Artificial Intelligence 172 (2008) 1809–1832

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

A new approach to estimating the expected first hitting time of evolutionary algorithms

Yang Yu, Zhi-Hua Zhou *

National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China

ARTICLE INFO

Article history:
Received 12 December 2007
Received in revised form 8 July 2008
Accepted 9 July 2008
Available online 12 July 2008

Keywords:
Evolutionary algorithms
Expected first hitting time
Convergence rate
Computational complexity

ABSTRACT

Evolutionary algorithms (EA) have been shown to be very effective in solving practical problems, yet many important theoretical issues of them are not clear. The *expected first hitting time* is one of the most important theoretical issues of evolutionary algorithms, since it implies the average computational time complexity. In this paper, we establish a bridge between the expected first hitting time and another important theoretical issue, i.e., *convergence rate*. Through this bridge, we propose a new general approach to estimating the expected first hitting time. Using this approach, we analyze EAs with different configurations, including three mutation operators, with/without population, a recombination operator and a time variant mutation operator, on a hard problem. The results show that the proposed approach is helpful for analyzing a broad range of evolutionary algorithms. Moreover, we give an explanation of what makes a problem hard to EAs, and based on the recognition, we prove the hardness of a general problem.

1. Introduction

Evolutionary algorithms (EAs) are a kind of optimization technique, inspired by the natural evolution process. Despite many different implementations [1], e.g., *genetic algorithm*, *genetic programming* and *evolutionary strategies*, traditional evolutionary algorithms can be summarized below by four steps:

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

In the evolutionary process, a population of randomly initialized solutions is maintained and evolved. Mutation and recombination are two popular operators for reproduction in Step 2. A fitness function is employed to guide Step 3. The evolutionary repetition stops when, e.g., an optimal solution is found or time runs out.

EAs solve problems in straightforward ways and do not require, for example, continuous or differentiable functions or inversable matrices. So, EAs have been applied to bioinformatics [17], circuit design [3], data mining [9], information retrieval [4], etc. Despite the remarkable success achieved by EAs on practice problems, EAs are often criticized for the lack of a solid theoretical foundation. Actually, such a theoretical foundation is very desired in order to gain deep understanding of the strength and weakness of current EAs and thus develop better EAs.

* Corresponding author.
E-mail address: zhouzh@nju.edu.cn (Z.-H. Zhou).

0004-3702/$ – see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.artint.2008.07.001

The *first hitting time* of EAs is the time that, in a run, EAs find an optimal solution for the first time, and the *expected first hitting time* (EFHT) is the average time that EAs require to find an optimal solution, which implies the average computational time complexity of EAs. It is evident that the EFHT is one of the most important theoretical issues of EAs.

Many papers have been devoted to the analysis of simple EAs. The $(1+1)$-EA, i.e., EA without population, has been studied on the long path problem [22], the OneMax problem [23], the uni-model functions [5,7] and linear functions [6,7]. Another EA without population has been studied on the OneMax problem [10]. More details can be found in Beyer et al.'s survey [2]. Owing to these efforts, several theoretical properties of EAs become more clear. In these works, however, ad hoc approaches were used to analyze simple EAs on simple problems, yet a general approach that can be used to analyze wider kinds of EAs to gain deeper insights is more desired. Recently, several works [13–15] have been devoted to developing general analysis approaches, which are summarized in the latest survey [19].

He and Yao [13,15] have developed a general approach to analyzing a wide class of EAs based on *drift analysis* [11], which is a significant advance. Intuitively, if we know the length of the whole path toward the optimum and the length of the drift of the EA at each step, we can estimate the EFHT by dividing the path length by the step drift. However, no practical measure of these quantities is known.

He and Yao [14] have developed another framework based on the analytical solution of EFHT to analyze and compare EAs. Under this framework, two hard problem classes (i.e., problems that can only be solved in exponential time), the 'wide gap' problem class and the 'long path' problem class, were identified. Since the analytical framework is derived from homogeneous Markov chain models, only EAs with static reproduction operators can be analyzed, although EAs with time-variant operators or adaptive operators are very popular and powerful [8].

The *convergence rate* is another important theoretical issue of EAs, which implies how close the current state is to the optimal area at each step. The convergence issue has been studied for years [12,16,21,23,25]. He and Yu [16] did a thorough study based on the *minorization method* [20].

In this paper, we present the first study on the relationship between the EFHT and the convergence rate, and establish a bridge between them. Through this bridge, we propose a new general approach to estimating the expected first hitting time. In contrast to previous researches where easy problems (i.e., problems that can be solved in polynomial time) [6,15,23] were studied, we use the proposed approach to analyze EAs on a hard problem. The analyzed EAs involve various configurations, including three mutation operators, with/without population, a recombination operator and a time variant mutation operator. The results show that the proposed approach is helpful for analyzing a broad range of EAs. Moreover, we give an explanation of what makes a problem hard to an EA, and based on the recognition, we prove the hardness of a general problem.

The rest of this paper is organized as follows. In Section 2, we briefly review some related work and introduce how to model EAs using Markov chains. In Section 3, we introduce a new approach to estimating the EFHT, which is the main result of this paper. In Section 4, we analyze several EAs on a hard problem using the proposed approach, which is followed by discussions in Section 5. Finally, in Section 6, we conclude the paper.

2. Modeling EAs using Markov chain

EAs evolve solutions from generation to generation. Each generation stochastically depends on the previous one, except the initial generation which is randomly generated. This *conditional independence* can be modeled natually by Markov chains [12–14,18,24,25].

Combinatorial optimization problems are among the most common problems in practice, whose solutions can be represented by a sequence of symbols. In this paper, we use EAs to tackle them. To model this kind of EAs, we construct Markov chains with discrete state space. The key to construct such a Markov chain is to bijectively map the populations of an EA to the states of the Markov chain. A popular mapping [12,16,25] enables one state of the Markov chain to correspond to one possible population of the EA. Suppose an EA encodes a solution in a vector of length L, each component of the vector is drawn from an alphabet set $\mathcal{B}$, and each population contains M solutions. Let $\mathcal{S}$ denote the solution space. There are $|\mathcal{S}| = |\mathcal{B}|^L$ number of different solutions. Let X denote the population space. There are $|X| = \binom{M+|\mathcal{B}|^L-1}{M}$ number of different possible populations [25]. A Markov chain which models the EA is constructed by taking X as the state space, i.e., a chain $\{\xi_t\}_{t=0}^{+\infty}$ is built where $\xi_t \in X$.

A population is called an *optimal population* if it contains at least one optimal solution. Let X^* $(\in X)$ denote the set of all optimal populations. The goal of EAs is to reach X^* from an initial population. Thus, the process of an EA which seeks X^* can be analyzed by studying the corresponding Markov chain [12,16].

In the rest of this section, we introduce several notations and definitions. Given a Markov chain $\{\xi_t\}_{t=0}^{+\infty}(\xi_t \in X)$ and a target subspace $X^* \subset X$, let μ_t $(t = 0, 1, \ldots)$ denote the probability of ξ_t being in X^*, i.e.,

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

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

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

In [16], *convergence rate* is measured by $1-\mu_t$ at step t, which is equivalent to that used in [25]. Therefore, we also use $1-\mu_t$ as the measure of convergence rate in this paper.

Definition 2 *(Convergence rate).* Given a Markov chain $\{\xi_t\}_{t=0}^{+\infty}$ $(\xi_t \in X)$ and a target subspace $X^* \subset X$, the convergence rate to X^* at time t is $1-\mu_t$.

Definition 3 *(Absorbing Markov chain).* Given a Markov chain $\{\xi_t\}_{t=0}^{+\infty}(\xi_t \in X)$ and a target subspace $X^* \subset X$, $\{\xi_t\}_{t=0}^{+\infty}$ is said to be an absorbing chain, if

$$\forall t \in \{0,1,\ldots\}: \quad P\big(\xi_{t+1} \notin X^* \mid \xi_t \in X^*\big) = 0. \tag{3}$$

We use absorbing Markov chains to model all the EAs studied in this paper, because absorbing Markov chains have good theoretical properties and can be practically achieved. An EA can be modeled by an absorbing Markov chain if it never loses an optimal solution once found. Actually, most EAs for real problems satisfy this condition because, if an optimal solution can be identified, the EA will stop when it finds them; otherwise, when optimal solutions cannot be identified, the commonly used strategy of keeping the best-so-far solution in every generation can make the condition be satisfied. Moreover, EAs that can be modeled by absorbing Markov chains converge to optimal solutions with certain operators [16], which is a desirable property in practice.

Definition 4 *(Expected first hitting time, EFHT).* Given a Markov chain $\{\xi_t\}_{t=0}^{+\infty}$ $(\xi_t \in X)$ and a target subspace $X^* \subset X$, let a random variable τ denote the events

$$\begin{aligned}
&\tau = 0\text{: } \xi_0 \in X^*,\\
&\tau = 1\text{: } \xi_1 \in X^* \wedge \xi_i \notin X^* \quad (i=0),\\
&\tau = 2\text{: } \xi_2 \in X^* \wedge \xi_i \notin X^* \quad \big(\forall i \in \{0,1\}\big),\\
&\ldots\\
&\tau = t\text{: } \xi_t \in X^* \wedge \xi_i \notin X^* \quad \big(\forall i \in \{0,1,\ldots,t-1\}\big), \ldots.
\end{aligned}$$

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

This definition of EFHT is equivalent to those used in [13,14]. The EFHT of an EA is the average time in which it finds an optimal solution, which is its average computational time complexity.

Markov chains model the essential of the corresponding EA processes, thus the convergence, convergence rate and EFHT of EAs can be obtained by analyzing the corresponding Markov chains. So, in the rest of the paper, we do not distinguish the convergence, convergence rate and EFHT of EAs and those of the corresponding Markov chains.

3. Deriving expected first hitting time from convergence rate

The convergence rate has been studied for many years [12,25] and recently a general bound has been developed in [16] through the *minorization condition* method [20]. Since we focus on using EAs to solve combinatorial optimization problems in this paper, a discrete space version of Theorem 4 in [16] is proven below.

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

$$\prod_{t=0}^{+\infty}(1-\alpha_t) = 0 \tag{4}$$

and

$$\beta_t \geqslant \sum_{x \notin X^*} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big)\frac{P(\xi_t = x)}{1-\mu_t} \geqslant \alpha_t, \tag{5}$$

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

$$(1-\mu_0)\prod_{i=0}^{t-1}(1-\alpha_i) \geqslant 1-\mu_t \geqslant (1-\mu_0)\prod_{i=0}^{t-1}(1-\beta_i). \tag{6}$$

Proof. From Eqs. (1) and (3), it follows that

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

and by applying Eq. (5) we get

$$(1 - \mu_{t-1})\alpha_{t-1} \leqslant \mu_t - \mu_{t-1} \leqslant (1 - \mu_{t-1})\beta_{t-1},$$
$$(1 - \mu_{t-1})(1 - \alpha_{t-1}) \geqslant 1 - \mu_t \geqslant (1 - \mu_{t-1})(1 - \beta_{t-1}),$$

by applying this inequality recursively, we have

$$(1 - \mu_0)\prod_{i=0}^{t-1}(1 - \alpha_i) \geqslant 1 - \mu_t \geqslant (1 - \mu_0)\prod_{i=0}^{t-1}(1 - \beta_i). \qquad \square$$

Lemma 1 implies that as far as the probability of an EA 'jumping' into the set of optimal solutions can be estimated for each step, the bounds of its convergence rate can be derived. The only requirement is that the EA can be modeled by an absorbing Markov chain, i.e., the EA satisfies Eq. (3). As mentioned before, most EAs used in real problems meet this requirement.

In Definition 4, the EFHT is the mathematical expectation of the random variable τ. Meanwhile, the probability distribution of τ is the probability of an optimal solution being found before step t $(t = 0, 1, \ldots)$. Thus, as long as the EA can be modeled by an absorbing Markov chain, it holds that

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

This implies that the probability distribution of τ is equal to μ_t, which is one minus the convergence rate. So, the convergence rate and the EFHT are two sides of a coin.

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

Lemma 2. *Let u and v denote two discrete random variables that are nonnegative integers with limited expectation, and $D_u(\cdot)$ and $D_v(\cdot)$ denote their distribution functions, respectively, i.e.,*

$$D_u(t) = P(u \leqslant t) = \sum_{i=0}^{t} P(u = i),$$
$$D_v(t) = P(v \leqslant t) = \sum_{i=0}^{t} P(v = i).$$

If $D_u(t) \geqslant D_v(t)$ $(\forall t = 0, 1, \ldots)$, then the expectations of the random variables satisfy

$$\mathbb{E}[u] \leqslant \mathbb{E}[v], \tag{7}$$

where $\mathbb{E}[u] = \sum_{t=0,1,\ldots} tP(u = t)$ and $\mathbb{E}[v] = \sum_{t=0,1,\ldots} tP(v = t)$.

Proof. Since D_u is the distribution of u,

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

and same for v. Thus,

$$\begin{aligned}\mathbb{E}[u]-\mathbb{E}[v] &= \sum_{i=0}^{+\infty}\big(1-D_u(i)\big)-\sum_{i=0}^{+\infty}\big(1-D_v(i)\big)\\ &= \sum_{i=0}^{+\infty}\big(D_v(i)-D_u(i)\big)\\ &\leqslant 0. \quad \square\end{aligned}$$

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

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

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

$$\prod_{t=0}^{+\infty}(1-\alpha_t)=0 \tag{8}$$

and

$$\beta_t \geqslant \sum_{x\notin X^*} P\big(\xi_{t+1}\in X^* \mid \xi_t=x\big)\frac{P(\xi_t=x)}{1-\mu_t} \geqslant \alpha_t, \tag{9}$$

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

$$\mathbb{E}[\tau] \leqslant \alpha_0+\sum_{t=2}^{+\infty} t\alpha_{t-1}\prod_{i=0}^{t-2}(1-\alpha_i) \tag{10}$$

and

$$\mathbb{E}[\tau] \geqslant \beta_0+\sum_{t=2}^{+\infty} t\beta_{t-1}\prod_{i=0}^{t-2}(1-\beta_i). \tag{11}$$

Proof. Applying Lemma 1 with Eq. (9), we have

$$1-\mu_t \leqslant (1-\mu_0)\prod_{i=0}^{t-1}(1-\alpha_i).$$

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

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

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

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

Since $D_\tau(t) \geqslant D_\eta(t)$, according to Lemma 2, $\mathbb{E}[\tau] \leqslant \mathbb{E}[\eta]$. Thus, the upper bound of the EFHT is

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

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

The lower bound of the EFHT can be derived similarly. $\square$

Two points in Theorem 1 remain to be clarified. Firstly, ‘starting from non-optimal solutions’ is just a theoretical assumption that is used to make the result easy to read. Practically, for the problems where EAs are applied, the probability of a randomly generated solution being optimal is exponentially small. In such case, this assumption will not affect the result of the asymptotic analysis. Secondly, Theorem 1 is written in a compact form, i.e., we will have both lower and upper bounds of the EFHT if we have both β_t and α_t in Eq. (9). Actually, it is also applicable when we only have one of them. We will have a lower bound of the EFHT if we have β_t, and we will have an upper bound of the EFHT if we have α_t.

The bounds of EFHT, i.e., Eqs. (10) and (11), have an intuitive explanation. The part $\alpha_{t-1}\prod_{i=0}^{t-2}(1-\alpha_i)$ (or replacing α by β) indicates the probability of the event that the EA finds an optimal solution at the tth step, but does not find it at any earlier step.

Theorem 1 shows that we can have bounds of the EFHT from the bounds of the formula

$$\sum_{x\notin X^*} P\big(\xi_{t+1}\in X^* \mid \xi_t = x\big)\frac{P(\xi_t = x)}{1-\mu_t}. \tag{12}$$

The first part of the formula $P(\xi_{t+1}\in X^* \mid \xi_t = x)$ is the probability of the EA 'jumping' into an optimal population, which we call as *success probability*. The second part $\frac{P(\xi_t=x)}{1-\mu_t}$ is a *normalized distribution* over non-optimal states. As long as these two parts are estimated, the bounds of the EFHT can be derived. The more accurate the estimated probability, the tighter the derived bounds.

4. Case study on a hard problem

In this section, we will prove that the *Trap problem* is hard (i.e., can only be solved in exponential time) for several EAs, using our proposed approach. The Trap problem is defined below.

Definition 5 *(Trap problem).* Given a set of n positive values, i.e., $W=\{w_i\}_{i=1}^n$, and a capacity value c, to find x^* from

$$x^* = \mathop{\arg\max}_{x\in\{0,1\}^n} \sum_{i=1}^n w_i\cdot x_i$$

$$\text{s.t.}\quad \sum_{i=1}^n w_i\cdot x_i \leqslant c,$$

where $w_1 = w_2 = \cdots = w_{n-1} > 1$, $w_n = (\sum_{i=1}^{n-1} w_i)+1$ and $c = w_n$.

Trap problem has one optimal solution $x^* = (000\ldots01)$. A solution is a feasible solution if it satisfies the constraint, otherwise it is an infeasible solution.

We try to tackle the Trap problem using several EAs which are configured commonly as below. The *Reproduction* will be implemented by concrete operators later.

- *Encoding:* Each solution is encoded by a string with n binary bits, where the ith bit is 1 if w_i is included and 0 otherwise.
- *Initial:* Randomly generate a population of M solutions encoded by binary strings.
- *Reproduction:* Generate M new solutions from the current population.
- *Selection:* Select the best M solutions among the current population and the reproduced solutions, which is also called plus-selection, to form the population of the next generation. The selected M solutions are with the best fitness value (according to the definition below).
- *Fitness:* The fitness of a solution $x=(x_1x_2\ldots x_n)$ is defined as

$$\text{Fitness}(x) = \theta\sum_{i=1}^n w_i x_i - c, \tag{13}$$

  where $\theta = 1$ when x is a feasible solution, i.e., $\sum_{i=1}^n w_i x_i \leqslant c$, and $\theta = 0$ otherwise. The fitness function is to be *maximized*, and the larger the fitness is, the better the solution is. Here, the maximum fitness value is zero.
- *Stop criterion:* If the largest fitness value in population is zero, stop and output the solution with the maximum fitness.

To implement the Reproduction operator, we use several popular operators, listed below.

Mutation#1 (bitwise mutation with constant probability): Independently flip each bit of each solution with an constant probability $p_m \in (0, 0.5]$.

Mutation#2 (bitwise mutation with probability 1/n): Independently flip each bit of each solution with probability $p_m = \frac{1}{n}$. This may be the most commonly used mutation operator.

Mutation#3 (one-bit mutation): Randomly flip one bit of each solution.

Mutation#4 (time-variant mutation): Independently flip each bit of each solution with probability $(0.5-d)e^{-t}+d$, where $d\in(0,0.5]$ and $t = 0,1,\ldots$.

Recombination (one-point crossover): Exchange the leading σ bits of randomly selected two solutions, where σ is drawn randomly from $\{1,2,\ldots,n-1\}$.

Mutation#1 seems like a special case of Mutation#2. However, there is a significant difference, that is, the mutation probability of Mutation#2 is adapted to the problem size while that of Mutation#1 is a constant. So, the asymptotic behaviors, as $n \to +\infty$, of the two are different.

To focus on the order of the asymptotic complexity of EFHT of EAs, we use the $\Omega(\cdot)$ representation. For two functions $f(\cdot)$ and $g(\cdot)$, we write $f(n) = \Omega(g(n))$ to represent that $g(n)$ is an asymptotic lower bound of $f(n)$, if and only if

$$\lim_{n\to+\infty} \frac{f(n)}{g(n)} > 0,$$

and meanwhile, we write $g(n) = O(f(n))$.

4.1. Static mutation without population

First, we show how the three static mutation operators, Mutations #1, #2 and #3, perform on the Trap problem, with population size 1, that is, $(1+1)$-EA. Since a population is equal to a solution, the population state space is equal to the solution state space, i.e., $X = \mathcal{S}$.

Proposition 1. *Solving the Trap problem using the EA with Reproduction implemented by Mutation#1 (bitwise mutation with constant probability) and with a population size 1, i.e. $(1+1)$-EA, if starting from non-optimal populations, the EFHT is bounded by*

$$\mathbb{E}[\tau] = \Omega(\theta^n), \tag{14}$$

where $\theta = (1 - p_m)^{-1} \in (1, 2]$ is a constant and n is the problem size.

To prove this proposition, we need to find an upper bound of formula (12) applying Theorem 1. We first investigate the part of success probability of formula (12), $P(\xi_{t+1} \in X^* \mid \xi_t = x)$. Assuming a solution has k bits different from the optimal solution, the probability of the solution being mutated to be the optimal solution is $p_m^k(1 - p_m)^{n-k}$ using Mutation#1. So the maximum probability of a solution being mutated to be the optimal solution is $p_m(1 - p_m)^{n-1}$, which means that there is only one bit difference. Therefore, we have $P(\xi_{t+1} \in X^* \mid \xi_t = x) \leqslant p_m(1 - p_m)^{n-1}$. Then, by applying Theorem 1 with this upper bound, we get this proposition.

Proof. Since $P(\xi_{t+1} \in X^* \mid \xi_t = x) \leqslant p_m(1 - p_m)^{n-1}$, we have

$$\begin{aligned}
&\sum_{x \notin X^*} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) \frac{P(\xi_t = x)}{1 - \mu_t} \\
&\quad \leqslant \sum_{x \notin X^*} p_m(1 - p_m)^{n-1} \frac{P(\xi_t = x)}{1 - \mu_t} \\
&\quad = p_m(1 - p_m)^{n-1} \frac{\sum_{x \notin X^*} P(\xi_t = x)}{1 - \mu_t} \\
&\quad = p_m(1 - p_m)^{n-1}. \quad \text{\%\% by Eq. (1)}
\end{aligned}$$

Let $\beta_t = p_m(1 - p_m)^{n-1}$, by Theorem 1,

$$\mathbb{E}[\tau] \geqslant \beta_0 + \sum_{t=2}^{+\infty} t\beta_{t-1} \prod_{i=0}^{t-2} (1 - \beta_i) = \frac{1}{p_m}\left(\frac{1}{1 - p_m}\right)^{n-1} = \frac{1}{p_m}\theta^{n-1}.$$

Considering that p_m is a constant, $\mathbb{E}[\tau] = \Omega(\theta^n)$. □

Proposition 2. *Solving the Trap problem using the EA with Reproduction implemented by Mutation#2 (bitwise mutation with probability $1/n$) and with a population size 1, i.e. $(1+1)$-EA, if starting from non-optimal populations, the EFHT is bounded by*

$$\mathbb{E}[\tau] = \Omega\big(2^n\big), \tag{15}$$

where n is the problem size.

If we follow the idea in the proof of Proposition 1 to obtain an upper bound of success probability, i.e., $P(\xi_{t+1} \in X^* \mid \xi_t = x) \leqslant 1/(en)$, we can only obtain an $\Omega(n)$ lower bound for the EFHT, which is too loose.

To get a tighter bound, we take two steps. First, we show that formula (12) is $O(1/2^n)$ at the beginning, i.e., $t = 0$. Second, we show that formula (12) decreases as t increases. So $O(1/2^n)$ is an upper bound of formula (12). By applying Theorem 1 we get Proposition 1.

There is a trick for calculating formula (12). We divide the state space into subspaces, in which states share some common properties, and treat each subspace as a whole. We first divide the state space into $n + 1$ subspaces $\{X_i\}_{i=0}^n$, where

X_i contains all the solutions that have exactly i identical bits with the optimal solution, so that the solutions in each subspace have the same probability of being mutated to be the optimal solution. By this division, the success probability at $t=0$ is calculated. We then divide the state space into the optimal space X^*, the feasible space X_F and the infeasible space X_I, according to whether solutions satisfy the constraint, and combine this division with the previous one. By this division, we find that formula (12) decreases as t increases.

Proof. Let

$$X_i = \{x \in X \mid \|x - x^*\|_H = n - i\},$$

where $\|\cdot\|_H$ is Hamming distance and x^* is the optimal solution, which means solutions in X_i have i bits identical with the optimal solution, and that $X = \bigcup_{i=0}^{n} X_i$, $|X_i| = \binom{n}{i}$, and $X_n = X^*$.

Then, by applying Mutation#2, we calculate the success probability

$$\forall x \in X_i\colon \quad P(\xi_{t+1} \in X^* \mid \xi_t = x) = \left(\frac{1}{n}\right)^{n-i}\left(1 - \frac{1}{n}\right)^{i}.$$

At $t=0$, we have

$$\begin{aligned}
&\sum_{x \notin X^*} P(\xi_1 \in X^* \mid \xi_0 = x)\frac{P(\xi_0 = x)}{1-\mu_0} \\
&\quad = \sum_{x \notin X^*} P(\xi_1 \in X^* \mid \xi_0 = x) P(\xi_0 = x) \quad \text{%% by assumption } \mu_0 = 0 \\
&\quad = \sum_{i=0}^{n-1} \sum_{x \in X_i} \big(P(\xi_1 \in X^* \mid \xi_0 = x) P(\xi_0 = x)\big) \quad \text{%% by } X = \bigcup_{i=0}^{n} X_i \\
&\quad = \sum_{i=0}^{n-1} \binom{n}{i}\left(\frac{1}{n}\right)^{n-i}\left(1-\frac{1}{n}\right)^{i}\frac{1}{2^n} \quad \text{%% by } \forall x\colon P(\xi_0 = x) = \frac{1}{2^n} \\
&\quad = \left(1 - \left(\frac{n-1}{n}\right)^{n}\right)\frac{1}{2^n} \\
&\quad \sim \frac{e-1}{e}\frac{1}{2^n}.
\end{aligned}$$

Let $X = X^* \cup X_F \cup X_I$, where X^* contains the optimal solutions, X_F contains all non-optimal feasible solutions whose last bit is 0, and X_I contains all the infeasible solutions whose last bit is 1. Denote

$$X_i^F = X_i \cap X_F.$$

According to the fitness function, we have

$$\begin{aligned}
&\forall x_0 \in X_0^F, x_1 \in X_1^F, \ldots, x_{n-1} \in X_{n-1}^F, x_I \in X_I\colon \\
&\quad f(x^*) > f(x_0) > f(x_1) > \cdots > f(x_{n-1}) > f(x_I)
\end{aligned}$$

and due to the selection behavior, i.e., the solutions with the largest fitness will be selected, we have

$$\begin{aligned}
&\forall j, q \ (n-1 \geqslant j > q \geqslant 0)\colon \\
&\quad P(\xi_{t+1} \in X_j^F \mid \xi_t \in X_q^F) = 0, \qquad P(\xi_{t+1} \in X_I \mid \xi_t \in X_F) = 0, \\
&\forall j, q\colon \quad \frac{P(\xi_{t+1} \in X_j^F \mid \xi_t \in X_I)}{P(\xi_{t+1} \in X_q^F \mid \xi_t \in X_I)} = \frac{P(\xi_0 \in X_j^F)}{P(\xi_0 \in X_q^F)},
\end{aligned}$$

where the last equation is by that, since every infeasible solution has the same lowest fitness, there is no selection pressure on the leading $n-1$ bits when the solution is infeasible, and thus each of the leading $n-1$ bits has probability 0.5 to be either zero or one.

For all $k \in \{0, 1, \ldots, n-1\}$, denoting $X_{A_k} = \bigcup_{i=0}^{k} X_i^F$ and $X_{B_k} = \bigcup_{i=k+1}^{n-1} X_i^F$, at time $t=0$, for all $k \in \{0, 1, \ldots, n-1\}$, there exists $\eta_{A_k,t}, \eta_{B_k,t}, \eta_{F,t}$ and $\eta_{I,t}$ such that

$$\begin{aligned}
&\sum_{x \in X_{A_k}} P(\xi_{t+1} \in X^* \mid \xi_t = x) P(\xi_t = x) = \eta_{A_k,t} P(\xi_t \in X_{A_k}), \\
&\sum_{x \in X_{B_k}} P(\xi_{t+1} \in X^* \mid \xi_t = x) P(\xi_t = x) = \eta_{B_k,t} P(\xi_t \in X_{B_k}),
\end{aligned}$$

$$\sum_{x \in X_F} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) P(\xi_t = x) = \eta_{F,t} P(\xi_t \in X_F),$$
$$\sum_{x \in X_I} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) P(\xi_t = x) = \eta_{I,t} P(\xi_t \in X_I).$$

On the relationship between $\eta_{A_k,t}$ and $\eta_{B_k,t}$, it holds that

$$\eta_{A_k,t} < \eta_{B_k,t},$$

by $P(\xi_{t+1} \in X^* \mid \xi_t \in X_{A_k}) \leqslant P(\xi_{t+1} \in X^* \mid \xi_t \in X_k^F) \leqslant P(\xi_{t+1} \in X^* \mid \xi_t \in X_{B_k})$.

On the relationship between $\eta_{A_k,t}$ and $\eta_{I,t}$, it holds that at $t = 0$,

$$\eta_{A_k,0} < \eta_{I,0},$$

by, first, $\eta_{A_k,0} < \eta_{F,0}$, which is by $\eta_{F,0} = \eta_{A_k,0} \frac{P(\xi_0 \in X_A)}{P(\xi_0 \in X_F)} + \eta_{B_k,0} \frac{P(\xi_0 \in X_B)}{P(\xi_0 \in X_F)}$ and $\eta_{A_k,0} < \eta_{B_k,0}$, and second, $\eta_{F,0} < \eta_{I,0}$, which is by $\forall x_1, x_2 \in X$: $P(\xi_0 = x_1) = P(\xi_0 = x_2)$ and

$$\frac{\sum_{x \in X_F} P(\xi_1 \in X^* \mid \xi_0 = x)}{P(\xi_0 \in X_F)} < \frac{\sum_{x \in X_I} P(\xi_1 \in X^* \mid \xi_0 = x)}{P(\xi_0 \in X_I)}.$$

And for $t > 0$,

$$\eta_{A_k,t} < \eta_{I,0},$$

by that, since

$$\begin{aligned}
P(\xi_{t+1} \in X_{A_k}) &= P(\xi_t \in X_{A_k}) + \sum_{x \in X_{B_k} \cup X_I} P(\xi_{t+1} \in X_{A_k} \mid \xi_t = x) P(\xi_t = x) - \sum_{x \in X_{A_k}} P\big(\xi_{t+1} \in X^* \cup X_{B_k} \cup X_I \mid \xi_t = x\big) P(\xi_t = x) \\
&= P(\xi_t \in X_{A_k}) + \sum_{x \in X_{B_k} \cup X_I} P\big(\xi_{t+1} \in X_{A_k} \mid \xi_t = x\big) P(\xi_t = x) - \sum_{x \in X_{A_k}} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) P(\xi_t = x) \\
&\qquad \text{\%\% by } P\big(\xi_{t+1} \in X_{B_k} \cup X_I \mid \xi_t \in X_{A_k}\big) = 0 \\
&> \big(1 - P\big(\xi_{t+1} \in X^* \mid \xi_t \in X_k^F\big)\big) P(\xi_t \in X_{A_k}) + \sum_{x \in X_I} P\big(\xi_{t+1} \in X_{A_k} \mid \xi_t = x\big) P(\xi_t = x) \\
&\qquad \text{\%\% by } -P\big(\xi_{t+1} \in X^* \mid \xi_t \in X_A\big) \geqslant -P\big(\xi_{t+1} \in X^* \mid \xi_t \in X_k^F\big) \\
&> P\big(\xi_0 \in X_{A_k}\big) \prod_{i=0}^{t} \big(1 - P(\xi_{i+1} \in X^* \mid \xi_i \in X_k^F)\big) \\
&\quad + \sum_{i=0}^{t} \left(\sum_{x \in X_I} P\big(\xi_{i+1} \in X_{A_k} \mid \xi_i = x\big) P(\xi_i = x) \right) \left(\prod_{j=i}^{t} \big(1 - P(\xi_{j+1} \in X^* \mid \xi_j \in X_k^F)\big) \right) \\
&= P(\xi_0 \in X_{A_k}) \prod_{i=0}^{t} \big(1 - P(\xi_{i+1} \in X^* \mid \xi_i \in X_k^F)\big) \\
&\quad + \sum_{i=0}^{t} P\big(\xi_{i+1} \in X_{A_k} \mid \xi_i \in X_I\big) P(\xi_i \in X_I) \left(\prod_{j=i}^{t} \big(1 - P(\xi_{j+1} \in X^* \mid \xi_j \in X_k^F)\big) \right),
\end{aligned}$$

and similarly,

$$\begin{aligned}
P(\xi_{t+1} \in X_{B_k}) &< P\big(\xi_0 \in X_{B_k}\big) \prod_{i=0}^{t} \big(1 - P(\xi_{i+1} \in X^* \mid \xi_i \in X_k^F)\big) \\
&\quad + \sum_{i=0}^{t} P\big(\xi_{i+1} \in X_{B_k} \mid \xi_i \in X_I\big) P(\xi_i \in X_I) \left(\prod_{j=i}^{t} \big(1 - P(\xi_{j+1} \in X^* \mid \xi_j \in X_k^F)\big) \right),
\end{aligned}$$

and

$$\frac{P(\xi_{t+1} \in X_j^F \mid \xi_t \in X_I)}{P(\xi_{t+1} \in X_q^F \mid \xi_t \in X_I)} = \frac{P(\xi_0 \in X_j^F)}{P(\xi_0 \in X_q^F)},$$

we thus have

$$\forall k\colon \quad \frac{P(\xi_{t+1} \in X_{A_k})}{P(\xi_{t+1} \in X_{B_k})} > \frac{P(\xi_0 \in X_{A_k})}{P(\xi_0 \in X_{B_k})};$$

since $P(\xi_{t+1} \in X^* \mid \xi_t \in X_{A_k}) < P(\xi_{t+1} \in X^* \mid \xi_t \in X_{B_k})$, by enumerating k, we have

$$\frac{\sum_{x\in X_{A_k}} P(\xi_{t+1} \in X^* \mid \xi_t = x)P(\xi_t = x)}{P(\xi_t \in X_{A_k})} < \frac{\sum_{x\in X_{A_k}} P(\xi_1 \in X^* \mid \xi_t = x)P(\xi_0 = x)}{P(\xi_0 \in X_{A_k})},$$

which is $\eta_{A_k,t} < \eta_{A_k,0}$, and by $\eta_{A_k,0} < \eta_{I,0}$, it holds $\eta_{A_k,t} < \eta_{I,0}$.

Then, we have

$$\begin{aligned}
\frac{P(\xi_{t+1} \in X_{A_k})}{1-\mu_{t+1}} &= \frac{P(\xi_t \in X_{A_k}) - \sum_{x\in X_{A_k}} P(\xi_{t+1} \in X^* \mid \xi_t = x)P(\xi_t = x)}{1-\mu_t - \sum_{x\in X_{A_k}\cup X_{B_k}\cup X_I} P(\xi_{t+1} \in X^* \mid \xi_t = x)P(\xi_t = x)} \\
&= \frac{(1-\eta_{A_k,t})P(\xi_t \in X_{A_k})}{1-\mu_t - \eta_{A_k,t}P(\xi_t \in X_{A_k}) - \eta_{B_k,t}P(\xi_t \in X_{B_k}) - \eta_{I,t}P(\xi_t \in X_I)} \\
&> \frac{(1-\eta_{A_k,t})P(\xi_t \in X_{A_k})}{1-\mu_t - \eta_{A_k,t}P(\xi_t \in X_{A_k}) - \eta_{A_k,t}P(\xi_t \in X_{B_k}) - \eta_{A_k,t}P(\xi_t \in X_I)} \\
&\qquad \%\%\ \text{by } \eta_{A_k,t} < \eta_{B_k,t},\ \eta_{A_k,t} < \eta_{I,0} \text{ and } \eta_{I,t} = \eta_{I,0} \\
&= \frac{(1-\eta_{A_k,t})P(\xi_t \in X_{A_k})}{(1-\eta_{A_k,t})(1-\mu_t)} \\
&= \frac{P(\xi_t \in X_{A_k})}{1-\mu_t},
\end{aligned}$$

which is $\forall n-1 \geqslant k \geqslant 0{:}\ \frac{P(\xi_{t+1}\in\bigcup_{i=0}^k X_i^F)}{1-\mu_{t+1}} \geqslant \frac{P(\xi_t\in\bigcup_{i=0}^k X_i^F)}{1-\mu_t}$ by writing back $X_{A_k} = \bigcup_{i=0}^k X_i^F$. So, we have

$$\begin{aligned}
&\sum_{x\notin X^*} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big)\frac{P(\xi_t = x)}{1-\mu_t} \\
&\quad= \sum_{x\in X_F} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big)\frac{P(\xi_t = x)}{1-\mu_t} + \sum_{x\in X_I} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big)\frac{P(\xi_t = x)}{1-\mu_t} \\
&\quad\leqslant \sum_{x\in X_F} P\big(\xi_1 \in X^* \mid \xi_0 = x\big)\frac{P(\xi_0 = x)}{1-\mu_0} + \sum_{x\in X_I} P\big(\xi_1 \in X^* \mid \xi_0 = x\big)\frac{P(\xi_0 = x)}{1-\mu_0} \\
&\qquad \%\%\ \text{by both that } \forall n-1 \geqslant k \geqslant 0{:}\ \frac{P(\xi_{t+1} \in \bigcup_{i=0}^k X_i^F)}{1-\mu_{t+1}} \geqslant \frac{P(\xi_t \in \bigcup_{i=0}^k X_i^F)}{1-\mu_t} \text{ and} \\
&\qquad \%\%\ P\left(\xi_{t+1} \in X^* \mid \xi_t \in \bigcup_{i=0}^k X_i^F\right) \leqslant P\left(\xi_{t+1} \in X^* \mid \xi_t \in \bigcup_{i=0}^k X_i^F\right) \\
&\quad= \sum_{x\notin X^*} P\big(\xi_1 \in X^* \mid \xi_0 = x\big)\frac{P(\xi_0 = x)}{1-\mu_0} \\
&\quad= \frac{e-1}{e}\frac{1}{2^n}.
\end{aligned}$$

Let $\beta_t = \frac{e-1}{e}\frac{1}{2^n}$, by Theorem 1,

$$\mathbb{E}[\tau] \geqslant \frac{e}{e-1}2^n,$$

that is, $\mathbb{E}[\tau] = \Omega(2^n)$. □

Proposition 3. *Solving the Trap problem using the EA with Reproduction implemented by Mutation#3 (one-bit mutation) and with a population size 1, i.e. $(1+1)$-EA, if starting from non-optimal populations, the EFHT is bounded by*

$$\mathbb{E}[\tau] = \Omega(2^n), \tag{16}$$

where n is the problem size.

We can follow the idea of the proof to Proposition 2. First, at $t = 0$, formula (12) is calculated to be $O(1/2^n)$. Then we find that formula (12) reduces as t increases, which leads to an upper bound $O(1/2^n)$ of formula (12). By Theorem 1, the EFHT has a lower bound $\Omega(1/2^n)$. The difference to the proof of Proposition 2 is that the solution space is divided into subspaces in a different way, according to the characteristic of Mutation#3. To arrive at the proof, we divide the state space into subspaces, in each subspace solutions have the same Hamming distance to the optimal solution. By this division, we find that only the solutions, which have only one bit different from the optimal solution, have non-zero probability of being mutated to be the optimal solution. Thus, formula (12) is calculated.

Proof. Let

$$X_i = \{x \in X \mid \|x - x^*\|_H = n - i\},$$

where $\|\cdot\|_H$ is Hamming distance and x^* is the optimal solution, such that $X = \bigcup_{i=0}^{n} X_i$, $|X_i| = \binom{n}{i}$, and $X_n = X^*$. Then, by applying Mutation#3, the success probability is

$$\forall x \in X_i\colon \quad P(\xi_{t+1} \in X^* \mid \xi_t = x) = \begin{cases} \frac{1}{n}, & \text{if } i = n-1, \\ 0, & \text{otherwise.} \end{cases}$$

Noticing that X_{n-1} contains the one feasible solution, which has the lowest fitness among feasible solutions, and $n-1$ infeasible solutions, which has the lowest fitness among all solutions, we have

$$P(\xi_{t+1} \in X_{n-1} \mid \xi_t \in X - X_{n-1} - X^*) < P(\xi_{t+1} \in X - X_{n-1} - X^* \mid \xi_t \in X_{n-1}).$$

At $t = 0$, we have

$$\begin{aligned}
&\sum_{x \notin X^*} P(\xi_1 \in X^* \mid \xi_0 = x) \frac{P(\xi_0 = x)}{1 - \mu_0} \\
&\quad = \sum_{x \in X_{n-1}} P(\xi_1 \in X^* \mid \xi_0 = x) \frac{P(\xi_0 = x)}{1 - \mu_0} \quad \text{\%\% by subspace dividing} \\
&\quad = \sum_{x \in X_{n-1}} \frac{1}{n} \frac{P(\xi_0 = x)}{1 - \mu_0} \quad \text{\%\% by } P(\xi_{t+1} \in X^* \mid \xi_t \in X_{n-1}) = \frac{1}{n} \\
&\quad = \frac{1}{n} \frac{n}{2^n} = \frac{1}{2^n}. \quad \text{\%\% by } P(\xi_0 \in X_{n-1}) = \frac{|X_{n-1}|}{2^n} = \frac{1}{2^n}
\end{aligned}$$

At time $t+1$, on the relationship between μ_t and μ_{t+1}, we have

$$\begin{aligned}
\mu_{t+1} &= \mu_t + P(\xi_{t+1} \in X^* \mid \xi_t \in X - X_{n-1} - X^*) + P(\xi_{t+1} \in X^* \mid \xi_t \in X_{n-1}) \\
&= \mu_t + P(\xi_{t+1} \in X^* \mid \xi_t \in X_{n-1}). \\
&\qquad \text{\%\% by } P(\xi_{t+1} \in X^* \mid \xi_t \in X - X_{n-1} - X^*) = 0
\end{aligned}$$

On the relationship between $P(\xi_t \in X_{n-1})$ and $P(\xi_{t+1} \in X_{n-1})$, we have

$$\begin{aligned}
P(\xi_{t+1} \in X_{n-1}) &= P(\xi_t \in X_{n-1}) + P(\xi_{t+1} \in X_{n-1} \mid \xi_t \in X - X_{n-1} - X^*) \\
&\quad - P(\xi_{t+1} \in X^* \mid \xi_t \in X_{n-1}) - P(\xi_{t+1} \in X - X_{n-1} - X^* \mid \xi_t \in X_{n-1}) \\
&\leqslant P(\xi_t \in X_{n-1}) - P(\xi_{t+1} \in X^* \mid \xi_t \in X_{n-1}). \\
&\qquad \text{\%\% by } P(\xi_{t+1} \in X_{n-1} \mid \xi_t \in X - X_{n-1} - X^*) < P(\xi_{t+1} \in X - X_{n-1} - X^* \mid \xi_t \in X_{n-1})
\end{aligned}$$

Considering the above two relationships together, we have

$$\begin{aligned}
\frac{P(\xi_{t+1} \in X_{n-1})}{1 - \mu_{t+1}} &< \frac{P(\xi_t \in X_{n-1}) - P(\xi_{t+1} \in X^* \mid \xi_t \in X_{n-1})}{1 - \mu_t - P(\xi_{t+1} \in X^* \mid \xi_t \in X_{n-1})} \\
&\leqslant \frac{P(\xi_t \in X_{n-1})}{1 - \mu_t} \quad \text{\%\% by } \forall t\colon 1 - \mu_t \geqslant P(\xi_t \in X_{n-1}) \geqslant 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_{x \notin X^*} P(\xi_{t+1} \in X^* \mid \xi_t = x) \frac{P(\xi_t = x)}{1 - \mu_t} \\
&\quad \leqslant \sum_{x \notin X^*} P(\xi_1 \in X^* \mid \xi_0 = x) \frac{P(\xi_0 = x)}{1 - \mu_0} \quad \text{\%\% by } \frac{P(\xi_{t+1} \in X_i)}{1 - \mu_{t+1}} \leqslant \frac{P(\xi_t \in X_i)}{1 - \mu_t} \\
&\quad = \frac{1}{2^n}.
\end{aligned}$$

So, let $\beta_t = \frac{1}{2^n}$, by Theorem 1, the EFHT is lower bounded

$$\mathbb{E}[\tau] \geqslant 2^n,$$

that is, $\mathbb{E}[\tau] = \Omega(2^n)$. $\square$

4.2. Static mutation with population

Now we study how the three static mutation operators perform with population size larger than 1. Specifically, let us consider the case where the population size is equal to the problem size n, i.e. $(n+n)$-EA, which is a practical strategy. In this case, the population state space consists of solution state spaces, which means a population $x \in X$ contains n solutions from solution space $\mathcal{S}$.

We can consider each population as a set of solutions without order. Denoting '$(\cdots)$' as a solution and '$\{\cdots\}$' as a population set, the two populations, which consist of the same solutions with different orders, are equal, e.g., $\{(001),(011),(110)\} = \{(110),(011),(001)\}$. By this consideration, there are $|X| = \binom{2^n+n-1}{n}$ number of population states [25]. But if we generate a population by generating each bit of each solution independently from an uniform distribution, we will have different probabilities to choose different states, e.g., $P(\{(001),(001),(001)\}) = 0.5^9$ but $P(\{(001),(011),(110)\}) = 6 \times 0.5^9$. Meanwhile, we can equivalently consider each population as an ordered set of solutions. Denoting '$[\cdots]$' as a population with order, the two populations, which consist of the same solutions with different orders are unequal, e.g., $[(001),(011),(110)] \neq [(110),(011),(001)]$. By this consideration, there are $|X| = 2^{n\times n}$ number of different population states, and the probability of randomly generating every population is exactly $1/|X|$. We use the second consideration in the follows, such that the calculation will be simple.

Proposition 4. *Solving the Trap problem using the EA with Reproduction implemented by Mutation#1* (*bitwise mutation with constant probability*) *and with a population size equals to the problem size, i.e.* $(n+n)$*-EA, if starting from non-optimal populations, the EFHT is bounded by*

$$\mathbb{E}[\tau] = \Omega\left(\frac{\theta^n}{n}\right), \tag{17}$$

where $\theta = (1-p_m)^{-1} \in (1,2]$ *is a constant and n is the problem size.*

The proof of this proposition is the same as of Proposition 1, except that the state level is upgraded to the population states. We know that the maximum probability of a solution being mutated to be the optimal solution by Mutation#1 is $p_m(1-p_m)^{n-1}$, which leads to that the maximum probability of a population being mutated to be an optimal population is $1-(1-p_m(1-p_m)^{n-1})^n$. Therefore, we can have an upper bound of formula (12). By Theorem 1, we get this proposition.

Proof. Since $P(\xi_{t+1} \in X^* \mid \xi_t = x) \leqslant 1-(1-p_m(1-p_m)^{n-1})^n$, we have

$$\begin{aligned}
&\sum_{x\notin X^*} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big)\frac{P(\xi_t = x)}{1-\mu_t} \\
&\quad \leqslant \sum_{x\notin X^*} p_m(1-p_m)^{n-1}\frac{P(\xi_t = x)}{1-\mu_t} \\
&\quad = \big(1-\big(1-p_m(1-p_m)^{n-1}\big)^n\big)\frac{\sum_{x\notin X^*} P(\xi_t = x)}{1-\mu_t} \\
&\quad = 1-\big(1-p_m(1-p_m)^{n-1}\big)^n \quad \text{\%\% by Eq. (1)} \\
&\quad \sim np_m(1-p_m)^{n-1}. \quad \text{\%\% asymptotically equal}
\end{aligned}$$

Let $\beta_t = np_m(1-p_m)^{n-1}$, by Theorem 1,

$$\mathbb{E}[\tau] \geqslant \beta_0 + \sum_{t=2}^{+\infty} t\beta_{t-1}\prod_{i=0}^{t-2}(1-\beta_i) = \frac{1}{n}\frac{1}{p_m}\left(\frac{1}{1-p_m}\right)^{n-1} = \frac{\theta^n}{np_m}.$$

Considering that p_m is a constant, $\mathbb{E}[\tau] = \Omega(\frac{\theta^n}{n})$. □

Proposition 5. *Solving the Trap problem using the EA with Reproduction implemented by Mutation#2* (*bitwise mutation with probability* $1/n$) *and with a population size equals to the problem size, i.e.* $(n+n)$*-EA, if starting from non-optimal populations, the EFHT is bounded by*

$$\mathbb{E}[\tau] = \Omega\left(\frac{2^n}{n^2}\right), \tag{18}$$

where n is the problem size.

Following the proof of Proposition 2, we divide the population state space $X = \{0,1\}^{n*n}$ into $n+1$ subspaces $\{X_i\}_{i=0}^n$, where X_n contains all optimal populations, X_i ($i \in \{0,\ldots,n-1\}$) contains non-optimal populations, and the solution with the *worst* fitness in each X_i has i identical bits with the optimal solution. By doing so, the state subspaces hold the same properties as those in the proof of Proposition 2, which leads to the calculation of formula (12). The only difference from the proof of Proposition 2 is that the upper bound of formula (12) is calculated at the population level but not at the solution level, which results in $O(\frac{n^2}{2^n})$. Therefore a lower bound $\Omega(\frac{2^n}{n^2})$ is obtained by Theorem 1.

Proof. Let

$$X_i = \left\{x \in X \mid \min_{s\in x} \|s - s^*\|_H = n - i\right\},$$

where $\|\cdot\|_H$ is Hamming distance, x denotes a population, s denotes a solution, and s^* is the optimal solution.

Denote

$$\tilde{X}_i = \left\{x \in X_i \mid \forall s \in x\colon \|s - s^*\|_H = n - i\right\}.$$

By applying Mutation#2, the success probability is

$$\begin{aligned}
\forall x \in X_i\colon \quad & P\left(\xi_{t+1} \in X^* \mid \xi_t = x\right) \leqslant P\left(\xi_{t+1} \in X^* \mid \xi_t \in \tilde{X}_i\right) \\
& \quad \text{%% by considering the worst population in the subspace } X_i \\
& = 1 - \left(1 - \left(\frac{1}{n}\right)^{n-i}\left(1 - \frac{1}{n}\right)^{i}\right)^{n} \\
& \sim n\left(\frac{1}{n}\right)^{n-i}\left(1 - \frac{1}{n}\right)^{i}. \quad \text{%% asymptotically equal}
\end{aligned}$$

Because all the solutions in every population of subspace X_i have at most i bits identical to the optimal solution, and there is at least one solution holding the exact i bits of that, we have the probability at the initialization that

$$P(\xi_0 \in X_i) = \binom{n}{1} \cdot \left(\binom{n}{i}\frac{1}{2^n}\right) \cdot \left(\sum_{j=0}^{i}\binom{n}{j}\frac{1}{2^n}\right)^{n-1} \leqslant n\binom{n}{i}\frac{1}{2^n}.$$

At $t = 0$, we have

$$\begin{aligned}
& \sum_{x \notin X^*} P\left(\xi_1 \in X^* \mid \xi_0 = x\right)\frac{P(\xi_0 = x)}{1 - \mu_0} \\
& \quad = \sum_{x \notin X^*} P\left(\xi_1 \in X^* \mid \xi_0 = x\right)P(\xi_0 = x) \quad \text{%% by assumption } \mu_0 = 0 \\
& \quad = \sum_{i=0}^{n-1}\sum_{x \in X_i}\left(P\left(\xi_1 \in X^* \mid \xi_0 = x\right)P(\xi_0 = x)\right) \quad \text{%% by } X = \bigcup_{i=0}^{n} X_i \\
& \quad \leqslant \sum_{i=0}^{n-1} n^2\left(\frac{1}{n}\right)^{n-i}\left(1 - \frac{1}{n}\right)^{i}\binom{n}{i}\frac{1}{2^n} \\
& \quad = n^2\left(1 - \left(\frac{n-1}{n}\right)^{n}\right)\frac{1}{2^n} \\
& \quad \sim \frac{e-1}{e}\frac{n^2}{2^n}.
\end{aligned}$$

Let the solution space S be divided into three subspaces $S = S^* \cup S_F \cup S_I$, where S^* contains the optimal solution, S_F contains all the non-optimal feasible solutions whose last bit is 0, and S_I contains all the infeasible solutions whose last bit is 1. Denote

$$X_F = \{x \in X \mid \forall s \in x\colon s \in S_F\}, \qquad X_I = X - X_F - X^*$$

and then denote

$$X_i^F = X_i \cap X_F, \qquad \tilde{X}_i^F = \tilde{X}_i \cap X_F.$$

According to the selection behavior, i.e., the solutions with the largest fitness will be selected, we have

$$\begin{aligned}
& \forall j, q (n-1 \geqslant j \geqslant q \geqslant 0)\colon \quad P\left(\xi_{t+1} \in X_j^F \mid \xi_t \in X_q^F\right) = 0, \qquad P\left(\xi_{t+1} \in X_I \mid \xi_t \in X_F\right) = 0, \\
& \forall j, q\colon \quad \frac{P(\xi_{t+1} \in X_j^F \mid \xi_t \in X_I)}{P(\xi_{t+1} \in X_q^F \mid \xi_t \in X_I)} = \frac{P(\xi_0 \in X_j^F)}{P(\xi_0 \in X_q^F)},
\end{aligned}$$

where the last equation is by that, since every infeasible solution has the same lowest fitness, there is no selection pressure on the leading $n-1$ bits when there is at least one solution in the current population is infeasible, and thus each of the leading $n-1$ bits has probability 0.5 to be either zero or one.

For all $k \in \{0, 1, \dots, n-1\}$, denoting $X_{A_k} = \bigcup_{i=0}^{k} X_i^F$ and $X_{B_k} = \bigcup_{i=k+1}^{n-1} X_i^F$, at time 0, for all $k \in \{0, 1, \dots, n-1\}$, there exists $\eta_{A_k,t}, \eta_{B_k,t}, \eta_{F,t}$ and $\eta_{I,t}$ such that

$$\sum_{x \in X_{A_k}} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) P(\xi_t = x) = \eta_{A_k,t} P(\xi_t \in X_{A_k}),$$

$$\sum_{x \in X_{B_k}} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) P(\xi_t = x) = \eta_{B_k,t} P(\xi_t \in X_{B_k}),$$

$$\sum_{x \in X_F} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) P(\xi_t = x) = \eta_{F,t} P(\xi_t \in X_F),$$

$$\sum_{x \in X_I} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) P(\xi_t = x) = \eta_{I,t} P(\xi_t \in X_I).$$

On the relationship between $\eta_{A_k,t}$ and $\eta_{B_k,t}$, it holds that

$$\eta_{A_k,t} < \eta_{B_k,t},$$

by $P(\xi_{t+1} \in X^* \mid \xi_t \in X_{A_k}) \leqslant P(\xi_{t+1} \in X^* \mid \xi_t \in X_{B_k})$, which is by

$$P\big(\xi_{t+1} \in X^* \mid \xi_t \in X_{A_k}\big) \leqslant P\big(\xi_{t+1} \in X^* \mid \xi_t \in \tilde{X}_k^F\big)$$

according to the definition of $\tilde{X}_k^F$, and

$$P\big(\xi_{t+1} \in X^* \mid \xi_t \in \tilde{X}_k^F\big) \leqslant P\big(\xi_{t+1} \in X^* \mid \xi_t \in X_{B_k}\big)$$

when $n \to +\infty$.

On the relationship between $\eta_{A_k,t}$ and $\eta_{I,t}$, it holds that at $t = 0$,

$$\eta_{A_k,0} < \eta_{I,0},$$

by, first, $\eta_{A_k,0} < \eta_{F,0}$, which is by $\eta_{F,0} = \eta_{A_k,0} \frac{P(\xi_0 \in X_A)}{P(\xi_0 \in X_F)} + \eta_{B_k,0} \frac{P(\xi_0 \in X_B)}{P(\xi_0 \in X_F)}$ and $\eta_{A_k,0} < \eta_{B_k,0}$, and second, $\eta_{F,0} < \eta_{I,0}$, which is by

$$\forall x_1, x_2 \in X: \quad P(\xi_0 = x_1) = P(\xi_0 = x_2) \quad \text{and} \quad \frac{\sum_{x \in X_F} P(\xi_1 \in X^* \mid \xi_0 = x)}{P(\xi_0 \in X_F)} < \frac{\sum_{x \in X_I} P(\xi_1 \in X^* \mid \xi_0 = x)}{P(\xi_0 \in X_I)}.$$

And for $t > 0$,

$$\eta_{A_k,t} < \eta_{I,0},$$

by that, since

$$\begin{aligned}
P(\xi_{t+1} \in X_{A_k}) &= P(\xi_t \in X_{A_k}) + \sum_{x \in X_{B_k} \cup X_I} P(\xi_{t+1} \in X_{A_k} \mid \xi_t = x) P(\xi_t = x) - \sum_{x \in X_{A_k}} P\big(\xi_{t+1} \in X^* \cup X_{B_k} \cup X_I \mid \xi_t = x\big) P(\xi_t = x) \\
&= P(\xi_t \in X_{A_k}) + \sum_{x \in X_{B_k} \cup X_I} P(\xi_{t+1} \in X_{A_k} \mid \xi_t = x) P(\xi_t = x) - \sum_{x \in X_{A_k}} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) P(\xi_t = x) \\
&\qquad \text{\%\% by } P(\xi_{t+1} \in X_{B_k} \cup X_I \mid \xi_t \in X_{A_k}) = 0 \\
&> \big(1 - P\big(\xi_{t+1} \in X^* \mid \xi_t \in X_k^F\big)\big) P(\xi_t \in X_{A_k}) + \sum_{x \in X_I} P(\xi_{t+1} \in X_{A_k} \mid \xi_t = x) P(\xi_t = x) \\
&\qquad \text{\%\% by } -P\big(\xi_{t+1} \in X^* \mid \xi_t \in X_A\big) \geqslant -P\big(\xi_{t+1} \in X^* \mid \xi_t \in X_k^F\big) \\
&> P(\xi_0 \in X_{A_k}) \prod_{i=0}^{t} \big(1 - P(\xi_{i+1} \in X^* \mid \xi_i \in X_k^F)\big) \\
&\quad + \sum_{i=0}^{t} \left(\sum_{x \in X_I} P\big(\xi_{i+1} \in X_{A_k} \mid \xi_i = x\big) P(\xi_i = x) \right) \left(\prod_{j=i}^{t} \big(1 - P(\xi_{j+1} \in X^* \mid \xi_j \in X_k^F)\big) \right) \\
&= P(\xi_0 \in X_{A_k}) \prod_{i=0}^{t} \big(1 - P(\xi_{i+1} \in X^* \mid \xi_i \in X_k^F)\big) \\
&\quad + \sum_{i=0}^{t} P\big(\xi_{i+1} \in X_{A_k} \mid \xi_i \in X_I\big) P(\xi_i \in X_I) \left(\prod_{j=i}^{t} (1 - P(\xi_{j+1} \in X^* \mid \xi_j \in X_k^F)) \right),
\end{aligned}$$

and similarly,

$$P(\xi_{t+1} \in X_{B_k}) < P(\xi_0 \in X_{B_k}) \prod_{i=0}^{t} \left(1 - P(\xi_{i+1} \in X^* \mid \xi_i \in X_k^F)\right)$$
$$+ \sum_{i=0}^{t} P\left(\xi_{i+1} \in X_{B_k} \mid \xi_i \in X_I\right) P(\xi_i \in X_I) \left(\prod_{j=i}^{t} \left(1 - P(\xi_{j+1} \in X^* \mid \xi_j \in X_k^F)\right) \right),$$

and $\frac{P(\xi_{t+1} \in X_j^F \mid \xi_t \in X_I)}{P(\xi_{t+1} \in X_q^F \mid \xi_t \in X_I)} = \frac{P(\xi_0 \in X_j^F)}{P(\xi_0 \in X_q^F)}$, we thus have $\forall k$: $\frac{P(\xi_{t+1} \in X_{A_k})}{P(\xi_{t+1} \in X_{B_k})} > \frac{P(\xi_0 \in X_{A_k})}{P(\xi_0 \in X_{B_k})}$; since $P(\xi_{t+1} \in X^* \mid \xi_t \in X_{A_k}) < P(\xi_{t+1} \in X^* \mid \xi_t \in X_{B_k})$, by enumerating k, we have

$$\frac{\sum_{x \in X_{A_k}} P(\xi_{t+1} \in X^* \mid \xi_t = x) P(\xi_t = x)}{P(\xi_t \in X_{A_k})} < \frac{\sum_{x \in X_{A_k}} P(\xi_1 \in X^* \mid \xi_t = x) P(\xi_0 = x)}{P(\xi_0 \in X_{A_k})},$$

which is $\eta_{A_k,t} < \eta_{A_k,0}$, and by $\eta_{A_k,0} < \eta_{I,0}$, it holds $\eta_{A_k,t} < \eta_{I,0}$.

Then, we have

$$\begin{aligned}
\frac{P(\xi_{t+1} \in X_{A_k})}{1 - \mu_{t+1}} &> \frac{P(\xi_t \in X_{A_k}) - \sum_{x \in X_{A_k}} P(\xi_{t+1} \in X^* \mid \xi_t = x) P(\xi_t = x)}{1 - \mu_t - \sum_{x \in X_{A_k} \cup X_{B_k} \cup X_I} P(\xi_{t+1} \in X^* \mid \xi_t = x) P(\xi_t = x)} \\
&= \frac{(1 - \eta_{A_k,t}) P(\xi_t \in X_{A_k})}{1 - \mu_t - \eta_{A_k,t} P(\xi_t \in X_{A_k}) - \eta_{B_k,t} P(\xi_t \in X_{B_k}) - \eta_{I,t} P(\xi_t \in X_I)} \\
&> \frac{(1 - \eta_{A_k,t}) P(\xi_t \in X_{A_k})}{1 - \mu_t - \eta_{A_k,t} P(\xi_t \in X_{A_k}) - \eta_{A_k,t} P(\xi_t \in X_{B_k}) - \eta_{A_k,t} P(\xi_t \in X_I)} \\
&\qquad \text{\%\% by } \eta_{A_k,t} < \eta_{B_k,t},\ \eta_{A_k,t} < \eta_{I,0} \text{ and } \eta_{I,t} = \eta_{I,0} \\
&= \frac{(1 - \eta_{A_k,t}) P(\xi_t \in X_{A_k})}{(1 - \eta_{A_k,t})(1 - \mu_t)} \\
&= \frac{P(\xi_t \in X_{A_k})}{1 - \mu_t},
\end{aligned}$$

which is $\forall n-1 \geqslant k \geqslant 0 : \frac{P(\xi_{t+1} \in \bigcup_{i=0}^k X_i^F)}{1-\mu_{t+1}} \geqslant \frac{P(\xi_t \in \bigcup_{i=0}^k X_i^F)}{1-\mu_t}$ by writing back $X_{A_k} = \bigcup_{i=0}^k X_i^F$. So, we have

$$\begin{aligned}
&\sum_{x \notin X^*} P\left(\xi_{t+1} \in X^* \mid \xi_t = x\right) \frac{P(\xi_t = x)}{1 - \mu_t} \\
&= \sum_{x \in X_F} P\left(\xi_{t+1} \in X^* \mid \xi_t = x\right) \frac{P(\xi_t = x)}{1 - \mu_t} + \sum_{x \in X_I} P\left(\xi_{t+1} \in X^* \mid \xi_t = x\right) \frac{P(\xi_t = x)}{1 - \mu_t} \\
&\leqslant \sum_{x \in X_F} P\left(\xi_1 \in X^* \mid \xi_0 = x\right) \frac{P(\xi_0 = x)}{1 - \mu_0} + \sum_{x \in X_I} P\left(\xi_1 \in X^* \mid \xi_0 = x\right) \frac{P(\xi_0 = x)}{1 - \mu_0} \\
&\qquad \text{\%\% by both that } \forall n-1 \geqslant k \geqslant 0\text{:}\ \frac{P(\xi_{t+1} \in \bigcup_{i=0}^k X_i^F)}{1 - \mu_{t+1}} \geqslant \frac{P(\xi_t \in \bigcup_{i=0}^k X_i^F)}{1 - \mu_t} \\
&\qquad \text{\%\% and } P\left(\xi_{t+1} \in X^* \mid \xi_t \in \bigcup_{i=0}^{k} X_i^F\right) \leqslant P\left(\xi_{t+1} \in X^* \mid \xi_t \in \bigcup_{i=k+1}^{n-1} X_i^F\right) \\
&= \sum_{x \notin X^*} P\left(\xi_1 \in X^* \mid \xi_0 = x\right) \frac{P(\xi_0 = x)}{1 - \mu_0} \\
&= \frac{e-1}{e} \frac{n^2}{2^n}.
\end{aligned}$$

Let $\beta_t = \frac{e-1}{e} \frac{n^2}{2^n}$, by Theorem 1,

$$\mathbb{E}[\tau] \geqslant \frac{e}{e-1} \frac{2^n}{n^2},$$

that is, $\mathbb{E}[\tau] = \Omega(\frac{2^n}{n^2})$. □

Proposition 6. *Solving the Trap problem using the EA with Reproduction implemented by Mutation#3 (one-bit mutation) and with a population size equals to the problem size, i.e. $(n+n)$-EA, if starting from non-optimal populations, the EFHT is bounded by*

$$\mathbb{E}[\tau] = \Omega\left(\frac{2^n}{n^2}\right), \tag{19}$$

where n is the problem size.

As in the proof of Proposition 5, we can divide the population state space $X = \{0,1\}^{n*n}$ into $n+1$ subspaces $\{X_i\}_{i=0}^n$, where X_n contains all the optimal populations, and X_i $(i \in \{0,\dots,n-1\})$ contains non-optimal populations, among which the solution with the *worst* fitness in the population has $n-i$ bits different from the optimal solution. By doing so, the state subspaces hold the same properties as those in the proof of Proposition 2, which leads to the calculation of formula (12). The only difference from the proof of Proposition 2 is that the upper bound of formula (12) is calculated at the population level but not at the solution level, which results in $O(\frac{n^2}{2^n})$. Therefore a lower bound $\Omega(\frac{2^n}{n^2})$ is obtained by Theorem 1.

Proof. Let

$$X_i = \left\{x \in X \mid \min_{s \in x} \|s - s^*\|_H = n - i\right\},$$

where $\|\cdot\|_H$ is Hamming distance, x denotes a population, s denotes a solution, and s^* is the optimal solution.

By applying the one-bit mutation, the success probability is

$$P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) \begin{cases} \leqslant 1 - (1-\frac{1}{n})^n, & x \in X_{n-1}, \\ = 0, & x \in X - X_{n-1} - X^*, \end{cases}$$

by considering that only populations in X_{n-1} have chance to mutate to be optimal, and that the best case that all solutions in X_{n-1} have only 1 bit different from the optimal solution. Since there are n solutions that have one bit different from the optimal solution, the probability of being in X_{n-1} at initialization is

$$P(\xi_0 \in X_{n-1}) = 1 - \left(1 - \frac{n}{2^n}\right)^n.$$

Noticing that X_{n-1} consists of populations that contain either the feasible solution that has the lowest fitness among feasible solutions, or $n-1$ infeasible solutions that are with the lowest fitness among all solutions, we have

$$P\big(\xi_{t+1} \in X_{n-1} \mid \xi_t \in X - X_{n-1} - X^*\big) < P\big(\xi_{t+1} \in X - X_{n-1} - X^* \mid \xi_t \in X_{n-1}\big).$$

At $t = 0$, we have

$$\begin{aligned}
&\sum_{x \notin X^*} P\big(\xi_1 \in X^* \mid \xi_0 = x\big)\frac{P(\xi_0 = x)}{1-\mu_0} \\
&\quad = \sum_{x \notin X^*} P\big(\xi_1 \in X^* \mid \xi_0 = x\big)P(\xi_0 = x) \qquad \text{\%\% by assumption } \mu_0 = 0 \\
&\quad = \sum_{x \in X_{n-1}} P\big(\xi_1 \in X^* \mid \xi_0 = x\big)P(\xi_0 = x) \\
&\qquad\quad \text{\%\% by } P\big(\xi_{t+1} \in X^* \mid \xi_t \in X - X_{n-1} - X^*\big) = 0 \\
&\quad \leqslant \left(1 - \left(1 - \frac{1}{n}\right)^n\right)\left(1 - \left(1 - \frac{n}{2^n}\right)^n\right) \\
&\quad \sim \frac{n^2}{2^n}.
\end{aligned}$$

At time $t+1$, on the relationship between μ_t and μ_{t+1}, we have

$$\begin{aligned}
\mu_{t+1} &= \mu_t + \sum_{x \in (X - X_{n-1} - X^*)} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big)P(\xi_t = x) + \sum_{x \in X_{n-1}} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big)P(\xi_t = x) \\
&= \mu_t + \sum_{x \in X_{n-1}} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big)P(\xi_t = x) \\
&\qquad \text{\%\% by } P\big(\xi_{t+1} \in X^* \mid \xi_t \in X - X_{n-1} - X^*\big) = 0.
\end{aligned}$$

On the relationship between $P(\xi_t \in X_{n-1})$ and $P(\xi_{t+1} \in X_{n-1})$, we have

$$\begin{aligned}
P(\xi_{t+1} \in X_{n-1}) &= P(\xi_t \in X_{n-1}) + \sum_{x \in X - X_{n-1} - X^*} P(\xi_{t+1} \in X_{n-1} \mid \xi_t = x) P(\xi_t = x) \\
&\quad - \sum_{x \in X_{n-1}} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) P(\xi_t) - \sum_{x \in X_{n-1}} P\big(\xi_{t+1} \in X - X_{n-1} - X^* \mid \xi_t = x\big) P(\xi_t = x) \\
&< P(\xi_t \in X_{n-1}) - \sum_{x \in X_{n-1}} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) P(\xi_t) \\
&\qquad \text{\%\% by } P\big(\xi_{t+1} \in X_{n-1} \mid \xi_t \in X - X_{n-1} - X^*\big) < P\big(\xi_{t+1} \in X - X_{n-1} - X^* \mid \xi_t \in X_{n-1}\big).
\end{aligned}$$

Considering the above two relationships together, we have

$$\begin{aligned}
\frac{P(\xi_{t+1} \in X_{n-1})}{1 - \mu_{t+1}} &< \frac{P(\xi_t \in X_{n-1}) - \sum_{x \in X_{n-1}} P(\xi_{t+1} \in X^* \mid \xi_t = x) P(\xi_t)}{1 - \mu_t - \sum_{x \in X_{n-1}} P(\xi_{t+1} \in X^* \mid \xi_t = x) P(\xi_t = x)} \\
&\leqslant \frac{P(\xi_t \in X_{n-1})}{1 - \mu_t}. \quad \text{\%\% by } \forall t: 1 - \mu_t \geqslant P(\xi_t \in X_{n-1}) \geqslant 0
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\sum_{x \notin X^*} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) \frac{P(\xi_t = x)}{1 - \mu_t} \\
&\quad \leqslant \sum_{x \notin X^*} P\big(\xi_1 \in X^* \mid \xi_0 = x\big) \frac{P(\xi_0 = x)}{1 - \mu_0} \quad \text{\%\% by } \frac{P(\xi_{t+1} \in X_{n-1})}{1 - \mu_{t+1}} \leqslant \frac{P(\xi_t \in X_{n-1})}{1 - \mu_t} \\
&\quad \leqslant \frac{n^2}{2^n}.
\end{aligned}$$

So, let $\beta_t = \frac{n^2}{2^n}$, by Theorem 1, the EFHT is lower bounded

$$\mathbb{E}[\tau] \geqslant \frac{2^n}{n^2},$$

that is, $\mathbb{E}[\tau] = \Omega(\frac{2^n}{n^2})$. □

4.3. Mutation and recombination with population

We implement the Reproduction by a mutation operator and a recombination operator together as follows, where ξ_{t+1}^M and ξ_{t+1}^C are defined as the sets of solutions produced by mutation and recombination from ξ_t, respectively:

1. the population ξ_t contains n solutions,
2. apply the mutation on ξ_t to produce another n solutions ξ_{t+1}^M,
3. apply the recombination on ξ_t to produce n more solutions ξ_{t+1}^C,
4. choose the best n solutions from $\xi_t \cup \xi_{t+1}^M \cup \xi_{t+1}^C$.

Considering the fitness function of the Trap problem, when the last bit of a solution is 1, it is either the optimal solution or a solution with the worst fitness. We can divide the solution space $\mathcal{S}$ into $\mathcal{S}_F$ and $\mathcal{S}_I$, where the last bit of solutions in $\mathcal{S}_F$ is 0, and the last bit of solutions in $\mathcal{S}_I$ is 1. This separation is helpful for our analysis on recombination. We first give a lemma, based on which the lower bounds of three EAs are then derived.

Lemma 3. *Solving the Trap problem by the EA with population size equals to the problem size, i.e. $(n+n)$-EA. Let $\Phi : X \to Z$ be a function that cuts off the last bit of all solutions in the population from X, and Z is another population space $Z = \{0,1\}^{n*(n-1)}$ where each population contains n solutions of $n-1$ bits long. Then, as long as there is at least one solution in the current population ξ_t is from $\mathcal{S}_I$, we have*

$$\forall \tilde{Z} \subseteq Z: \quad P\big(\Phi(\xi_t) \in \tilde{Z}\big) = \frac{|\tilde{Z}|}{|Z|}.$$

This lemma tells that how the leading $n-1$ bits of solutions in the population distribute, given that there is at least one solution whose last bit is 1. To prove the lemma, first, we find that, by initialization, each of the leading $n-1$ bits of every solution has probability 0.5 to be either zero or one. Second, we notice that applying of the mutation and the recombination operators does not change that distribution. Finally, the selection does not change the distribution of those $n-1$ bits so far as at least one solution from $\mathcal{S}_I$ remains in the population after the selection operation. Thus, the lemma is proved.

Proof. At $t = 0$, we have

$$\forall \tilde{Z} \subseteq Z: \quad P\big(\Phi(\xi_0) \in \tilde{Z}\big) = \frac{|\tilde{Z}|}{|Z|},$$

due to random initialization.

At time $t + 1$, the mutation operators and the recombination operator are all symmetric, i.e.,

$$\begin{cases} P(\xi^M_{t+1} = y \mid \xi_t = x) = P(\xi^M_{t+1} = x \mid \xi_t = y), \\ P(\xi^R_{t+1} = y \mid \xi_t = x) = P(\xi^R_{t+1} = x \mid \xi_t = y), \end{cases}$$

where ξ^M_{t+1} and ξ^R_{t+1} denote the populations of ξ_t after the mutation and the recombination operators, respectively.

By applying an arbitrary symmetric operator, i.e., one for which it holds $P(\xi_{t+1} = y \mid \xi_t = x) = P(\xi_{t+1} = x \mid \xi_t = y)$ for arbitrary $x \in X$ and $y \in X$, we have

$$\begin{aligned} \forall y: \quad & P(\xi_1 = y) - P(\xi_0 = y) \\ &= \sum_{x \in X - \{y\}} P(\xi_1 = y \mid \xi_0 = x) P(\xi_0 = x) - \sum_{x \in X - \{y\}} P(\xi_1 = x \mid \xi_0 = y) P(\xi_0 = y) \\ &= \sum_{x \in X - \{y\}} P(\xi_1 = y \mid \xi_0 = x)(P(\xi_0 = x) - P(\xi_0 = y)) \\ &\qquad \%\% \text{ by } P(\xi_1 = y \mid \xi_0 = x) = P(\xi_1 = x \mid \xi_0 = y) \\ &= \sum_{x \in X - \{y\}} P(\xi_1 = y \mid \xi_0 = x)\left(\frac{1}{|X|} - \frac{1}{|X|}\right) \\ &= 0. \end{aligned}$$

Therefore, we have

$$\forall \tilde{X} \subseteq X: \quad P\big(\xi^M_1 \in \tilde{X}\big) = \frac{|\tilde{X}|}{|X|}, \qquad P\big(\xi^C_1 \in \tilde{X}\big) = \frac{|\tilde{X}|}{|X|}$$

for the mutation operators and the recombination operator, respectively. This leads to

$$\forall \tilde{Z} \subseteq Z: \quad P\big(\Phi\big(\xi^M_1\big) \in \tilde{Z}\big) = \frac{|\tilde{Z}|}{|Z|}, \qquad P\big(\Phi\big(\xi^C_1\big) \in \tilde{Z}\big) = \frac{|\tilde{Z}|}{|Z|}$$

by noticing the universal quantifier $\forall$.

The Selection operation generates ξ_{t+1} by choosing the best n solutions from $\xi_t \cup \xi^M_{t+1} \cup \xi^R_{t+1}$. Since there is at least one solution from $\mathcal{S}_I$ survives in ξ_{t+1}, we have

$$\forall s \in \mathcal{S}_F: \quad P\big(s \in \xi_{t+1} \mid s \in \xi_t \cup \xi^M_{t+1} \cup \xi^R_{t+1}\big) = 1,$$

by considering that every solution in $\mathcal{S}_F$ has a better fitness value than all the non-optimal solutions in $\mathcal{S}_I$. Since all the non-optimal solutions in $\mathcal{S}_I$ have the same fitness values, they have the same probability to survive in ξ_{t+1}, i.e.,

$$\forall s_1, s_2 \in \mathcal{S}_I: \quad P\big(s_1 \in \xi_{t+1} \mid s_1 \in \xi_t \cup \xi^M_{t+1} \cup \xi^R_{t+1}\big) = P\big(s_2 \in \xi_{t+1} \mid s_2 \in \xi_t \cup \xi^M_{t+1} \cup \xi^R_{t+1}\big).$$

Therefore,

$$\forall \tilde{Z} \in Z: \quad P\big(\Phi(\xi_{t+1}) \in \tilde{Z}\big) = \frac{|\tilde{Z}|}{|Z|}. \qquad \square$$

Proposition 7. *Solving the Trap problem using the EA with a population size equals to the problem size, i.e.* $(n + n)$*-EA, if starting from non-optimal populations,*

(a) *if the Reproduction is implemented by Mutation#1 (bitwise mutation with constant probability) and Recombination, the EFHT is bounded by*

$$\mathbb{E}[\tau] = \Omega\left(\frac{2^n}{n^3}\right), \tag{20}$$

(b) *if the Reproduction is implemented by Mutation#2 (bitwise mutation with probability* $1/n$*) and Recombination, the EFHT is bounded by*

$$\mathbb{E}[\tau] = \Omega\left(\frac{2^n}{n^3}\right), \tag{21}$$

(c) *if the Reproduction is implemented by Mutation#3 (one-bit mutation) and Recombination, the EFHT is bounded by*

$$\mathbb{E}[\tau] = \Omega\left(\frac{2^n}{n^3}\right), \tag{22}$$

where n is the problem size.

To prove the proposition, we first notice that the recombination operator could not generate the optimal solution from a population that does not contain any solutions from $\mathcal{S}_I$, which derives an upper bound for when the recombination has non-zero success probability. Then we find that the probability of being in populations that contain solutions from $\mathcal{S}_I$ is decreasing. Thus we get an upper bound of formula (12), which leads to the proposition.

Proof. Considering that the mutation and recombination are applied independently and the Selection operation does not generate new solutions, we have

$$\begin{aligned}&\sum_{x\notin X^*} P\big(\xi_{t+1}\in X^* \mid \xi_t = x\big)\frac{P(\xi_t = x)}{1-\mu_t}\\&\quad = \sum_{x\notin X^*} P\big(\xi^M_{t+1}\in X^* \mid \xi_t = x\big)\frac{P(\xi_t = x)}{1-\mu_t} + \sum_{x\notin X^*} P\big(\xi^R_{t+1}\in X^* \mid \xi_t = x\big)\frac{P(\xi_t = x)}{1-\mu_t},\end{aligned}$$

where ξ^M_{t+1} is the population reproduced from ξ_{t+1} with only mutation, and ξ^R_{t+1} is that with only recombination. Let

$$\begin{aligned}&X_F = \{x \in X \mid \forall s \in x\colon\ s \in \mathcal{S}_F\},\\&X_I = X - X_F - X^*,\\&X_I^P = \big\{x \in X_I \mid P\big(\xi^R_{t+1} \in X^* \mid \xi_t = x\big) > 0\big\}.\end{aligned}$$

Then, we have

$$\begin{aligned}&\forall t\colon\quad P(\xi_{t+1} \in X_I \mid \xi_t \in X_F) = 0,\\&\forall t\colon\quad P\big(\xi^R_{t+1} \in X^* \mid \xi_t \in X_F\big) = 0,\end{aligned}$$

by considering the behavior of the selection and the recombination.

When $\xi_t \in X_F$, we have exactly the same results as with using mutation operators only, since the recombination is not useful.

When $\xi_t \in X_I$, for the mutation operators we have

$$\sum_{x\notin X^*} P\big(\xi^M_{t+1}\in X^* \mid \xi_t = x\big)\frac{P(\xi_t = x)}{1-\mu_t} \leqslant \begin{cases} np_m(1-p_m)^{n-1}, & \text{for Mutation\#1},\\ \frac{e-1}{e}\frac{n^2}{2^n}, & \text{for Mutation\#2},\\ \frac{n^2}{2^n}, & \text{for Mutation\#3}.\end{cases}$$

For the recombination operator, we have

$$\begin{aligned}&\sum_{x\in X_I} P\big(\xi^R_{t+1}\in X^* \mid \xi_t = x\big)\frac{P(\xi_t = x)}{1-\mu_t}\\&\quad = \sum_{x\in X_I^P} P\big(\xi^R_{t+1}\in X^* \mid \xi_t = x\big)\frac{P(\xi_t = x)}{1-\mu_t} \quad \text{\%\% by the definition of } X_I^P\\&\quad \leqslant \sum_{x\in X_I^P} P\big(\xi^R_{t+1}\in X^* \mid \xi_t = x\big)\frac{P(\Phi(\xi_t) = \Phi(x))(1-\mu_t)}{1-\mu_t}\\&\qquad \text{\%\% by optimistic assumption that the last bits of all solutions in } x \text{ are all } 1\\&\quad \leqslant \sum_{x\in X_I^P} P\big(\Phi(\xi_t) = \Phi(x)\big) \quad \text{\%\% by } P\big(\xi^R_{t+1}\in X^* \mid \xi_t = x\big) \leqslant 1\\&\quad = P\big(\Phi(\xi_t) \in \{\Phi(x) \mid x \in X_I^P\}\big)\\&\quad = \frac{|\{\Phi(x) | x\in X_I^P\}|}{|Z|}, \quad \text{\%\% by Lemma 3}\end{aligned}$$

where $\Phi(\cdot)$ and Z are defined in Lemma 3.

By the definition of X_I^P, every population in X_I^P should contain at least one solution whose leading d bits are the same as those of the optimal solution, and at least one solution whose tailing $n-1-d$ bits are the same as those of the optimal solution. Thus we have

$$\frac{|\{\Phi(x)|x \in X_I^P\}|}{|Z|} = \sum_{d=1}^{n-2}\binom{n}{1}\left(\left(\frac{1}{2}\right)^{d+1}\left(1-\left(1-\left(\frac{1}{2}\right)^{n-d}\right)^{n}\right)\right) \leqslant n\sum_{d=1}^{n-2}\left(\left(\frac{1}{2}\right)^{d+1} n\left(\frac{1}{2}\right)^{n-d}\right) = \frac{n^2(n-2)}{2^{n+1}}.$$

Then, we get

$$\sum_{x\in X_I} P\left(\xi_{t+1}^R \in X^* \mid \xi_t = x\right)\frac{P(\xi_t = x)}{1-\mu_t} \leqslant \frac{n^2(n-2)}{2^{n+1}}.$$

Therefore, for Mutation#1,

$$\sum_{x\in X} P\left(\xi_{t+1} \in X^* \mid \xi_t = x\right)\frac{P(\xi_t = x)}{1-\mu_t} \leqslant np_m(1-p_m)^{n-1} + \frac{n^2(n-2)}{2^{n+1}}.$$

Let $\beta = np_m(1-p_m)^{n-1} + \frac{n^2(n-2)}{2^{n+1}}$, the EFHT is obtained,

$$\mathbb{E}[\tau] = \left(np_m(1-p_m)^{n-1} + \frac{n^2(n-2)}{2^{n+1}}\right)^{-1} = \Omega\left(\frac{2^n}{n^3}\right).$$

For Mutation#2,

$$\sum_{x\in X} P\left(\xi_{t+1}^C \in X^* \mid \xi_t = x\right)\frac{P(\xi_t = x)}{1-\mu_t} \leqslant \frac{e-1}{e}\frac{n^2}{2^n} + \frac{n^2(n-2)}{2^{n+1}}.$$

Let $\beta = \frac{e-1}{e}\frac{n^2}{2^n} + \frac{n^2(n-2)}{2^{n+1}}$, the EFHT is obtained,

$$\mathbb{E}[\tau] = \left(\frac{e-1}{e}\frac{n^2}{2^n} + \frac{n^2(n-2)}{2^{n+1}}\right)^{-1} = \Omega\left(\frac{2^n}{n^3}\right).$$

For Mutation#3,

$$\sum_{x\in X} P\left(\xi_{t+1}^C \in X^* \mid \xi_t = x\right)\frac{P(\xi_t = x)}{1-\mu_t} \leqslant \frac{n^2}{2^n} + \frac{n^2(n-2)}{2^{n+1}}.$$

Let $\beta = \frac{n^2}{2^n} + \frac{n^2(n-2)}{2^{n+1}}$, the EFHT is obtained,

$$\mathbb{E}[\tau] = \left(\frac{n^2}{2^n} + \frac{n^2(n-2)}{2^{n+1}}\right)^{-1} = \Omega\left(\frac{2^n}{n^3}\right). \qquad \square$$

4.4. Time variant mutation

Proposition 8. *Solving the Trap problem using the EA with Reproduction implemented by Mutation#4 and with a population size* 1*, i.e.* $(1+1)$*-EA, if starting from non-optimal population, the EFHT is bounded by*

$$\mathbb{E}[\tau] = \Omega\left(\theta^n\right) \tag{23}$$

where $\theta \in (1, 2]$ *is a constant and n is the problem size.*

Since we model the evolution process by a non-homogeneous Markov chain, we can easily model the time-variant mutation, and simply reduce it to a homogeneous Markov chain to prove the proposition following the proof of Proposition 1.

Proof. Applying Mutation#4, for x such that $\|x - x^*\|_H = k$, we have

$$P\left(\xi_{t+1} \in X^* \mid \xi_t = x\right) = \left((0.5-d)e^{-t} + d\right)^k\left(1-(0.5-d)e^{-t} - d\right)^{n-k}.$$

Since $(0.5-d)e^{-t} + d \in (0, 0.5]$, we have

$$P\left(\xi_{t+1} \in X^* \mid \xi_t \in X - X^*\right) \leqslant \left((0.5-d)e^{-t} + d\right)\left(1-(0.5-d)e^{-t} - d\right)^{n-1}.$$

Thus,

$$\begin{aligned}
&\sum_{x \notin X^*} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) \frac{P(\xi_t = x)}{1-\mu_t} \\
&\quad \leqslant \big((0.5-d)e^{-t}+d\big)\big(1-(0.5-d)e^{-t}-d\big)^{n-1} \sum_{x \notin X^*} \frac{P(\xi_t = x)}{1-\mu_t} \\
&\quad = \big((0.5-d)e^{-t}+d\big)\big(1-(0.5-d)e^{-t}-d\big)^{n-1} \quad \text{\%\% by Eq. (1)} \\
&\quad \leqslant 0.5(1-d)^{n-1}. \quad \text{\%\% further relax}
\end{aligned}$$

Therefore, $\beta_t = 0.5(1-d)^{n-1}$, by Theorem 1, the EFHT is lower bounded

$$\mathbb{E}[\tau] \geqslant \beta_0 + \sum_{t=2}^{+\infty} t\beta_{t-1} \prod_{i=0}^{t-2}(1-\beta_i) = 2(1-d)^{-n+1} = 2\theta^{n-1},$$

that is, $\mathbb{E}[\tau] = \Omega(\theta^n)$. □

5. Discussion

In the previous section, we have proved that it needs exponential time to obtain the optimal solution to the Trap problem using several variations of EAs. To arrive the proof of that the Trap problem is hard to be solved by the EA using Mutation#1, we need only to bound the part of success probability of formula (12). In the same way, we can also prove that any problem with exponential size of solution space, even easy problems such as the OneMax problem, is hard for the EA using Mutation#1. This suggests that a non-adaptive mutation rate is not suitable for any problem which is with exponential size of solution space, and in those cases an adaptive mutation rate is preferred.

From the proofs for EAs using Mutations #2, #3 and Recombination, we find a common trick. At first the success probability at the initial step is exponentially small, then the EA goes toward a wrong direction which makes the success probability even lower. To disclose what is behind this trick, we re-investigate formula (12), i.e.,

$$\sum_{x \notin X^*} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) \frac{P(\xi_t = x)}{1-\mu_t} = \sum_{x \notin X^*} P\big(s^* \in \xi^M_{t+1} \cup \xi^R_{t+1} \mid \xi_t = x\big) \frac{P(\xi_t = x)}{1-\mu_t},$$

%% by that Selection does not generate solutions

where ξ^M_{t+1} and ξ^R_{t+1} are the populations reproduced by mutation and recombination operators from ξ_t, respectively, and s^* is the optimal solution.

The above formula consists of two parts. The success probability part

$$P\big(s^* \in \xi^M_{t+1} \cup \xi^R_{t+1} \mid \xi_t = x\big)$$

is determined by the Reproduction operators. Once the operators of the EA are fixed, the success probability is determined. In other words, this part is at the *algorithm* side. The normalized distribution part $\frac{P(\xi_t = x)}{(1-\mu_t)}$ is calculated by a recursive equation

$$P(\xi_{t+1} = x) = P(\xi_t = x) + \sum_{y \notin X^*} P(\xi_{t+1} = x \mid \xi_t = y)P(\xi_t = y) - \sum_{y \notin X^*} P(\xi_{t+1} = y \mid \xi_t = x)P(\xi_t = x),$$

in which, once the Reproduction operators have been fixed, the non-recursive terms $P(\xi_{t+1} = x \mid \xi_t = y)$ and $P(\xi_{t+1} = y \mid \xi_t = x)$ are determined by how solutions are favored, which is dominated by the fitness. Considering that when we apply an existing EA to tackle a problem, the Reproduction operator is fixed before we see the problem, while the fitness fully depends on the problem. So, the normalized distribution part is at the *problem* side.

So, our explanation to the question that what makes a problem hard to an EA is, on such problem the EA will run in a direction, which causes that the probability of being in good areas (i.e., having a large success probability) to decrease while that of being in bad areas (i.e., having a small success probability) to increase. In other words, the algorithm side mismatches the problem side.

Motivated by this recognition, we raise a question: how large is a problem class which, for any EA, contains at least one problem instance that cannot be solved by the EA in polynomial time? The answer to this question implies a general bound on the effectiveness of EAs.

At the first glance, this question seems related to the *No Free Lunch* theorem [26] which indicates that if all possible problems are considered, and if every problem instance has equal chance to be encountered (which is arguable), any two algorithms will have equal average performance. But note that the No Free Lunch theorem only considers whether one algorithm is relatively better than another algorithm, and an algorithm performs poor on a problem does not mean other algorithms will also perform poor on this problem. While, what we want to know is, whether there is some problem class which is hard for all EAs, that is, for any EA the problem class contains at least one hard problem instance.

We find that, for any EAs there must exist a problem instance which is hard in an exponential size *general problem*.

Definition 6 *(General problem).* A general problem with a solution space $\mathcal{S}$ and solution space size $|\mathcal{S}|$ is a finite set of problem instances, where every function in a bijective function cluster $\mathcal{F}:\mathcal{S}\to\{1,2,\ldots,|\mathcal{S}|\}$ corresponds to the fitness function of a problem instance.

In other words, a general problem contains $|\mathcal{S}|!$ number of problem instances, each is a permutation of solutions. Thus, every point in the solution space can be found as the optimal solution to an instance of the general problem.

We denote $\exp(n)$ as the exponential order of n, and $\mathrm{poly}(n)$ as the polynomial order of n, omitting the exact components of that order.

Theorem 2. *Given that*

(a) *a general problem with a solution space size no smaller than* $\exp(n)$,
(b) *an EA with a population size no larger than* $\mathrm{poly}(n)$,
(c) *every solution has an equal probability to appear in the initial population,*

there exists at least one problem instance on which the EFHT of the EA is no smaller than $\exp(n)/\mathrm{poly}(n)$, *where n is the problem size.*

Proof. Denote $\mathrm{Reprod}(\xi_t)$ as the solution set generated by Reproduction operators, of which the size is no more than $\mathrm{poly}(n)$ (otherwise it already costs $\exp(n)$ time). Considering that $|\mathcal{S}|=\exp(n)$ and population size is no more than $\mathrm{poly}(n)$, the population state space $|X|=\exp(n)^{\mathrm{poly}(n)}$. We have

$$\sum_{s\in\mathcal{S}} P\big(s\in\mathrm{Reprod}(\xi_t)\mid \xi_t=x\big)\leqslant \big|\mathrm{Reprod}(\xi_t)\big|\leqslant \mathrm{poly}(n)$$
$$\Rightarrow\ \sum_{s\in\mathcal{S}}\sum_{x\in X} P\big(s\in\mathrm{Reprod}(\xi_t)\mid \xi_t=x\big)\leqslant \mathrm{poly}(n)\exp(n)^{\mathrm{poly}(n)}$$
$$\Rightarrow\ \exists\tilde{s}\in\mathcal{S}:\quad \sum_{x\in X} P\big(\tilde{s}\in\mathrm{Reprod}(\xi_t)\mid \xi_t=x\big)\leqslant \mathrm{poly}(n)\exp(n)^{\mathrm{poly}(n)}\frac{1}{\exp(n)}.$$
%% otherwise the sum over $\mathcal{S}$ will exceed $\mathrm{poly}(n)\exp(n)^{\mathrm{poly}(n)}$

Let $\tilde{s}$ denote the optimal solution s^*.

At time $t=0$, we have

$$\sum_{x\notin X^*} P\big(\xi_1\in X^*\mid \xi_0=x\big)\frac{P(\xi_0=x)}{1-\mu_0}$$
$$=\sum_{x\notin X^*} P\big(s^*\in\mathrm{Reprod}(\xi_0)\mid \xi_0=x\big)\frac{P(\xi_0=x)}{1-\mu_0}$$
%% by that the selection does not generate new solutions
$$=\sum_{x\notin X^*} P\big(s^*\in\mathrm{Reprod}(\xi_0)\mid \xi_0=x\big)\frac{1}{\exp(n)^{\mathrm{poly}(n)}}\quad \text{%% by assumption } \mu_0=0$$
$$\leqslant \mathrm{poly}(n)\exp(n)^{\mathrm{poly}(n)}\frac{1}{\exp(n)}\frac{1}{\exp(n)^{\mathrm{poly}(n)}}$$
$$=\frac{\mathrm{poly}(n)}{\exp(n)}.$$

At time $t+1$, we sort all non-optimal population states into a sequence $\{x_i\}$ $(x_i\notin X^*)$, such that

$$P\big(\xi_{t+1}\in X^*\mid \xi_t=x_i\big)\leqslant P\big(\xi_{t+1}\in X^*\mid \xi_t=x_{i+1}\big).$$

Afterwards, we have

$$\sum_{x\in X} P\big(\xi_{t+1}\in X^*\mid \xi_t=x\big)=\sum_{x\in X} P\big(s^*\in\mathrm{Reprod}(\xi_t)\mid \xi_t=x\big)\leqslant \frac{\mathrm{poly}(n)\exp(n)^{\mathrm{poly}(n)}}{\exp(n)}$$
$$\Rightarrow\ \exists\hat{x}\in X:\quad P\big(\xi_{t+1}\in X^*\mid \xi_t=\hat{x}\big)\leqslant \frac{\mathrm{poly}(n)}{\exp(n)}.$$
%% otherwise the sum over X will exceed $\dfrac{\mathrm{poly}(n)\exp(n)^{\mathrm{poly}(n)}}{\exp(n)}$

Therefore,

$$P\big(\xi_t^R = x^* \mid \xi_t = x_0\big) \leqslant P\big(\xi_t^R = x^* | \xi_t = \hat{x}\big) \leqslant \frac{\text{poly}(n)}{\exp(n)},$$

by considering that x_0 has the lowest success probability.

Now, we choose a fitness function $f(\cdot)$ such that

$$f(x^*) > f(x_0) > \cdots > f(x_i) > f(x_{i+1}) > \ldots.$$

Since the solutions with larger fitness values will have higher probability to survive from the selection, we have

$$\begin{aligned} &\forall k\colon \quad \frac{P(\xi_{t+1} \in \{x_0, \ldots, x_k\})}{P(\xi_{t+1} \in X - \{x_0, \ldots, x_k\} - \{x^*\})} > \frac{P(\xi_t \in \{x_0, \ldots, x_k\})}{P(\xi_t \in X - \{x_0, \ldots, x_k\} - \{x^*\})} \\ &\quad \Rightarrow \ \forall k\colon \quad \frac{P(\xi_{t+1} \in \{x_0, \ldots, x_k\})}{1 - \mu_{t+1}} \geqslant \frac{P(\xi_t \in \{x_0, \ldots, x_k\})}{1 - \mu_t}. \end{aligned}$$

Then, we have

$$\begin{aligned} &\sum_{x \notin X^*} P\big(\xi_{t+1} \in X^* \mid \xi_t = x\big) \frac{P(\xi_t = x)}{1 - \mu_t} \\ &\quad \leqslant \sum_{x \notin X^*} P\big(\xi_1 \in X^* \mid \xi_0 = x\big) \frac{P(\xi_0 = x)}{1 - \mu_t} \leqslant \frac{\text{poly}(n)}{\exp(n)}, \end{aligned}$$

which makes $\mathbb{E}[\tau] \geqslant \frac{\exp(n)}{\text{poly}(n)}$ by Theorem 1. $\square$

6. Conclusion

This paper extends our preliminary research [27]. We establish a bridge between two of the most important theoretical issues of evolutionary algorithms (EAs), that is, the expected first hitting time (EFHT) and the convergence rate. With this bridge, we propose a new approach for analyzing the EFHT of EAs. The proposed approach bases on non-homogeneous Markov chains, and thus it is suitable for analyzing a broad range of EAs.

Using the proposed approach, we proved that a problem is hard (i.e., can only be solved in exponential time) for several EAs under various settings, including three static mutation operators, with/without population, a recombination operator and a time-variant mutation operator. It is noteworthy that the time-variant operator was hard to analyze before, while the proposed approach is naturally useful for this situation.

We gave an explanation to the question that what makes a problem hard to EA, that is, the *algorithm* part and the *problem* part mismatch. EAs are usually considered as general optimization approaches, or in other words, they are problem independent. Thus, when the parameters of an EA are fixed, the EA may run toward a wrong direction on some problems, which makes the problems hard for the EA. Based on this recognition, we proved that a *general problem* is hard if it has an exponentially large state space.

In the future, we intend to extend our approach to optimization problems for real-valued functions.

Acknowledgements

We want to thank the anonymous reviewers for their helpful comments and suggestions, and want to thank Tianshi Chen, Xiang-Nan Kong, Chao Qian and De-Chuan Zhan for proofreading the paper. We also want to thank the editor Raymond Perrault for his generous help in polishing the final version of the paper. This research was supported by the National Science Foundation of China (60635030, 60721002).

References

[1] T. Bäck, Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms, Oxford University Press, Oxford, UK, 1996.
[2] H.-G. Beyer, H.-P. Schwefel, I. Wegener, How to analyse evolutionary algorithms, Theoretical Computer Science 287 (1) (2002) 101–130.
[3] S.-J. Chang, H.-S. Hou, Y.-K. Su, Automated passive filter synthesis using a novel tree representation and genetic programming, IEEE Transactions on Evolutionary Computation 10 (1) (2006) 93–100.
[4] Y.-C. Chang, S.-M. Chen, A new query reweighting method for document retrieval based on genetic algorithms, IEEE Transactions on Evolutionary Computation 10 (5) (2006) 617–622.
[5] S. Droste, T. Jansen, I. Wegener, On the optimization of unimodal functions with the $(1+1)$ evolutionary algorithm, in: Proceedings of the 5th International Conference on Parallel Problem Solving from Nature (PPSN V), Amsterdam, Netherlands, 1998.
[6] S. Droste, T. Jansen, I. Wegener, A rigorous complexity analysis of the $(1+1)$ evolutionary algorithm for linear functions with boolean inputs, Evolutionary Computation 6 (2) (1998) 185–196.
[7] S. Droste, T. Jansen, I. Wegener, On the analysis of the $(1+1)$ evolutionary algorithm, Theoretical Computer Science 276 (1–2) (2002) 51–81.

[8] A.E. Eiben, R. Hinterding, Z. Michalewicz, Parameter control in evolutionary algorithms, IEEE Transactions on Evolutionary Computation 3 (2) (1999) 124–141.
[9] A.A. Freitas, A survey of evolutionary algorithms for data mining and knowledge discovery, in: A. Ghosh, S. Tsutsui (Eds.), Advances in Evolutionary Computing: Theory and Applications, Springer-Verlag, New York, NY, 2003, pp. 819–845.
[10] J. Garnier, L. Kallel, M. Schoenauer, Rigorous hitting times for binary mutations, Evolutionary Computation 7 (2) (1999) 173–203.
[11] B. Hajek, Hitting time and occupation time bounds implied by drift analysis with applications, Advances in Applied Probability 14 (1982) 502–525.
[12] J. He, L. Kang, On the convergence rates of genetic algorithms, Theoretical Computer Science 229 (1–2) (1999) 23–39.
[13] J. He, X. Yao, Drift analysis and average time complexity of evolutionary algorithms, Artificial Intelligence 127 (1) (2001) 57–85.
[14] J. He, X. Yao, Towards an analytic framework for analysing the computation time of evolutionary algorithms, Artificial Intelligence 145 (1–2) (2003) 59–97.
[15] J. He, X. Yao, A study of drift analysis for estimating computation time of evolutionary algorithms, Natural Computing 3 (1) (2004) 21–35.
[16] J. He, X. Yu, Conditions for the convergence of evolutionary algorithms, Journal of Systems Architecture 47 (7) (2001) 601–612.
[17] P.C.H. Ma, K.C.C. Chan, X. Yao, D.K.Y. Chiu, An evolutionary clustering algorithm for gene expression microarray data analysis, IEEE Transactions on Evolutionary Computation 10 (3) (2006) 296–314.
[18] A.E. Nix, M.D. Vose, Modeling genetic algorithms with Markov chains, Annals of Mathematics and Artificial Intelligence 5 (1) (1992) 77–88.
[19] P.S. Oliveto, J. He, X. Yao, Time complexity of evolutionary algorithms for combinatorial optimization: A decade of results, International Journal of Automation and Computing 4 (3) (2007) 281–293.
[20] J.S. Rosenthal, Minorization conditions and convergence rates for Markov chain Monte Carlo, Journal of the American Statistical Association 90 (430) (1995) 558–566.
[21] G. Rudolph, Convergence analysis of canonical genetic algorithms, IEEE Transactions on Neural Networks 5 (1) (1994) 96–101.
[22] G. Rudolph, How mutation and selection solve long path problems in polynomial expected time, Evolutionary Computation 4 (2) (1996) 195–205.
[23] G. Rudolph, Convergence Properties of Evolutionary Algorithms, Verlag Dr. Kovač, Hamburg, Germany, 1997.
[24] G. Rudolph, Finite Markov chain results in evolutionary computation: A tour d'horizon, Fundamenta Informaticae 35 (1–4) (1998) 67–89.
[25] J. Suzuki, A Markov chain analysis on simple genetic algorithms, IEEE Transactions on Systems, Man and Cybernetics 25 (4) (1995) 655–659.
[26] D. Wolpert, W.G. Macready, No free lunch theorems for optimization, IEEE Transactions on Evolutionary Computation 1 (1) (1997) 67–82.
[27] Y. Yu, Z.-H. Zhou, A new approach to estimating the expected first hitting time of evolutionary algorithms, in: Proceeding of the 21st National Conference on Artificial Intelligence, Boston, WA, 2006, pp. 555–560.