both the (A, ρ, ϵ)-takeover time and (B, ρ, ϵ)-takeover time describe the time sufficient for the EA to accumulate enough promising individuals for generating better individuals. Restricted on $\mathscr{S}^*$, the evolution of individuals towards the global optimum can be characterized as the so-called repeated takeover-upgrade process: first, the promising individuals are accumulated by the overall impact of the selection and mutation operators; When there are enough promising individuals in the population, the probability of generating better individuals becomes large enough, thus one or more individuals will soon upgrade to better individuals. The evolution of individuals restricted on $\mathscr{S}_0 \cup (\mathscr{S}_1 \setminus \mathscr{S}^*)$ can be characterized in a similar way, though the evolution will eventually lead to the local optimum (instead of the global optimum) if no individual belonging to $\mathscr{S}^*$ has been generated.

Formally, given $0 \le \rho_1 \le \rho_2 \le n-2$ and generation index t_{ρ_1} satisfying that the population $\xi_{t_{\rho_1}}$ belongs to $E_{\rho_1}^{(A)}$, define the mean first hitting time to $E_{\rho_2}^{(A)}$ of the $(N+N)$ EA starting from $E_{\rho_1}^{(A)}$ be

$$\tau_{\rho_1,\rho_2}^{(A)} = \min\left\{t - t_{\rho_1}; \xi_t \in \left(\bigcup_{i=\rho_2}^{n-2} E_i^{(A)}\right), t_{\rho_1} < \infty\right\}. \tag{7}$$

By drift analysis [7] utilized in the proof of Proposition 1 in [2], we can easily combine the aforementioned repeated takeover-upgrade processes together and obtain the lemmas concerning the mean first hitting time to different population subset:

Lemma 4. *On the* TRAPZEROS *problem, the mean first hitting time to $E_{\rho_2}^{(A)}$ of the $(N+N)$ EA starting from $E_{\rho_1}^{(A)}$ (where $0 \le \rho_1 \le \rho_2 \le \ln^2 n$ holds), conditional on that no individual belonging to $\mathscr{S}_0$ has ever been generated before the first time the EA reaches $E_{\rho_2}^{(A)}$, satisfies:*

$$\mathbb{E}\left[\tau_{\rho_1,\rho_2}^{(A)} \middle| \bigwedge_{t=1}^{\tau_{\rho_1,\rho_2}^{(A)}} (\mathscr{S}_0 \cap \xi_{t+1} = \emptyset)\right] = O\left[\sum_{k=\rho_1+1}^{\rho_2} \left(\bar{\eta}_\epsilon^{(A)} + 1/u_{k,\epsilon}^{(A)}\right)\right], \tag{8}$$

where $\bar{\eta}_\epsilon^{(A)}$ is defined by Eq. (5).

The mean first hitting time to $E_{\rho_2}^{(A)}$ of the $(N+N)$ EA starting from $E_{\rho_1}^{(A)}$ (where $\ln^2 n \leq \rho_1 \leq \rho_2 \leq n-2$ holds) satisfies:

$$\mathbb{E}\left[\tau_{\rho_1,\rho_2}^{(A)}\right] = O\left[\sum_{k=\rho_1+1}^{\rho_2}\left(\bar{\eta}_{\epsilon}^{(A)} + 1/u_{k,\epsilon}^{(A)}\right)\right]; \tag{9}$$

Interested readers can refer to the appendix of this paper for the detailed proof of the above lemma. On the other hand, combining the takeover-upgrade processes for the LOIBs, we have the following lemma:

Lemma 5. *On the* TRAPZEROS *problem, given the population size $N = \Omega(n/\ln n)$, the first hitting time to $E_{\rho_2}^{(B)}$ of the $(N+N)$ EA starting from $E_{\rho_1}^{(B)}$ (where $0 \leq \rho_1 \leq \rho_2 \leq n-2$ holds), conditional on that no individual belonging to $\mathcal{S}^*$ has ever been generated before the first time the EA reaches $E_{\rho_2}^{(B)}$, satisfies the following inequality with an overwhelming probability:*

$$\tau_{\rho_1,\rho_2}^{(B)} \leq (\rho_2-\rho_1)\left(\hat{\eta}_{\epsilon}^{(B)} + \ln^3 n\right), \tag{10}$$

where $\tau_{\rho_1,\rho_2}^{(B)}$ is the first hitting time to $E_{\rho_2}^{(B)}$ starting from $E_{\rho_1}^{(B)}$, $\hat{\eta}_{\epsilon}^{(B)}$ satisfies the following condition: $\forall i \in \{0,\ldots,n-2\}, \forall \xi_{t_i} \in E_i^{(B)} : \mathbb{P}\left(\eta_{i,\epsilon}^{(B)}(\xi_{t_i}) \leq \hat{\eta}_{\epsilon}^{(B)}\right) \succ 1 - 1/SuperPoly(n)$, i.e., it is super-polynomially close to 1.

The detailed proof is given in the appendix.

In this section, we have provided the analytical tools for the theoretical investigations of the $(N+N)$ EA. In the following parts of the paper, we will apply the tools to study the performance of the $(N+N)$ EA with different population sizes, so as to demonstrate the impact of population size on the performance of EA.

4. $(N+N)$ EA with $N=1$

In this section, we analyze the performance of $(1+1)$ EA on TRAPZEROS, where the $(1+1)$ EA can be considered as a degenerate case of the $(N+N)$ EA. By employing the aforementioned solvable rate as a measure, we obtain the following result:

Theorem 6. *The first hitting time τ of the $(1+1)$ EA on* TRAPZEROS *is $O(n^2)$ with the probability of $\frac{1}{4} - O\left(\frac{\ln^2 n}{n}\right)$. In other words, the solvable rate κ of the $(1+1)$ EA on* TRAPZEROS *is at least $\frac{1}{4} - O\left(\frac{\ln^2 n}{n}\right)$.*

Proof. After initialization, with a probability of $1/4$, the first and second bits of the initial individual both take the value 1, i.e., the initial individual belongs to $\mathcal{S}_1$. At any generation, the probability that the EA generates an offspring belonging to $\mathcal{S}_0$ is $1/n^2$. Noting that the fitness of individual belonging to the schemata $\{(1,0,*,\ldots,*)\}$ and $\{(0,1,*,\ldots,*)\}$ is strictly smaller than that of any individual belonging to $\mathcal{S}_1$, the probability of avoiding finding any individual belonging to $\mathcal{S}_0$ at no later than the tth generation is at least $(1 - 1/n^2)^t/4$. Hence, for any $t \leq (2en-1)\ln^2 n$, the above probability is at least $(1 - 1/n^2)^{2en\ln^2 n}/4 > (1 - 4e\ln^2 n/n)/4$.

Given the condition that the EA does not find any individual belonging to $\mathcal{S}_0$ before the $((2en-1)\ln^2 n)^{th}$ generation, we then estimate the time sufficient for the EA to find some solutions belonging to the schema $\mathcal{S}^*$. The reason of concerning the schema $\mathcal{S}^*$ is that, once some individual belonging to $\mathcal{S}^*$ is found, the elitist selection operator will not accept any individual belonging to $\mathcal{S}_0$, which will eventually lead to the global optimum. It is easy to see that, at t^{th} $(t \leq (2en-1)\ln^2 n)$ generation of optimization, the EA takes the probability of no less than $(1/n)(1 - 1/n)^{n-3} \geq 1/(e^2 n)$ to find an individual with better fitness, conditional on the event that the EA does not find any individual belonging to $\mathcal{S}_0$ before the $((2en-1)\ln^2 n)^{th}$ generation. Hence, according to Chernoff bounds, with an extra probability of $1 - e^{-\Theta(\ln^2 n)}$ the EA can find some individual belonging to $\mathcal{S}^*$ no later than the $((2en-1)\ln^2 n)^{th}$ generation.

Once the EA has found a solution belonging to $\mathcal{S}^*$, it will continue to find better solutions and eventually find the global optimum. In this phase, the EA takes the probability of no less than $(1/n)(1 - 1/n)^{n-3} \geq 1/(e^2 n)$ to find an individual with better fitness, and in the worst case the Hamming distance between the individual and global optimum is $n - \ln^2 n - 2$. Hence, by Chernoff bounds, with an extra probability of $1 - e^{-\Theta(n)}$, the EA can find the global optimum with extra $O(n^2)$ generations.

Combining the above propositions together, we have proved that the first hitting time τ of the $(1+1)$ EA on TRAPZEROS is $O(n^2)$ with a probability of $(1/4)(1 - 2e\ln^2 n/n)(1 - e^{-\Theta(\ln^2 n)})(1 - e^{-\Theta(n)}) = 1/4 - O(\ln^2 n/n)$. $\square$

The above theorem demonstrates that, when the population size is extremely small, the $(N+N)$ EA can find the global optimum of TRAPZEROS with a constant probability.

5. $(N+N)$ EA with $N=O(\ln n)$

In this section, the population size of the $(N+N)$ EA grows to $N=O(\ln n)$. The investigation begins with the following lemma concerning the number of LOIAs:

Lemma 7. *Let X_t be the total number of LOIAs at the end of the tth generation. For the $(N + N)$ EA with truncation selection, we have*

$$\mathbb{P}\left[X_{t+1} > (1 + ch)X_t \,\middle|\, X_t < \frac{N}{1 + ch}, \, \mathcal{S}_0 \cap \xi_{t+1} = \emptyset\right] > 1 - \frac{1 - h}{(1 - c)^2 X_t h},$$

where h is the probability that an old LOIA generates a new LOIA, $c \in (0, 1)$ is a constant.

The proof of the above lemma is given in the appendix. Similar to Lemma 7 focusing on LOIAs, the following lemma about LOIBs holds:

Lemma 8. *Let Y_t be the total number of LOIBs at the end of the tth generation. For the $(N + N)$ EA with truncation selection, we have*

$$\mathbb{P}\left[Y_{t+1} > (1 + ch')Y_t \,\middle|\, Y_t < \frac{N}{1 + ch'}, \, \mathcal{S}^* \cap \xi_{t+1} = \emptyset\right] > 1 - \frac{1 - h'}{(1 - c)^2 Y_t h'},$$

where h' is the probability that an old LOIB generates a new LOIB, $c \in (0, 1)$ is a constant.

For the sake of brevity, we do not offer the detailed proof of the above lemma. Interested readers can refer to the proof of Lemma 7 for details.

By Lemma 7, we are able to bound the takeover time from above, which enables us to prove the following result:

Theorem 9. *Given the population size N satisfying $N = O(\ln n)$ and $N = \omega(1)$, the first hitting time τ of the $(N + N)$ EA on* TRAPZEROS *is $O\left(\frac{n^2}{N}\right)$ with a probability of $1/Poly(n)$ [1]. In other words, the solvable rate κ of the $(N + N)$ EA on* TRAPZEROS *is at least $1/Poly(n)$, which is the reciprocal of some polynomial function of the problem size n.*

Proof. The proof requires two major steps which aim at proving the following propositions respectively:

9.1 The probabilities of the following two events are both $1/Poly(n)$: (1) The initial individuals all belong to $\mathcal{S}_1$; (2) The EA does not find any individual belonging to $\mathcal{S}_0$ within $n \ln^3 n/N$ generations.

9.2 The maximal $(A, \rho, 5/(5 + c))$-takeover time ($\rho \in \{1, \ldots, n - 2\}$, $c \in (0, 1)$ is a constant) of the $(N + N)$ EA on TRAPZEROS, conditional on the above two events, is upper bounded.

Given the upper bound of the takeover times obtained when proving Theorem 9.2, we can utilize Lemma 4 to obtain a conditional expected first hitting time of the EA, which immediately leads to the theorem according to 3 (Markov inequality).

Proof of Theorem 9.1

The proof begins with the discussions on Theorem 9.1. After the initialization of the EA, with the probability of $1/4^N$, the first and second bits of all initial individuals all take the value of 1, i.e., the initial individuals all belong to $\mathcal{S}_1$. In the meantime, the probability for an individual (belonging to $\mathcal{S}_1$) to generate an offspring belonging to $\mathcal{S}_0$ is $1/n^2$. On the other hand, since the fitness of every individual belonging to the schemata $\{(1, 0, *, \ldots, *)\}$ or $\{(0, 1, *, \ldots, *)\}$ is strictly smaller than that of any individual belonging to $\mathcal{S}_1$, the truncation selection of the $(N + N)$ EA will always immediately eliminate all the newly generated individual belonging to the schemata $\{(1, 0, *, \ldots, *)\}$ and $\{(0, 1, *, \ldots, *)\}$. As a consequence of the above facts, the probability of avoiding finding any individual belonging to $\mathcal{S}_0$ before the tth generation is at least $(1 - 1/n^2)^{Nt}/4^N$. In other words, we have

$$\mathbb{P}\left[\bigwedge_{t=1}^{n \ln^3 n/N} (\mathcal{S}_0 \cap \xi_t = \emptyset)\right] \geq \frac{\left(1 - \frac{1}{n^2}\right)^{n \ln^3 n}}{4^N} > \frac{1 - \frac{2 \ln^3 n}{n}}{4^N}. \tag{11}$$

So far we have proved Theorem 9.1.

Proof of Theorem 9.2

Concerning Theorem 9.2, we focus on the (A, ρ, ϵ)-takeover time, where $\rho \in \{1, \ldots, n - 2\}$ holds. Here we need to consider two cases according to the value of ρ: (1) $\rho \in \{1, \ldots, \ln^2 n\}$; (2) $\rho \in \{\ln^2 n + 1, \ldots, n - 2\}$.

Let us study the first case first, where $\rho \in \{1, \ldots, \ln^2 n\}$ holds. We note that at every generation the probability that an old LOIA generates a new LOIA is no smaller than $(1 - p_m)^n = (1 - 1/n)^n > 1/5$ ($n \geq 2$), where $p_m = 1/n$ is the mutation rate of the EA and $(1 - p_m)^n$ is the probability that an old LOIA generates an offspring that is the same to itself. Hence, in the $(A, \rho, 5/(5 + c))$-takeover process (where we let $\epsilon = 5/(5 + c)$, and $c \in (0, 1 - \sqrt{\frac{4}{5}})$ is a positive constant) starting with $X_t = 1$ (we consider the population with a unique LOIA at the beginning of the $(A, \rho, 5/(5 + c))$-takeover process in our worst-case analysis), the expected number of generations spent by the EA to accumulate 5 LOIAs is less than $5/(1/5) = 25$ generations.

---

[1] In this paper, $1/Poly(n)$ refers to some positive function (of the problem size n), whose reciprocal is bounded from above by a polynomial function of the problem size n.

On the other hand, for $X_t \in [5, 5N/(5 + c)]$, according to Lemma 7, we further obtain the following inequality

$$\mathbb{P}\left[X_{t+1} > (1 + c/5)X_t \,\middle|\, X_t \in [5, 5N/(5 + c)], \, \mathscr{S}_0 \cap \xi_{t+1} = \emptyset\right]$$

$$> \mathbb{P}\left[X_{t+1} > (1 + ch)X_t \,\middle|\, X_t \in [5, 5N/(5 + c)], \, \mathscr{S}_0 \cap \xi_{t+1} = \emptyset\right] > 1 - \frac{1 - h}{(1 - c)^2 X_t h}$$

$$\geq 1 - \frac{4}{5(1 - c)^2} = \sigma, \tag{12}$$

where $c \in (0, 1 - \sqrt{\frac{4}{5}})$ is a constant, $h \geq (1 - 1/n)^n > 1/5$ is the probability that an old LOIA generates a new LOIA, and $\sigma = 1 - 4/(5(1 - c)^2) < 1$ is a positive constant. Now we can estimate upper bound for the takeover times, conditional on the event

$$\bigwedge_{t=1}^{n \ln^3 n/N} (\mathscr{S}_0 \cap \xi_t = \emptyset). \tag{13}$$

Here we call the event "$X_{t+1} > (1 + c/5)X_t$" a success. Let $\hat{l}$ be the number of successes sufficient for the population to (A, ρ, ϵ)-takeover, which can be estimated by

$$\hat{l} = \min\left\{l; 5 \cdot \left(1 + \frac{c}{5}\right)^l \geq \lceil \epsilon N \rceil\right\}, \tag{14}$$

where $\epsilon = 5/(5 + c)$. As a consequence, we obtain

$$\hat{l} \in \left(\frac{\ln 5 - \ln \lceil \epsilon N \rceil}{\ln \epsilon} - 1, \frac{\ln 5 - \ln \lceil \epsilon N \rceil}{\ln \epsilon} + 1\right).$$

Hence, $\hat{l} = \Theta(\ln N)$ successes are sufficient for $(A, \rho, 5/(5 + c))$-takeover (in addition to an average of 25 generations that are sufficient for accumulating 5 LOIAs). According to Eq. (12), the expected number of generations that is sufficient for $(A, \rho, 5/(5 + c))$-takeover ($\rho \in \{1, \ldots, \ln^2 n\}$), conditional on the event described in Eq. (13), is at most $\hat{l}/\sigma + 25$.

Similar to the case of $\rho \in \{1, \ldots, \ln^2 n\}$, we study the $(A, \rho, 5/(5 + c))$-takeover time for the case of $\rho \in \{\ln^2 n + 1, \ldots, n - 2\}$. The only difference between the two cases is that the former is carried out in the context of the condition that the individuals belonging to $\mathscr{S}_0$ have not been generated at every generation (which is summarized in Eq. (13)) while the latter does not require such a condition. The obtained takeover time is similar: the expected number of generations sufficient for any $(A, \rho, 5/(5 + c))$-takeover ($\rho \in \{\ln^2 n + 1, \ldots, n - 2\}$) is at most $\hat{l}/\sigma + 25$.

In the above analysis, we have analyzed in detail the $(A, \rho, 5/(5 + c))$-takeover times for any $\rho \in \{1, \ldots, n - 2\}$, which completes the proof of Theorem 9.2.

Proof of result

The rest of the proof is to estimate the $(A, \rho, 5/(5 + c))$-upgrade times $1/u^{(A)}_{\rho, 5/(5+c)}$ with respect to the $(A, \rho, 5/(5 + c))$-takeover times, and then follows the technique introduced Lemma 4. Concerning the former point, we still consider two cases as we have done for estimating the $(A, \rho, 5/(5 + c))$-takeover times. For the first case, where $\rho \in \{1, \ldots, \ln^2 n\}$ holds, we estimate the $(A, \rho, 5/(5 + c))$-upgrade time $1/u^{(A)}_{\rho, 5/(5+c)}$ under the condition described in Eq. (13): Once the population of the EA belongs to $E^{(A)}_{\rho, \lceil \epsilon N \rceil} \cup E^{(A)}_{\rho, \lceil \epsilon N \rceil + 1} \cup \ldots \cup E^{(A)}_{\rho, N}$, $u^{(A)}_{\rho, \epsilon}$ can be calculated as follows:

$$u^{(A)}_{\rho, \epsilon} = 1 - \left[1 - \frac{1}{n}\left(1 - \frac{1}{n}\right)^\rho\right]^{\lceil \epsilon N \rceil} \geq 1 - \left(1 - \frac{1}{e^2 n}\right)^{\lceil \epsilon N \rceil} = 1 - \left[\left(1 - \frac{1}{e^2 n}\right)^{e^2 n}\right]^{\frac{\lceil \epsilon N \rceil}{e^2 n}},$$

where $\epsilon = 5/(5 + c)$. Hence we have:

$$1/u^{(A)}_{\rho, \epsilon} \leq \frac{1}{1 - \left[\left(1 - \frac{1}{e^2 n}\right)^{e^2 n}\right]^{\frac{\lceil \epsilon N \rceil}{e^2 n}}} \leq \frac{e^{\frac{\lceil \epsilon N \rceil}{e^2 n}}}{e^{\frac{\lceil \epsilon N \rceil}{e^2 n}} - 1} \leq 1 + \frac{e^2 n}{\lceil \epsilon N \rceil}.$$

Similarly, for the second case $\rho \in \{\ln^2 n + 1, \ldots, n - 2\}$, we can also obtain the corresponding (A, ρ, ϵ)-upgrade time:

$$1/u^{(A)}_{\rho, \epsilon} \leq 1 + \frac{e^2 n}{\lceil \epsilon N \rceil} = O\left(\frac{n}{N}\right).$$

The only difference between the two cases is that the proposition for the first case ($\rho \in \{1, \ldots, \ln^2 n\}$) holds when the condition described in Eq. (13) holds, while the second case ($\rho \in \{\ln^2 n + 1, \ldots, n - 2\}$) does not require such a condition.

As a consequence, by Lemma 4, we can estimate the upper bound of the expected first hitting time (to the population subset $E^{(A)}_{\ln^2 n+1}$) of the EA, conditional on the event described in Eq. (13). Formally, let $\tau_{\mathscr{S}^*}$ be the first hitting time (to the population subset $E^{(A)}_{\ln^2 n+1}$), $\bar{\tau}_{\mathscr{S}^*}$ be the above conditional expectation. According to Lemma 4 derived from the technique presented in [2], the asymptotic order of such a conditional expectation is no larger than $\ln^2 n(\bar{\eta}_\epsilon + \max_{\rho \leq \ln^2 n}\{1/u^{(A)}_{\rho,\epsilon}\}) \leq \ln^2 n\left(\hat{l}/\sigma + 25 + O(n/N)\right) = \ln^2 n\left(O(\ln\ln n) + O(n/N)\right) = O(\ln^2 n\ln\ln n + n\ln^2 n/N)$. According to Lemma 3 (Markov inequality), under the condition described in Eq. (13), with the probability of at least 1/2, the first hitting time to the population subset $E^{(A)}_{\ln^2 n+1}$ (i.e., $\tau_{\mathscr{S}^*}$) is less than $2\bar{\tau}_{\mathscr{S}^*} = \omega(n\ln^3 n/N)$. Combining this result with the proof of Theorem 9.1 (which estimates the probability that the condition described in Eq. (13) holds), we know that with a probability of $(1 - 2\ln^3 n/n) \cdot (1/2)/4^N$, the EA has found an individual belonging to $\mathscr{S}^*$ within $2\bar{\tau}_{\mathscr{S}^*} = O(\ln^2 n\ln\ln n + n\ln^2 n/N)$ generations.

Afterward, following the same technique, we further bound from above the expected first hitting time to $E^{(A)}_{n-2}$ (the population subset consisting of all populations containing the global optimum), with the starting point (initial population) $\xi_1 \in \left(\bigcup_{j=\ln^2 n+1}^{n-2} E^{(A)}_j\right)$. Formally, given the above starting point of the EA, let τ' be the first hitting time to $E^{(A)}_{n-2}$, $\bar{\tau}'$ be the expectation of τ'. According to Lemma 4, the asymptotic order of $\bar{\tau}'$ is no larger than $(n - \ln^2 n - 2)(\bar{\eta}_\epsilon + 1/u^{(A)}_{\rho,\epsilon}) = (n - \ln^2 n - 2)\left(O(\ln\ln n) + O(n/N)\right) = O(n\ln\ln n + n^2/N)$. By Markov inequality, with a probability of at least 1/2, the first hitting time τ' is less than $2\bar{\tau}' = O(n\ln\ln n + n^2/N)$.

Combining the above results together, we obtain that the first hitting time of the $(N+N)$ EA (i.e., $\tau_{\mathscr{S}^*}+\tau'$) is upper bounded by $2\bar{\tau}_{\mathscr{S}^*} + 2\bar{\tau}' = O(n^2/N)$ with a probability of $(1 - 2\ln^3 n/n) \cdot (1/2) \cdot (1/2)/4^N = (1 - 2\ln^3 n/n)/4^{N+1}$. Noting that the population size satisfies $N = O(\ln n)$, the above probability is $1/Poly(n)$. Hence, we have proved the whole theorem. □

6. $(N + N)$ EA with $N = \Omega(n/\ln n)$

In this section, we further increase the population size, and study the performance of the EA on the basis of our previous results. We have the following result:

Theorem 10. *Given the polynomial population size N satisfying $N = \Omega(n/\ln n)$, the first hitting time τ of the $(N + N)$ EA on* TRAPZEROS *is super-polynomial with an overwhelming probability. In other words, the solvable rate κ of the $(N + N)$ EA on* TRAPZEROS *is super-polynomially close to 0.*

Proof. The proof of Theorem 10 contains several steps, in which we focus on different propositions required for proving the whole theorem:

10.1 After initialization, the probability that all the individuals in the population contain no more than $3\ln^2 n/4$ 1-bits among the leading $\ln^2 n + 2$ bits is super-polynomially close to 1.
10.2 With an overwhelming probability, the EA cannot find any individual belonging to $\mathscr{S}^*$ before the ηth generation, where $\eta = \ln N \cdot \ln\ln n$. With an overwhelming probability, there are still at least $\ln^2 n/16$ 0-bits between the $(\ln^2 n/8)^{th}$ and $(\ln^2 n + 3)^{th}$ bits of each individual at the end of the ηth generation.
10.3 No later than the ηth generation, the population will be taken over by the individuals belonging to $\mathscr{S}_0$ with an overwhelming probability.
10.4 After the ηth generations, the probability that an individual belonging to $\mathscr{S}^*$ is generated via direct mutation from an individual belonging to $\mathscr{S}_0$ is super-polynomially close to 0.

Proof of Theorem 10.1
The polynomial population size $N = \Omega(n/\ln n)$ implies that the initial population contains one or more individuals belonging to $\mathscr{S}_0$ with the probability of $1-(1-1/4)^N = 1-(3/4)^{\Omega(n/\ln n)}$, which is an overwhelming probability. Meanwhile, the probability of generating one or more individuals belonging to $\mathscr{S}^*$ is $1 - (1 - (1/2)^{\ln^2 n+2})^N \approx N/2^{\ln^2 n+2}$, which is super-polynomially close to 0 due to that N is a polynomial function of the problem size n. Hence, after initialization the best individual in the population belongs to $\mathscr{S}_0$ with an overwhelming probability. On the other hand, concerning the number of 1-bits among the leading $\ln^2 n + 2$ bits for every initial individual, we apply Chernoff bounds, and obtain that the probability of having no more than $3\ln^2 n/4$ 1-bits is $1 - e^{-\ln^2 n/32}$, which is an overwhelming probability. As a consequence, after initialization, the probability that all the individuals in the population contain no more than $3\ln^2 n/4$ 1-bits among the leading $\ln^2 n + 2$ bits is also an overwhelming one.

Proof of Theorem 10.2
As mentioned, we define $\eta = \ln N \cdot \ln\ln n = O(\ln n \cdot \ln\ln n)$ ($N = \Omega(n/\ln n)$), we now prove that within η generations, the probability of finding individuals belonging to $\mathscr{S}^*$ is super-polynomially close to 0. Denote by $\mathcal{C}$ the event "the first 0-bit among the leading $\ln^2 n + 2$ bits of an individual is flipped, while the 1-bits before the flipped 0-bit (i.e., the leading 1-bits) have not been flipped". Optimistically, we assume that event $\mathcal{C}$ always happens before the ηth generation. Under

the circumstance, with the overwhelming probability $1 - \Theta(1/n^{\ln \ln n})$, at most $\ln \ln n$ 0-bits among the leading $\ln^2 n + 2$ bits of an individual can be flipped simultaneously at each generation. Given a polynomial population size N, the above proposition implies that at every generation, the maximal number of 1-bits among the leading $\ln^2 n + 2$ bits of each individual in the population can increase by at most $\ln \ln n$ with an overwhelming probability. Combining the above fact with the result presented in Theorem 10.1, we know that in order to find an individual belonging to $\mathscr{S}^*$, we need at least $(\ln^2 n + 2 - 3\ln^2 n/4)/(\ln \ln n) > \ln^2 n/(4 \ln \ln n)$ generations with an overwhelming probability, which implies that $\eta = o(\ln^2 n/(4 \ln \ln n))$ generations are not enough for the EA to generate an individual belonging to $\mathscr{S}^*$ with an overwhelming probability.

In the meantime, the fact that (with an overwhelming probability) at most $\ln \ln n$ 0-bits among the leading $\ln^2 n + 2$ bits of an individual can be flipped simultaneously at each generation also implies that (with an overwhelming probability) there are still at least $7 \ln^2 n/8 + 2 - 3 \ln^2 n/4 - O(\eta \cdot (\ln \ln n)) > \ln^2 n/16$ 0-bits between the $(\ln^2 n/8)^{th}$ and $(\ln^2 n + 3)^{th}$ bits of each individual at the end of the ηth generation.

Proof Sketch of Theorem 10.3

Next, we need to prove that within η generations, the population will be taken over by the individuals belonging to $\mathscr{S}_0$ (i.e., $(B, 0, 1)$-takeover) with an overwhelming probability. This proposition can be proved by a technique similar to the one utilized in the proof of Theorem 9, i.e., by applying Chebyshev inequality (Lemma 8) for establishing the probability of "a success". More specifically, let Y_t be the number of LOIBs at the tth generation. For $Y_t \in [5, 5N/(5+c)]$, we define the event "$Y_{t+1} > (1 + c/5)Y_t$" be a "success". According to Lemma 8, we obtain

$$\mathbb{P}\left[Y_{t+1} > (1 + c/5)Y_t \Big| Y_t \in [5, 5N/(5+c)], \mathscr{S}^* \cap \xi_{t+1} = \emptyset\right]$$
$$> \mathbb{P}\left[Y_{t+1} > (1 + ch')Y_t \Big| Y_t \in [5, 5N/(5+c)], \mathscr{S}^* \cap \xi_{t+1} = \emptyset\right] > 1 - \frac{1 - h'}{(1-c)^2 Y_t h'}$$
$$\geq 1 - \frac{4}{5(1-c)^2} = \sigma',$$

where $c \in (0, 1 - \sqrt{\frac{4}{5}})$ is a constant, $h' \geq (1 - 1/n)^n > 1/5$ is the probability that an old LOIB generates a new LOIB, and $\sigma' = 1 - 4/(5(1-c)^2) < 1$ is a positive constant. For the sake of brevity, we omit the mathematical details of discussing the number of successes sufficient for $(B, 0, 1)$-takeover. One can refer to the proof of Theorem 9.2 for more details. Next we prove that the number of generations sufficient for $(B, 0, 1)$-takeover is no larger than η with an overwhelming probability.

Assume that we have already proved that the number of successes sufficient for $(B, 0, 1)$-takeover (using Lemma 8 and the proof idea of Theorem 9.2), denoted by $\hat{l}'$, satisfies $\hat{l}' \leq \ln N/\ln(1/\epsilon) + 3$ (where $\epsilon = 5/(5 + c)$, c is a constant), and the probability of achieving a success at each generation (belonging to the $(B, 0, 1)$-takeover process) is no less than a constant $\sigma' \in (0, 1)$. For the $(B, 0, 1)$-takeover process, let $suc(T)$ be the number of successes happened among T generations. When the number of successes is smaller than $\hat{l}'$ [2], the evolution at each generation can be considered as a Bernoulli trial: the probability of having a success is at least σ'. According to Chernoff bounds, the probability that among T generations the number of successes $suc(T)$ is no smaller than $(1 - \delta) \cdot \mathbb{E}[suc(T)] \geq (1 - \delta) \cdot (T\sigma')$ is at least $1 - e^{-T\sigma'\delta^2/2}$, where $\delta \in (0, 1)$ is a constant. By setting $T = \sqrt{\ln \ln n} \cdot \hat{l}'/((1 - \delta)\sigma') = \Theta(\ln N \cdot \sqrt{\ln \ln n})$, we obtain that, the probability that $T = \Theta(\ln N \cdot \sqrt{\ln \ln n})$ generations are sufficient for obtaining more than $\hat{l}' = \Theta(\ln N)$ successes is at least $1 - e^{-T\sigma'\delta^2/2} = 1 - e^{-\Theta(\ln N \cdot \sqrt{\ln \ln n})\sigma'\delta^2/2}$, which is an overwhelming probability. By summarizing the above discussions, we know that with an overwhelming probability it takes at most $\Theta(\ln N \cdot \sqrt{\ln \ln n})$ generations for the population to be $(B, 0, 1)$-taken over. Since $\eta = \ln N \cdot \ln \ln n = \omega(\ln N \cdot \sqrt{\ln \ln n})$, we have proved Theorem 10.3.

Here, it is worth noting that the above proposition, along with the truncation selection that kills the worst N individuals among the overall $2N$ parents and offsprings, is getting very close to the final conclusion of the theorem. More precisely, once the population has been $(B, \rho, 1)$-taken over by the individuals belonging to $\mathscr{S}_0$ (for any $\rho = 0, \ldots, n - 2$), the truncation selection operator will no longer accept offsprings that belonging to neither $\mathscr{S}^*$ nor $\mathscr{S}_0$, since these offsprings have lower fitness than all N parents belonging to $\mathscr{S}_0$. Hence, the only way to generate individuals belonging to $\mathscr{S}^*$ is via the direct mutations of those parent individuals belonging to $\mathscr{S}_0$. Next, we show that such a probability is super-polynomially small.

Proof of Theorem 10.4

Optimistically, here we can assume that the bits between the $(\ln^2 n/8)^{th}$ and $(\ln^2 n + 3)^{th}$ bits of each individual have become "free-riders" [4] when the population has just been $(B, \rho, 1)$-taken over (where $\rho \in \{0, \ldots, \ln^2 n/8 - 3\}$), i.e., the values of these bits are not influenced by selection pressure, and only genetic drift is considered (though some of them are

---

[2] When the number of successes $suc(T)$ reaches $\hat{l}'$, the population has been $(B, 0, 1)$-taken over. Afterward, if no individual belonging to $\mathscr{S}^*$ has been generated (the corresponding duration has been investigated by Theorem 10.2), the event "success" still makes sense except that the population has entered the $(B, \rho, 1)$-takeover process, where $\rho > 0$.

very likely to be influenced by selection pressure which tends to preserve 0-bits). By dividing the evolutions after the ηth generation into two stages, we need to prove the following propositions:

10.4.1 No later than the $(\eta + \alpha)^{th}$ generation, the population has been $(B, \rho, 1)$-taken over by individuals with no less than $\ln^2 n/8$ consecutive leading 0-bits, where $\alpha = \ln^5 n$.

10.4.2 After the ηth generation but before the population has been taken over by individuals with no less than $\ln^2 n/8$ consecutive leading 0-bits[3], with an overwhelming probability at least $\ln^2 n/32$ free-riders between the $(\ln^2 n/8)^{th}$ and $(\ln^2 n + 3)^{th}$ bits of each individual (in the population) will take the value of 0.

By Lemma 5, we can study the (B, ρ, ϵ)-takeover processes ($\rho \leq \ln^2 n/8 - 2$) and the corresponding (B, ρ, ϵ)-upgrade probability, and prove that the number of generations sufficient for the population to be $(B, \ln^2 n/8 - 2, 1)$-taken over by individuals with more than $\ln^2 n/8$ consecutive leading 0-bits is bounded from above by $(\ln^2 n/8)(\hat{\eta}_\epsilon^{(B)} + \max_{\rho \leq \ln^2 n/8-2}\{1/u_{\rho,1}^{(B)}\}) < (\ln^2 n/8)(\hat{\eta}_\epsilon^{(B)} + \ln^3 n)$ with an overwhelming probability, where $\hat{\eta}_\epsilon^{(B)} = o(\ln N \cdot \ln \ln n) = o(\ln n \cdot \ln \ln n)$. The proof follows the proof idea of Theorem 10.3 to study the takeover time $\hat{\eta}^{(B)}$, and then applies Lemma 5 directly. For the sake of brevity, here we do not provide the details. Moreover, the above result, along with the condition $N = \Omega(n/\ln n)$, implies that $\eta + (\ln^2 n/8)(\hat{\eta}_\epsilon^{(B)} + \ln^3 n) < \eta + \ln^5 n = \eta + \alpha$. Hence, we can reach Theorem 10.4.1.

On the other hand, recall that when the population has just been taken over by the individuals belonging to $\mathcal{S}_0$ (i.e., $(B, \ln^2 n/8 - 2, 1)$-takeover), with an overwhelming probability there are still at least $\ln^2 n/16$ free-riders, between the $(\ln^2 n/8)^{th}$ and $(\ln^2 n + 3)^{th}$ bits of each individual, taking the value of 0. Within $\ln^5 n$ generations, each of the free-riders will receive at most $\alpha = \ln^5 n$ mutations. Given any individual at the ηth generation, there are at least $\ln^2 n/16$ free-riders (between its $(\ln^2 n/8)^{th}$ and $(\ln^2 n+3)^{th}$ bits) taking the value of 0. For each of those free-riders, the probability that its value does not change within the α mutations is at least $(1 - 1/n)^\alpha \approx 1 - O(\ln^5 n/n)$. According to Chernoff bounds, among the aforementioned $\ln^2 n/16$ free-riders of each individual, with the overwhelming probability of $1 - e^{-\Theta(\ln^2 n)} = 1 - n^{-\Theta(\ln n)}$, there are still $\ln^2 n/32$ free-riders consistently taking the value of 0 between the ηth generation and the $(\eta + \alpha)^{th}$ generation. As a consequence, the fact that the population size N is a polynomial function of the problem size n yields Theorem 10.4.2.

By summarizing the above discussions on the number of free-riders, we know that before the population has been taken over by individuals with no less than $\ln^2 n/8$ consecutive leading 0-bits (but after the ηth generation), the probability for each individual to find an offspring belonging to $\mathcal{S}^*$ is at most $1/n^{\ln^2 n/32}$. On the other hand, after the population has been taken over by individuals with no less than $\ln^2 n/8$ consecutive leading 0-bits, the probability for each individual to find an offspring belonging to $\mathcal{S}^*$ is $1/n^{\ln^2 n/8}$, since at that time each individual in the population will contain at least $\ln^2 n/8$ consecutive leading 0-bits. Further, since the population size N is polynomial, the probability that some individual in the population generates an offspring belonging to $\mathcal{S}^*$ is still super-polynomially close to 0.

Combining the results presented in Theorems 10.1, 10.2, 10.3 and 10.4 together, we have proved the theorem. □

7. Discussion

So far we have seen three analytical results concerning the performance of a population-based EA with different population sizes. It is shown that with the increase of population size, the solvable rate of the $(N + N)$ EA will drop to an extremely low level on the TRAPZEROS problem. Although the study in this paper is carried out on a specific problem using a specific type of EAs, it has much wider implications. The analytical results presented in this paper actually demonstrate an interesting problem characteristic under which the population-based EAs may perform poorly: when a problem has an attraction basin leading to some local optimum, and the individuals at this basin are with relatively high fitness than most individuals, a large population may not be useful and even becomes harmful, since it will lead to a large probability of finding individuals at the local basin. The resultant takeover process at the mistaken basin will quickly eliminates other promising individuals that may lead to the global optimum. After that, only large step sizes can help to find promising individuals again, resulting in a long runtime of the EA due to the small probability of getting close to the global optimum.

The weakness of the population-based EAs without recombination on the above problem characteristic, shown in this paper, can partially be tackled by employing larger step sizes. For example, if some appropriate recombination strategy, which can generate large step sizes, is employed, the EA can probably provide much larger step sizes in comparison with bitwise mutation (adopting the commonly used mutation rate $1/n$). Moreover, some adaptive/self-adaptive mutation schemes might also be helpful to cope with the situation, since they can provide large step sizes in exploring the correct attraction basin, and small large sizes in exploiting the correct basin. As a consequence, even if the whole population has been trapped in a mistaken basin, it is still possible to find promising individuals in other basins of attraction. The related investigations will be left as our future work.

8. Conclusion

In this paper, we have investigated the performance of an $(N + N)$ EA with different population sizes on a multimodal problem, namely TRAPZEROS. The theoretical results have revealed a problem characteristic that may lead to poor

---

[3] $(B, \ln^2 n/8 - 2, 1)$-takeover, happened no later than the $(\eta + \alpha)^{th}$ generation with an overwhelming probability.

performances of population-based EAs, as mentioned in the last section. This is the first time that the influence of population size on an $(N+N)$ EA is analyzed theoretically. In addition, the proposed solvable rate, which is an intrinsic feature extracted from the probability distribution of the first hitting time, offers an alternative choice for measuring the performance of EA.

Deriving from a recently developed approach for analyzing EAs on unimodal problems [2] and following the well-known building block hypothesis, the utilized takeover-upgrade technique is capable of characterizing the evolution within a single basin of attraction as repeated takeover-upgrade processes that accumulate enough promising individuals and then generate better individuals by the accumulated individuals. In this paper, the successful application of this technique in modeling population-based EAs on a multimodal problem has shed some light on analyzing more complicated population-based EAs using similar techniques. To utilize such techniques, an elaborate procedure, as shown in the proof of Theorem 9, is required to estimate the takeover times, which is directly related to the selection operator. In the future, we will further study the EAs with different parent and offspring sizes, i.e., the $(\mu + \lambda)$ EAs. Moreover, we will combine the techniques with other state-of-the-art analytical tools, so as to gain more insight into the impact of other operators (e.g., recombination) and the corresponding parameter settings on the performance of EA.

Acknowledgement

Appendix. Drift analysis

Drift analysis is a well-known technique for studying the time complexity of EAs [7,9]. Formally, let x^* be the unique optimum of the fitness function, let $V(X)$ be the function that measures the distance between the population X and the optimum x^*, then the one step mean drift at the tth generation of the EA, denoted by $\Delta(X, t)$, is given by

$$\Delta(X, t) = \mathbb{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = X]$$
$$= \sum_{Y \in E} [V(X) - V(Y)] \mathbb{P}(X, Y; t),$$

where E is the whole population set. The transition probability is not time dependent, if no self-adaptive strategy is utilized in the EA:

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

Under this circumstance, we simply use the notation $\Delta(X)$ to represent the one step mean drift. Meanwhile, if

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

holds for $t = 0, 1, \ldots$, then $\{V(\xi_t) : t = 0, 1, \ldots\}$ is called a super-martingale. According to He and Yao [7,9], 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 by the following lemma:

Lemma 11 (*Drift Theorem [7]*)**.** *Let $\{V(\xi_t) : t = 0, 1, \ldots\}$ be a super-martingale describing an EA, if for any time $t = 1, 2, \ldots,$ if $V(\xi_t) > 0$ and*

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

then the mean first hitting time satisfies

$$\mathbb{E}[\tau \mid \xi_1] \leq \frac{V(\xi_1)}{c_l},$$

where ξ_1 is the initial population of the EA.

One of the key steps when applying drift analysis is to specify an appropriate distance function $V(\cdot)$. The following lemma tells us that when the expected first hitting time of a homogeneous absorbing Markov chain, conditional on the starting state X, is defined as the distance function for the population state X, the one step mean drift equals 1:

Lemma 12 (*[9]*)**.** *Let $L : \{L_t, t = 0, 1, \ldots\}$ be a homogeneous absorbing Markov chain defined on the space M, $H \subset M$ be its absorbing subspace. For L, its first hitting time to H, which is formally defined by*

$$\tau = \min\{t \geq 0; L_t \in H\},$$

satisfies

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

where $\mathbb{E}[\tau|L_0 = X]$ is the expected first hitting time of the absorbing Markov chain L starting with the initial state $L_0 = X$.

The proof of Lemma 4 will utilize the above lemmas.

Proof of Lemma 4

We offer the detailed proof for Eq. (8) in Lemma 4. Before bounding the mean of the first hitting time $\tau_{\rho_1,\rho_2}^{(A)}$, we have to study the (A, k, ϵ)-takeover process $(\rho_1 \leq k \leq \rho_2 \leq \ln^2 n)$ by restricting the Markov chain on $E_k^{(A)}$ under the condition "*no individual belonging to $\mathcal{S}_0$ has ever been generated before the first time the EA reaches $E_{\rho_2}^{(A)}$*". Formally, the above condition is represented by

$$\bigwedge_{t=1}^{\tau_{\rho_1,\rho_2}^{(A)}} (\mathcal{S}_0 \cap \xi_{t+1} = \emptyset).$$

The rest of the proof will be presented in the context of the above condition (for the sake of brevity we omit it when presenting mathematical details).

The analysis follows a pessimistic style, which is quite similar to the proof of Proposition 1 in [2]. If the number of the LOIAs of a population Z is smaller than ϵN, we ignore all potential emergences of advanced LOIAs. Once some advanced LOIA is generated before the number of the LOIAs reaches ϵN, we assume pessimistically that each advanced LOIA is only a LOIA, i.e., it is replaced by some LOIA that is randomly selected from the LOIAs in the current population. In response to this step, the population Z is transformed to Y ($Y \in E_{k,i}^{(A)}$). Afterward, according to the number of LOIAs, we consider two different cases: $\lceil \epsilon N \rceil \leq i \leq N$ and $1 \leq i < \lceil \epsilon N \rceil$, and use the notation "$\to_k$" to represent the mapping from $\bigcup_{j=k+1}^{n-2} E_j^{(A)}$ to $E_k^{(A)}$:

- If $\lceil \epsilon N \rceil \leq i \leq N$, then $Z \to_k Y$. (an advanced LOIA is mapped to a LOIA.)
- If $1 \leq i < \lceil \epsilon N \rceil$, then $Z \to_k Y'$, where Y' is obtained by replacing the best $\lceil \epsilon N \rceil$ individuals of Y with $\lceil \epsilon N \rceil$ randomly selected LOIAs, $Y' \in E_{k,\lceil \epsilon N \rceil}^{(A)}$ (advanced LOIAs are ignored, and the population is considered to be (A, k, ϵ)-taken over by $\lceil \epsilon N \rceil$ LOIAs).

The aim of the above transformations is to restrict the whole (A, k, ϵ)-takeover process on the subspace $E_k^{(A)}$. The consequence is that we can utilize an auxiliary homogeneous absorbing Markov chain $(\zeta_t^{(k)}, t = 0, 1, \ldots)$ on $E_k^{(A)}$ to study the whole (A, k, ϵ)-takeover process. The transition probabilities of the auxiliary Markov chain are given by

$$\bar{\mathbb{P}}(X, Y) = \begin{cases} \mathbb{P}(X, Y), & X \neq Y, \quad X \notin \bigcup_{i=\lceil \epsilon N \rceil}^{N} E_{k,i}^{(A)}, Y \notin \bigcup_{i=\lceil \epsilon N \rceil}^{N} E_{k,i}^{(A)}; \\ \mathbb{P}(X, Y) + \sum_{Z \in \bigcup_{j=k+1}^{n-2} E_j^{(A)}, Z \to_k Y} \mathbb{P}(X, Z), & X \neq Y, \quad X \notin \bigcup_{i=\lceil \epsilon N \rceil}^{N} E_{k,i}^{(A)}, Y \in \bigcup_{i=\lceil \epsilon N \rceil}^{N} E_{k,i}^{(A)}; \\ 0, & X \neq Y, \quad X \in \bigcup_{i=\lceil \epsilon N \rceil}^{N} E_{k,i}^{(A)}, Y \in \bigcup_{i=\lceil \epsilon N \rceil}^{N} E_{k,i}^{(A)}; \\ 0, & X \neq Y, \quad X \in \bigcup_{i=\lceil \epsilon N \rceil}^{N} E_{k,i}^{(A)}, Y \notin \bigcup_{i=\lceil \epsilon N \rceil}^{N} E_{k,i}^{(A)}; \\ 1 - \sum_{Y \neq X} \bar{\mathbb{P}}(X, Y), & X = Y, \end{cases} \tag{15}$$

where $X, Y \in E_k^{(A)}$. According to the definitions of transition probabilities presented in Eq. (15), we have

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

Obviously, $\left(\bigcup_{i=\lceil \epsilon N \rceil}^{N} E_{k,i}^{(A)} \right) \cup \left(\bigcup_{j=k+1}^{n-2} E_j^{(A)} \right)$ is absorbing in $\zeta^{(k)}$. The other subspaces are $E_{k,1}^{(A)}, \ldots, E_{k,\lceil \epsilon N \rceil - 1}^{(A)}$. According to Lemma 12, we know that for $X \in \bigcup_{i=1}^{\lceil \epsilon N \rceil - 1} E_{k,i}^{(A)}$,

$$\sum_{Y \in E_k^{(A)}} \left[\bar{\eta}_{k,\epsilon}^{(A)}(X) - \bar{\eta}_{k,\epsilon}^{(A)}(Y) \right] \bar{\mathbb{P}}(X, Y) = 1. \tag{16}$$

On the other hand, let ξ_{t_k} be the population of the t_k^{th} generation. For $\xi_{t_k} \in E_k^{(A)}$, we define its first hitting time to the population set $\bigcup_{j=k+1}^{n-2} E_j^{(A)}$, starting from ξ_{t_k}:

$$\tau_{k,k+1}^{(A)} = \min\left\{ t_k' - t_k \geq 0; \xi_{t_k'} \in \bigcup_{j=k+1}^{n-2} E_j^{(A)}, \xi_{t_k} \in E_k^{(A)} \right\}.$$

The mean first hitting time (to the set $\bigcup_{j=k+1}^{n-2} E_j^{(A)}$) of the EA starting with the population $\xi_{t_k} = X$, denoted by $\tau_{k,k+1}^{(A)}$, is given by

$$\mathbb{E}\left[\tau_{k,k+1}^{(A)} \mid \xi_{t_k} = X \right] = \sum_{t=0}^{\infty} t \mathbb{P}\left[\tau_{k,k+1}^{(A)} = t \mid \xi_{t_k} = X \right].$$

Let $\mu_{t_k}(\cdot)$ specify the probability distribution of the population at the t_k^{th} generation (note that we have assumed that $\xi_{t_k} \in E_k^{(A)}$), then

$$\mathbb{E}\left[\tau_{k,k+1}^{(A)} \right] = \sum_{Y \in E_k^{(A)}} \mu_{t_k}(Y) \mathbb{E}\left[\tau_{k,k+1}^{(A)} \mid \xi_{t_k} = Y \right]$$

is called the mean first hitting time to population set $\bigcup_{j=k+1}^{n-2} E_j^{(A)}$. Now we utilize drift analysis to bound $\mathbb{E}\left[\tau_{k,k+1}^{(A)} \right]$. First, we define a distance function $V^{(k)}(X)$ for $X \in E_k^{(A)} \cup \left(\bigcup_{j=k+1}^{n-2} E_j^{(A)} \right)$ ($\rho_1 \leq k \leq \rho_2 \leq \ln^2 n$):

$$V^{(k)}(X) = \begin{cases} 0, & \forall X \in \bigcup_{j=k+1}^{n-2} E_j^{(A)}; \\ 1/u_{k,\epsilon}^{(A)} + \bar{\eta}_{k,\epsilon}^{(A)}(X), & \forall X \in E_k^{(A)}. \end{cases}$$

For each population set $E_k^{(A)}$ ($k = 1, \ldots, n$), we show that the one step mean drift of the populations is no less than some positive constant. Given a population $X \in E_k^{(A)}$, let $\Delta(X)$ be its one step mean drift. The estimation of one step mean drift involves two different cases (Eq. (17) and Eq. (19)).

For $X \in \left(\bigcup_{i=\lceil \epsilon N \rceil}^{N} E_{k,i}^{(A)} \right), Y \in E_k^{(A)} \cup \left(\bigcup_{j=k+1}^{n-2} E_j^{(A)} \right)$,

$$\Delta(X) = \sum_{Y \in \bigcup_{j=k+1}^{n-2} E_j^{(A)}} \left[V^{(k)}(X) - V^{(k)}(Y) \right] \mathbb{P}(X, Y) + \sum_{Y \in E_k^{(A)}} \left[V^{(k)}(X) - V^{(k)}(Y) \right] \mathbb{P}(X, Y)$$

$$\geq 1 - \sum_{Y \in \bigcup_{i=1}^{\lceil \epsilon N \rceil - 1} E_{k,i}^{(A)}} \bar{\eta}_{k,\epsilon}^{(A)}(Y) \mathbb{P}(X, Y) \tag{17}$$

$$\geq 1 - \max_{Y \in \bigcup_{i=1}^{\lceil \epsilon N \rceil - 1} E_{k,i}^{(A)}} \{\mathbb{P}(X, Y)\} \cdot \sum_{Y \in \bigcup_{i=1}^{\lceil \epsilon N \rceil - 1} E_{k,i}^{(A)}} \bar{\eta}_{k,\epsilon}^{(A)}(Y) = 1 \tag{18}$$

holds since the truncation selection operator always preserves the best N individuals among the $2N$ individuals in the parent and offspring populations of each generation.

For $X \in \left(\bigcup_{i=1}^{\lceil \epsilon N \rceil - 1} E_{k,i}^{(A)} \right), Y \in E_k^{(A)} \cup \left(\bigcup_{j=k+1}^{n-2} E_j^{(A)} \right)$,

$$\Delta(X) = \sum_{Y \in \bigcup_{j=k+1}^{n-2} E_j^{(A)}} \left[V^{(k)}(X) - V^{(k)}(Y) \right] \mathbb{P}(X, Y) + \sum_{Y \in E_k^{(A)}} \left[V^{(k)}(X) - V^{(k)}(Y) \right] \mathbb{P}(X, Y)$$

$$\geq \left(\sum_{Y \in \bigcup_{j=k+1}^{n-2} E_j^{(A)}} + \sum_{Y \in \bigcup_{i=\lceil \epsilon N \rceil}^{N} E_{k,i}^{(A)}} \right) \bar{\eta}_{k,\epsilon}^{(A)}(X) \mathbb{P}(X, Y) + \sum_{Y \in \bigcup_{i=1}^{\lceil \epsilon N \rceil - 1} E_{k,i}^{(A)}} \left[\bar{\eta}_{k,\epsilon}^{(A)}(X) - \bar{\eta}_{k,\epsilon}^{(A)}(Y) \right] \mathbb{P}(X, Y)$$

$$= \sum_{Y \in \bigcup_{i=\lceil \epsilon N \rceil}^{N} E_{k,i}^{(A)}} \bar{\eta}_{k,\epsilon}^{(A)}(X) \bar{\mathbb{P}}(X, Y) + \sum_{Y \in \bigcup_{i=1}^{\lceil \epsilon N \rceil - 1} E_{k,i}^{(A)}} \left[\bar{\eta}_{k,\epsilon}^{(A)}(X) - \bar{\eta}_{k,\epsilon}^{(A)}(Y) \right] \mathbb{P}(X, Y)$$

$$= \sum_{Y \in E_k^{(A)}} \left[\bar{\eta}_{k,\epsilon}^{(A)}(X) - \bar{\eta}_{k,\epsilon}^{(A)}(Y) \right] \bar{\mathbb{P}}(X, Y) = 1, \tag{19}$$

where Eq. (19) holds because of Eq. (16). The results presented in Eq. (17) and Eq. (19) show that the one step mean drift of X ($X \in E_k^{(A)}$) is always no less than 1. According to Lemma 11 and Eq. (5), we further have

$$\mathbb{E}\left[\tau_{k,k+1}^{(A)} \mid \xi_{t_k} = X \right] = 1/u_{k,\epsilon}^{(A)} + \bar{\eta}_{k,\epsilon}^{(A)}(X) = O\left(1/u_{k,\epsilon}^{(A)} + \bar{\eta}_{\epsilon}^{(A)} \right).$$

Then the first hitting time from $E_k^{(A)}$ to E_k^- satisfies:

$$\mathbb{E}\left[\tau_{k,k+1}^{(A)}\right] = \sum_{X \in E_k^{(A)}} \mu_{t_k}(X)\mathbb{E}\left[\tau_{k,k+1}^{(A)} \mid \xi_{t_k} = X\right] = O\left(1/u_{k,\epsilon}^{(A)} + \bar{\eta}_\epsilon^{(A)}\right) \sum_{X \in E_k^{(A)}} \mu_{t_k}(X)$$

$$= O\left(1/u_{k,\epsilon}^{(A)} + \bar{\eta}_\epsilon^{(A)}\right).$$

For $X \in E_{k'}^{(A)}$ ($k' \neq k$, $\rho_1 \leq k' \leq \rho_2 \leq \ln^2 n$), we can obtain the same upper bound by the same techniques. In the meantime, the truncation selection operator of the EA always preserves the best individual in every generation (ESS), thus once the population of the EA has reached $\bigcup_{j=k+1}^{n-2} E_j^{(A)}$, it will never return to $E_k^{(A)}$ again. Hence, combining the conditional means of $\tau_{k,k+1}^{(A)}$ with respect to different k, we have proved the first part of Lemma 4:

$$\mathbb{E}\left[\tau_{\rho_1,\rho_2}^{(A)} \,\middle|\, \bigwedge_{t=1}^{\tau_{\rho_1,\rho_2}^{(A)}}(\mathscr{S}_0 \cap \xi_{t+1} = \emptyset)\right] = O\left[\sum_{k=\rho_1+1}^{\rho_2}\left(\bar{\eta}_\epsilon^{(A)} + 1/u_{k,\epsilon}^{(A)}\right)\right];$$

The second part of Lemma 4 (Eq. (8)) can be proved in a similar way, except that the condition "no individual belonging to $\mathscr{S}_0$ has ever been generated before the first time the EA reaches $E_{\rho_2}^{(A)}$" is no longer required. □

Proof of Lemma 5

Proof. The idea of the proof is straightforward: in the worst case, the $(N + N)$ EA should spend $\rho_2 - \rho_1$ repeated takeover-upgrade processes in which no individual belonging to $\mathscr{S}^*$ has ever been generated before the first time the EA reaches $E_{\rho_2}^{(B)}$. Instead of considering the mean of the first hitting time from $E_{\rho_1}^{(B)}$ to $E_{\rho_2}^{(B)}$, we consider the upper bound of the first hitting time which holds with an overwhelming probability, i.e., a probability that is super-polynomially close to 1. Since the truncation selection is adopted by the EA, such an upper bound can be obtained by combining directly the upper bounds of the takeover and upgrade times together. As it is defined, with an overwhelming probability, $\hat{\eta}_\epsilon^{(B)}$ is an upper bound of the (B, i, ϵ)-takeover time $\eta_{i,\epsilon}^{(B)}(\xi_{t_i})$ for any population $\xi_{t_i} \in E_i^{(B)}$, where $i \in \{0, \ldots, n-2\}$ holds.

In addition, to prove the lemma, we have to prove that $\ln^3 n$ is an upper bound of the upgrade time with an overwhelming probability. Linking the above proposition to a variable X obeying the geometric distribution with parameter $u_{i,\epsilon}^{(B)}$, the probability that X is bounded from above by $\ln^3 n$ can be calculated as follows:

$$\mathbb{P}\left[X \leq \ln^3 n\right] = \sum_{k=1}^{\ln^3 n}\mathbb{P}[X = k] = 1 - \sum_{\ln^3 n+1}^{+\infty}\mathbb{P}[X = k]$$

$$= 1 - \sum_{k=\ln^3 n+1}^{+\infty}\left(1 - u_{i,\epsilon}^{(B)}\right)^{k-1}u_{i,\epsilon}^{(B)} = 1 - \left(1 - u_{i,\epsilon}^{(B)}\right)^{\ln^3 n}. \tag{20}$$

Concerning the upgrade probability $u_{i,\epsilon}^{(B)}$, we can estimate its lower bound under the condition $N = \Omega(n/\ln n)$. Given the population $\xi_{t_\rho} \in \left(\cup_{i=\epsilon N}^N E_{\rho,i}^{(A)}\right)$ containing at least ϵN LOIBs, $u_{i,\epsilon}^{(B)}$ is bounded from below:

$$u_{i,\epsilon}^{(B)} \geq 1 - \left[1 - \frac{1}{n}\left(1 - \frac{1}{n}\right)^{n-1}\right]^{\epsilon N} \geq 1 - \left(1 - \frac{1}{e^2 n}\right)^{\epsilon N} \geq 1 - e^{-\frac{\epsilon N}{e^2 n}}.$$

Noting that the population size N satisfies $N = \Omega(n/\ln n)$, the above inequality implies

$$u_{i,\epsilon}^{(B)} \geq 1 - e^{-\frac{\epsilon N}{e^2 n}} = 1 - e^{-\Omega(1/\ln n)} = \Omega\left(\frac{1}{\ln n}\right).$$

By inserting the above inequality into Eq. (20), we obtain

$$\mathbb{P}\left[X \leq \ln^3 n\right] = 1 - \left(1 - u_{i,\epsilon}^{(B)}\right)^{\ln^3 n} = 1 - \left[1 - \Omega\left(\frac{1}{\ln n}\right)\right]^{\ln^3 n} \geq 1 - e^{-\Omega(\ln^2 n)}, \tag{21}$$

which is super-polynomially close to 1. Given the population $\xi_{t_i} \in \left(\cup_{k=\epsilon N}^N E_{i,k}^{(A)}\right)$ at the t_i^{th} generation ($\rho_1 \leq i \leq \rho_2$), the (B, i, ϵ)-upgrade time be $\phi_{i,\epsilon}^{(B)}(\xi_{t_\rho})$ obeys the geometric distribution with parameter $u_{i,\epsilon}^{(B)}$. Hence, by replacing the variable X with $\phi_{i,\epsilon}^{(B)}(\xi_{t_i})$ in Eq. (21), we know that $\phi_{i,\epsilon}^{(B)}(\xi_{t_\rho})$ is bounded from above by $\ln^3 n$ with an overwhelming probability. Noting

that $\forall i \in \{0, \ldots, n-2\}, \forall \xi_{t_i} \in E_i^{(B)} : \mathbb{P}\left(\eta_{i,\epsilon}^{(B)}(\xi_{t_i}) \leq \hat{\eta}_\epsilon^{(B)}\right) \succ 1 - 1/SuperPoly(n)$ (condition of Lemma 5), following the proof idea presented at the beginning of the proof, we have proved that the upper bound of the first hitting time, shown in Eq. (10), holds with an overwhelming probability, i.e., a probability that is super-polynomially close to 1. □

Proof of Lemma 7

Let Z_{t+1} be the number of generated LOIAs at the $(t+1)^{th}$ generation. Given that $c \in (0, 1)$ is a constant, for $X_t < N/(1+ch)$ we have

$$\mathbb{P}\left[Z_{t+1} > c \cdot \mathbb{E}[Z_{t+1}|X_t, \mathscr{S}_0 \cap \xi_{t+1} = \emptyset]\Big| X_t < \frac{N}{1+ch}\right]$$

$$= \mathbb{P}\left[\mathbb{E}[Z_{t+1}|X_t, \mathscr{S}_0 \cap \xi_{t+1} = \emptyset] - Z_{t+1} < (1-c) \cdot \mathbb{E}[Z_{t+1}|X_t, \mathscr{S}_0 \cap \xi_{t+1} = \emptyset]\Big| X_t < \frac{N}{1+ch}\right]$$

$$> \mathbb{P}\left[\left|\mathbb{E}[Z_{t+1}|X_t, \mathscr{S}_0 \cap \xi_{t+1} = \emptyset] - Z_{t+1}\right| < (1-c) \cdot \mathbb{E}[Z_{t+1}|X_t, \mathscr{S}_0 \cap \xi_{t+1} = \emptyset]\Big| X_t < \frac{N}{1+ch}\right]$$

$$= 1 - \mathbb{P}\left[\left|Z_{t+1} - \mathbb{E}[Z_{t+1}|X_t, \mathscr{S}_0 \cap \xi_{t+1} = \emptyset]\right| \geq (1-c) \cdot \mathbb{E}[Z_{t+1}|X_t, \mathscr{S}_0 \cap \xi_{t+1} = \emptyset]\Big| X_t < \frac{N}{1+ch}\right]$$

$$\geq 1 - \frac{\text{Var}[Z_{t+1}|X_t < \frac{N}{1+ch}, \mathscr{S}_0 \cap \xi_{t+1} = \emptyset]}{(1-c)^2 \mathbb{E}^2[Z_{t+1}|X_t < \frac{N}{1+ch}, \mathscr{S}_0 \cap \xi_{t+1} = \emptyset]} = 1 - \frac{X_t h(1-h)}{(1-c)^2 X_t^2 h^2} = 1 - \frac{1-h}{(1-c)^2 X_t h}, \tag{22}$$

where in Eq. (22) we utilize the fact that the newly generated LOIAs can all be accepted[4] (where $X_t < N/(1+ch)$ holds), the number of newly generated LOIAs obeys the binomial distribution [15]. Hence, by considering the total number of LOIAs at the end of the $(t+1)^{th}$ generation, we know that

$$\mathbb{P}\left[X_{t+1} > (1+ch)X_t\Big| X_t < \frac{N}{1+ch}, \mathscr{S}_0 \cap \xi_{t+1} = \emptyset\right] > 1 - \frac{1-h}{(1-c)^2 X_t h} \tag{23}$$

holds, where $c \in (0, 1)$ is a constant. □

References

[1] T. Chen, J. He, G. Chen, X. Yao, Choosing selection pressure for wide-gap problems, Theoretical Computer Science 411 (6) (2010) 926–934.
[2] T. Chen, J. He, G. Sun, G. Chen, X. Yao, A new approach to analyzing average time complexity of population-based evolutionary algorithms on unimodal problems, IEEE Transactions on System, Man, and Cybernetics, Part B 39 (5) (2009) 1092–1106.
[3] T. Chen, K. Tang, G. Chen, X. Yao, Analysis of computational time of simple estimation of distribution algorithms, IEEE Transactions on Evolutionary Computation 14 (1) (2010) 1–22.
[4] S. Droste, T. Jansen, I. Wegener, On the analysis of the $(1 + 1)$ evolutionary algorithm, Theoretical Computer Science 276 (1–2) (2002) 51–81.
[5] T. Friedrich, P.S. Oliveto, D. Sudholt, C. Witt, Theoretical analysis of diversity mechanisms for global exploration, in: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'08), Atlanta, Georgia, USA, 2008, pp. 945–952.
[6] D.E. Goldberg, K. Deb, A comparative analysis of selection scheme used in genetic algorithms, in: Foundations of Genetic Algorithms, Morgan Kaufmann Publishers, 1991, pp. 69–93.
[7] J. He, X. Yao, Drift analysis and average time complexity of evolutionary algorithms, Artificial Intelligence 127 (1) (March 2001) 57–85.
[8] J. He, X. Yao, From an individual to a population: An analysis of the first hitting time of population-based evolutionary algorithms, IEEE Transactions on Evolutionary Computation 6 (5) (2002) 495–511.
[9] J. He, X. Yao, A study of drift analysis for estimating computation time of evolutionary algorithms, Natural Computing 3 (1) (2004) 21–35.
[10] T. Jansen, K.A.D. Jong, I. Wegener, On the choice of the offspring population size in evolutionary algorithms, Evolutionary Computation 13 (4) (2005) 413–440.
[11] P.K. Lehre, X. Yao, On the impact of mutation-selection balance on the runtime of evolutionary algorithms, in: IEEE Transactions on Evolutionary Computation, conditionally accepted on 12 November 2010.
[12] R. Motwani, P. Raghavan, Randomized Algorithms, Cambridge University Press, 1995.
[13] P.S. Oliveto, J. He, X. Yao, Evolutionary algorithms and the vertex cover problem, in: Proceedings of the IEEE Congress on Evolutionary Computation (CEC'07), Singapore, September 25-28, 2007, pp. 1870–1877.
[14] P.S. Oliveto, J. He, X. Yao, Analysis of the $(1 + 1)$-EA for finding approximate solutions to vertex cover problems, IEEE Transactions on Evolutionary Computation 13 (5) (2009) 1006–1029.
[15] A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, 1991.
[16] M.R. Spiegel, Theory and Problems of Probability and Statistics, McGraw-Hill, New York, 1992.
[17] C. Witt, An analysis of the $(\mu + 1)$ EA on simple pseudo-boolean functions, in: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'04), Seattle, Washington, USA, 2004, pp. 761–773.

---

[4] In our worst-case analysis, we ignore the generation of advanced LOIAs if the number of LOIAs has not reached a predefined level [2]