Towards an analytic framework for analysing the computation time of evolutionary algorithms ☆

Jun He, Xin Yao *

School of Computer Science, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Received 12 December 2001; received in revised form 15 August 2002

Abstract

In spite of many applications of evolutionary algorithms in optimisation, theoretical results on the computation time and time complexity of evolutionary algorithms on different optimisation problems are relatively few. It is still unclear when an evolutionary algorithm is expected to solve an optimisation problem efficiently or otherwise. This paper gives a general analytic framework for analysing first hitting times of evolutionary algorithms. The framework is built on the absorbing Markov chain model of evolutionary algorithms. The first step towards a systematic comparative study among different EAs and their first hitting times has been made in the paper.
© 2003 Published by Elsevier Science B.V.

Keywords: Evolutionary algorithms; Time complexity; Markov chain; First hitting time

1. Introduction

Evolutionary algorithms (EAs) are often used in optimisation [7,10,20,27,34]. Many experiments have been reported about successful applications of EAs in various combinatorial optimisation problems [18,35]. However, little is known about the computation time and time complexity of EAs for various optimisation problems except for a few cases [5,6,26,32].

Bäck [1] and Mühlenbein [21] studied the time complexity for the simple ONE-MAX problem. Rudolph [25] gave a comprehensive survey of the theoretical work up to 1997 and

* Corresponding author.
E-mail address: x.yao@cs.bham.ac.uk (X. Yao).
URL address: http://www.cs.bham.ac.uk/~xin (X. Yao).

provided an $O(n \log n)$ upper bound for the $(1 + 1)$ EA using the 1-bit-flip mutation for ONE-MAX problem. Garnier et al. [9] compared two different mutations in a $(1 + 1)$ EA, i.e., the c/n mutation and the 1-bit-flip mutation, when they are applied to the ONE-MAX problem, and showed that the bound on EA's average computation time in the former case is larger than that in the latter case by a factor of $(1 - e^{-c})^{-1}$. Droste et al. [4,5] improved these results by showing that these upper bounds can be reached: they proved that the time complexity of the $(1 + 1)$ EA with c/n mutation is indeed $\Theta(n \log n)$. Furthermore, they generalised this result to any linear binary functions. Some other unimodal functions, e.g., long path problems [8,24], have also been proved to be solvable in polynomial time. Most of the results mentioned above used $(1 + 1)$ EAs only. Recombination was rarely considered, although it may be helpful in some cases [16]. He and Yao [12–14] made one of the first attempts toward analysing EAs with recombination and with a population size greater than 1. This paper will examine both $(1 + 1)$ and $(N + N)$ EAs.

Markov chain models have been used widely in the analysis of EAs [22,29,31,33]. The first hitting time of a Markov chain plays an important role in analysing the computation time and time complexity of EAs. Analytic techniques from the passage time theory for Markov chains [15,30] are very useful in establishing a generic framework for studying EA's computation time. An initial idea first appeared in [2,28], but no further study has been carried out.

This paper builds a general framework for analysing the average hitting times of EAs based on their absorbing Markov chain models. Such a framework can facilitate answering questions at two different levels. Firstly, at the abstract level, it can help answering fundamental questions about EAs, such as, what kind of problems are easy (polynomial-time) or hard (exponential-time) for EAs. It can facilitate comparing first hitting times among different EAs so that insights can be gained on what makes a problem hard for an EA. Secondly, at the example level, the framework can facilitate the derivation of the average first hitting time or its bounds for a given EA and problem.

The following approach is adopted in the paper to establish our framework: We first study a simple $(1 + 1)$ EA which is easy to analyse, and discuss what kind of problems are difficult for it. Then by taking the simple EA as a starting point, we discuss other more complex EAs, which can all be regarded as an improvement to the simple EA. They should have some advantages over the simple EA.

In addition to the proposed general framework based on the absorbing Markov chain model, this paper has obtained several new results. First, the general approach has been applied to analysing different EAs by comparing their first hitting times, which were never done previously. Second, hard problems for EAs have been identified and grouped into two classes in theory, i.e., the "wide-gap" problem and the "long-path" problem. Third, explicit expressions of first hitting times for several $(1 + 1)$ EAs have been obtained. Previously, only bounds were known [5,26] except for two simple cases given by [8]. Finally, we have studied some special cases under our framework: one is to generalise the $O(n \log n)$ result for the linear function to a more general case and provide a new and shorter proof; the other is to show that an EA with crossover can solve a problem in polynomial time while the EA without crossover cannot.

The rest of this paper is organised as follows. Section 2 reviews some existing results on the first hitting time of Markov chains and also derives some new results on estimating

their bounds. Section 3 introduces the absorbing Markov chain model of EAs used in this paper and some definitions. Section 4 discusses $(1 + 1)$ EAs using different selection strategies and analyse what kinds of problems are difficult for them. Comparisons among these EAs are given. Section 5 analyses population-based EAs without and with crossover, and compares them with $(1 + 1)$ EAs. Finally, Section 6 concludes the paper with a short discussion of the results obtained in this paper.

2. Analytic approach to first hitting time of Markov chains

This section introduces some fundamental results on the first hitting times of Markov chains by the analytic approach. Details on this topic can be found in [15] for Markov chains with a discrete time parameter and in [30] for Markov processes with a continuous time parameter.

Our objective here is to obtain explicit expressions and bounds of the first hitting time of a Markov chain from its transition matrix. Such expressions and bounds will be used to analyse the first hitting time of EAs in later sections of this paper.

Let $(X_t;\ t = 0, 1, \ldots)$ be a homogeneous absorbing Markov chain defined on a probability space with a discrete state space S and a discrete time parameter with a canonical transition matrix [15],

$$\mathbf{P} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R} & \mathbf{T} \end{pmatrix}, \tag{1}$$

where T denotes the set of transient states and $H = S - T$ the set of absorbing states.

Definition 1. Given a Markov chain (X_t), define its *first hitting time* to the absorbing set H when starting from transient state i to be:

$$\tau_i = \min\{t; t \geqslant 0, X_t \in H \mid X_0 = i\},$$

with the infimum of the empty set taken to be ∞.

We are interested in the expectation $\mathbb{E}[\tau_i]$ of the first hitting time. Since it always holds: $\tau_i = 0$ for any $i \in H$, we only need to concentrate on the states outside set H.

Theorem 1. *Let τ_i be the first hitting time of Markov chain (X_t) when starting from transient state i, and denote $m_i = \mathbb{E}[\tau_i]$, $\mathbf{m} = [m_i]_{i \in T}$, then*

$$\mathbf{m} = (\mathbf{I} - \mathbf{T})^{-1}\mathbf{1}, \tag{2}$$

where $\mathbf{1}$ denotes the vector $(1, \ldots, 1)^{\mathrm{t}}$.

Proof. The theorem is part of Theorem 3.2 in [15]. $\quad\square$

For most transition matrices $\mathbf{P}$, it is difficult or even impossible to determine the inverse of matrix $\mathbf{I} - \mathbf{T}$. We can obtain an explicit expression of $\mathbf{m}$ only for some simple transition matrices, e.g., when $\mathbf{T}$ is a tridiagonal matrix or a lower triangular matrix. It is important to

estimate the bounds of **m** in practice. Here we give some new results based on Theorem 1 and drift analysis [12,19].

First we introduce some notations [17]. For vectors $\mathbf{a} = [a_i]$ and $\mathbf{b} = [b_i]$, if for all i, $a_i \geqslant b_i$, we shall write $\mathbf{a} \geqslant \mathbf{b}$. For matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$, if for all $i, j, a_{ij} \geqslant b_{ij}$, we shall write $\mathbf{A} \geqslant \mathbf{B}$.

For a Markov chain (X_t) and for each state $i \in S$, a function $d(i)$ or $d_i: S \to R$ is called *a metric function* if

$$d_i = \begin{cases} = 0, & \text{if } i \in H, \\ \geqslant 0, & \text{if } i \notin H. \end{cases}$$

For any state $i \in S$, we call

$$\sum_{j \in S} (d_i - d_j) p(i, j),$$

the *mean one-step drift of state i towards set H* [12,19].

The theorems below give an analytic way to estimate the bounds of the mean hitting time of an absorbing Markov chain from its transition matrix.

Theorem 2. *Let* $\mathbf{d} = [d_i]$ *with* $d_i \geqslant 0$. *If* $\mathbf{d} - \mathbf{Td} \leqslant \mathbf{1}$, *then* $\mathbf{m} \geqslant \mathbf{d}$.

Proof. From $\mathbf{d} - \mathbf{Td} \leqslant \mathbf{1}$, we get

$$\mathbf{1} - \mathbf{d} + \mathbf{Td} \geqslant 0.$$

Since **P** is a transition matrix, $\mathbf{T} \geqslant 0$ and the eigenvalues of **T** are less than 1, so $(\mathbf{I} - \mathbf{T})^{-1} \geqslant 0$ and we get

$$(\mathbf{I} - \mathbf{T})^{-1}(\mathbf{1} - \mathbf{d} + \mathbf{Td}) \geqslant 0.$$

Then we have $(\mathbf{I} - \mathbf{T})^{-1}\mathbf{1} \geqslant \mathbf{d}$ and then $\mathbf{m} \geqslant \mathbf{d}$. $\square$

Theorem 3. *Let* $\mathbf{d} = [d_i]$ *with* $d_i \geqslant 0$. *If* $\mathbf{d} - \mathbf{Td} \geqslant \mathbf{1}$, *then* $\mathbf{m} \leqslant \mathbf{d}$.

Furthermore, we have the following corollaries.

Corollary 1. *Let* $\mathbf{d} = [d_i]$ *with* $d_i \geqslant 0$. *If* $\mathbf{1} - \mathbf{d} + \mathbf{Td} \geqslant \mathbf{a} > 0$, *then* $\mathbf{m} \geqslant \mathbf{d} + (\mathbf{I} - \mathbf{T})^{-1}\mathbf{a}$.

Corollary 2. *Let* $\mathbf{d} = [d_i]$ *with* $d_i \geqslant 0$. *If* $\mathbf{d} - \mathbf{Td} - \mathbf{1} \geqslant \mathbf{a} > 0$, *then* $\mathbf{m} \leqslant \mathbf{d} - (\mathbf{I} - \mathbf{T})^{-1}\mathbf{a}$.

Since the cardinality of state space S is often very large, it is not easy to choose a proper d_i for each $i \in S$. We often decompose S into subspaces and use a common d for the states in the same subspace.

Corollary 3. *Assume that state space S is decomposed into* $L + 1$ *subspaces*:

$$S = \bigcup_{l=0}^{L} S_l, \qquad S_l \cap S_k = \emptyset, \quad \text{if } l \neq k,$$

where $S_0 = H$. There exists a sequence of $\{d_0, \ldots, d_L\}$ where $d_0 = 0$ and $d_l > 0$ for $l = 1, \ldots, L$. For any $i \in S_l$ ($l = 0, \ldots, L$), let $d_i = d_l$, if

$$\sum_{k=0}^{L} \sum_{j \in S_k} (d_l - d_k) p(i, j) \geqslant 1,$$

then for any $i \in S_l$ ($l = 0, \ldots L$), $m_i \leqslant d_i = d_l$.

Corollary 4. *Assume state space S is decomposed into $L + 1$ subspaces:*

$$S = \bigcup_{l=0}^{L} S_l, \qquad S_l \cap S_k = \emptyset, \quad if \, l \neq k,$$

where $S_0 = H$. There exists a sequence $\{d_0, \ldots, d_l\}$ where $d_0 = 0$ and $d_l > 0$ for $l > 0$. For any $i \in S_l$ ($l = 0, \ldots, L$), let $d_i = d_l$, if

$$\sum_{k=0}^{L} \sum_{j \in S_k} (d_l - d_k) p(i, j) \leqslant 1,$$

then for any $i \in S_l$ ($l = 0, \ldots, L$), it holds: $m_i \geqslant d_i = d_l$.

According to Theorems 2 and 3 and the above corollaries, we do not need to solve the linear systems when we estimate the bounds of the first hitting time.

3. Evolutionary algorithm, Markov chain model and first hitting time

A combinatorial optimisation problem can be described as follows [23]: Given a problem instance, i.e., a pair (S, f), where S is the set of feasible solutions (a discrete state space) and f is an objective function $f : S \to R$, the problem is to find an $x \in S$ such that

$$f(x) \geqslant f(y), \quad \forall y \in S,$$

where x is called a global optimal (maximum) solution to the given instance. Although not a requirement, the objective function will (or can be made to) take on only non-negative values. Minimisation problems can be defined similarly.

3.1. Absorbing Markov chain: A model of evolutionary algorithms

In EAs, a solution x in S is represented by an individual. A population is a collection of individuals. We always use $(x_1, \ldots, x_N)$ to indicate a population of N individuals. The population space E consists of all possible populations. In a $(1 + 1)$ EA, the population space E is the same as the state space S.

An EA for solving a combinatorial optimisation problem can be described as follows. Given an initial population X_0, let $X_t = (x_1, \ldots, x_N)$ in E be the population at time t (i.e., generation t). Offspring can then be reproduced as follows:

Recombination Individuals in population X_t are recombined, denoted by

$$rec : E \rightarrow E,$$

yielding an intermediate population $X_t^{(c)}$.

Mutation Individuals in population $X_t^{(c)}$ are mutated, denoted by

$$mut : E \rightarrow E,$$

yielding another intermediate population $X_t^{(m)}$.

Survival Selection Each individual in the original population X_t and mutated population $X_t^{(m)}$ is assigned a survival probability. N individuals will then be selected to survive into the next generation X_{t+1}. This operation can be denoted by

$$sel : E \times E \rightarrow E.$$

For most EAs, the state of population X_{t+1} depends only on population X_t. In this case, the stochastic process $(X_t; \ t = 0, 1, \ldots)$ is a Markov chain [3,11,26] whose state space is the population space. If no self-adaptation is used in EAs, the chain will be homogeneous. In this paper, we do not consider self-adaptation in EAs. In other words, all EAs discussed in this paper can be modeled by homogeneous Markov chains.

Because we are only interested in the first hitting time of a Markov chain and do not care about its behaviour after that, we can model the EA by an absorbing Markov chain. Let H be the set of populations that contain at least one global optimal solution to the optimisation problem. Define an absorbing Markov chain $(\overline{X}_t; \ t = 0, 1, 2, \ldots)$ whose transition probabilities is given by:

$$\bar{p}(i, j; t) = \begin{cases} 1, & i \in H, \ j = i, \\ 0, & i \in H, \ j \neq i, \\ \mathbb{P}(X_{t+1} = j \mid X_t = i), & i \notin H. \end{cases} \tag{3}$$

The transition probabilities of Markov chain $(\overline{X}_t)$ are different from those of chain (X_t) only in set H. By using the new absorbing Markov chain, it is more convenient for us to study the first hitting time [30]. In the reminder of the paper, we only consider the absorbing Markov chain. Without any confusion, we will still use (X_t) to represent the absorbing chain.

If the chain is homogeneous, i.e., $p(i, j; t)$ is independent of t, we write $p(i, j)$ in short.

3.2. First hitting time of evolutionary algorithms

Let $(X_t; \ t = 0, 1, \ldots)$ be the absorbing Markov chain associated with an EA, we can define its first hitting time to find an optimal solution as follows.

Definition 2. Let $H \subseteq E$ be the set of populations that contain at least one optimal solution for a given instance of a combinatorial optimisation problem. Then

$$\tau_i = \min\{t \geqslant 0; X_t \in H \mid X_0 = i\}$$

is defined as the *first hitting time* to H starting from state i, and

$$m_i = \mathbb{E}[\tau_i; \tau_i < \infty].$$

m_i indicates the average number of generations needed for an EA to find an optimal solution starting from initial population i.

Denote $\mathbf{m} = [m_i]_{i \in E}$. For vector $\mathbf{m}$, we can define a norm for it. Assume the initial population distribution is $\mu_0(i) = \mathbb{P}(X_0 = i)$, the norm will be

$$\|\mathbf{m}\| = \sum_{i \in E} m_i \mu_0(i).$$

Two special norms are

$$\|\mathbf{m}\|_1 = \frac{1}{|E|} \sum_{i \in E} m_i, \quad \text{and}$$

$$\|\mathbf{m}\|_\infty = \max_{i \in E} \{m_i\}.$$

Norms $\|\mathbf{m}\|_1$ and $\|\mathbf{m}\|_\infty$ represent two different performance measures used in the analysis of EAs: $\|\mathbf{m}\|_1$ is related to the average-case analysis, and $\|\mathbf{m}\|_\infty$ to the worst-case analysis. Starting from these vector norms, we can define the natural matrix norm induced to the vector norm $\|\mathbf{m}\|$. The time complexity of an EA can be measured by $\|\mathbf{m}\|$.

Based on $\mathbf{m}$, we can obtain the total number of operations used by an EA to find an optimal solution. Let n be the encoding length of an individual. For the binary representation, n is the string length. Assume that during one generation of an EA, $NO_r(n)$, $NO_m(n)$, $NO_s(n)$, and $NO_e(n)$ are the number of operations in recombination, mutation, selection, and fitness evaluation, respectively, then the mean number of operations for the EA to find an optimal solution will be

$$\big(NO_r(n) + NO_m(n) + NO_s(n) + NO_e(n)\big) \|\mathbf{m}\|,$$

and starting from state i, the mean number of operations will be

$$\big(NO_r(n) + NO_m(n) + NO_s(n) + NO_e(n)\big) m_i.$$

Usually $NO_r(n)$, $NO_m(n)$, $NO_s(n)$, and $NO_e(n)$ are not difficult to estimate and are polynomial in n (if one of them was exponential in n, the EA would be exponential in n; this EA should never be used in practice). So $\mathbf{m}$ is the decisive factor in analysing the time complexity of an EA. In this paper, we focus on estimating the first hitting time $\mathbf{m}$ rather than the total number of operations used by an EA.

Definition 3. Given a combinatorial optimisation problem, an EA is said to be of *polynomial time* (or *exponential time*) in n under norm $\|\mathbf{m}\|$ if $\|\mathbf{m}\|$ is polynomial (or exponential) in n.

Based on the mean first hitting time we can divide the states in E into two different classes:

$E(P)$: State i in space E belongs to class $E(P)$ if m_i is polynomial in n; and
$E(NP)$: State i belongs to class $E(NP)$ if m_i is exponential in n.

Assume the initial input is selected at random. Based on this classification, we come to two useful measures for the performance of EAs, that is,

$$\frac{|E(P)|}{|E|} \tag{4}$$

which is the probability of an EA to find an optimum in polynomial time, and

$$\sum_{i \in E(P)} m_i, \tag{5}$$

which is the mean first hitting time of an EA to find an optimum in polynomial time. Wegener et al. [5,16,32] analysed some Pseudo-Boolean optimisation problems using these two measures and showed that some multi-start EAs had a higher probability $|E(P)|/|E|$ and a much shorter time $\sum_{i \in E(P)} m_i$ than $(1 + 1)$ EAs for certain problems.

A fundamental question in the analysis of EAs is: For a given combinatorial optimisation problem, what is the mean first hitting time of an EA when it is used to solve the problem?

The following two sections provide an unified general framework to analyse the mean first hitting time of EAs. The framework enables us to derive both old and new results about EA's mean first hitting time in a systematic way. It also enables us to see the big picture and gain insights into future directions for analysing EA's mean first hitting time for other optimisation problems than those discussed here.

Since the number of entries in the transition matrix is often huge (exponential in individual and population sizes), an effective technique is to decompose the state space E into $L + 1$ subspaces, $E_0, E_1, \ldots, E_L$, where $E_0 = H$.

In this paper we always assume $\mathbb{E}[\tau_i] < \infty$ for any state $i \in E$, that is, the EA can find an optimal point in finite time (number of generations). This is an important assumption in our analysis. In other words, we do not consider EAs which cannot find the optimal solution in finite time. Such EAs have no practical use in optimisation.

4. Analysis of $(1 + 1)$ evolutionary algorithms

In this section, we restrict our discussion to the binary space $S = \{0, 1\}^n$. Denote $D(i, j)$ as the Hamming distance between two states i and j.

4.1. Elitist selection: Accepting only higher fitness individuals

Let's consider a $(1 + 1)$ EA which uses following operators:

Mutation I. Any proper mutation operator applied to population $X_t = \{x\}$. After mutation we get an intermediate population $X^{(m)}$.

Selection I. If $f(X^{(m)}) > f(X_t)$, then let $X_{t+1} = X^{(m)}$; otherwise, let $X_{t+1} = X_t$.

A proper mutation operator often used in EAs is

Mutation II. Given a population $X_t = \{x\}$, flip each of its bits with probability $p_m > 0$.

We start our analysis from this simple EA with the simple selection strategy.

Since S is a finite state space, fitness $f(x)$ takes only a finite number of values. Without the loss of generality, assume that there are $L + 1$ values. Let's sort them according to the descending order,

$$f_0 = f_{\max} > f_1 > \cdots > f_L = f_{\min}.$$

Define S_l to be $S_l = \{x \in S;\ f(x) = f_l\}$, $l = 0, \ldots, L$.

Then the state space S can be decomposed into $L + 1$ subspaces:

$$S = \bigcup_{l=0}^{L} S_l.$$

For the absorbing Markov chain (X_t) associated with the EA, the transition probability $p(i, j)$ between $i \in S_l$ and $j \in S_k$ ($l, k = 0, \ldots, L$), satisfies

$$p(i, j) = \begin{cases} 1, & i \in S_0,\ j = i, \\ 0, & i \in S_0,\ j \neq i, \\ \geqslant 0, & i \in S_l,\ j \in S_k,\ k < l, \\ \geqslant 0, & i \in S_l,\ j = i \\ 0, & \text{else.} \end{cases}$$

In matrix form, we have

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

where $\mathbf{T}$ is a lower-triangular matrix.

Theorem 4. *For an EA using Mutation* I *and Selection* I*, its first hitting time is given by*

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

and in more details

$$
\begin{aligned}
m_i &= 0, & i \in S_0, \\
m_i &= \frac{1}{\sum_{j \in S_0} p(i, j)}, & i \in S_1, \\
m_i &= \frac{1 + \sum_{k=1}^{l-1} \sum_{j \in S_k} p(i, j) m_j}{\sum_{k=0}^{l-1} \sum_{j \in S_k} p(i, j)}, & i \in S_l,\ l = 2, \ldots, L.
\end{aligned}
\tag{6}
$$

Proof. It comes directly from Theorem 1 and the fact that $\mathbf{T}$ is a lower-triangular matrix. $\square$

The theorem gives the explicit expression of the first hitting time of $(1 + 1)$ EAs using Selection I.

As an example, let us consider the EA using Mutation II and Selection I for the ONE-MAX problem.

Example 1. For $x = (s_1 \ldots s_n)$,

$$f(x) = \sum_{i=1}^{n} s_i. \tag{7}$$

Decompose the state space into $n + 1$ subspaces:

$$S_l = \left\{ x; n - \sum_{i=1}^{n} s_i = l \right\}, \quad l = 0, \ldots, n.$$

Let any $i \in S_l$ and any $j \in S_k$ ($l, k = 0, 1, \ldots, n$), and define

$$a_{l,k}(i) = \sum_{j \in S_k} p(i, j),$$

then the transition probability $a_{l,k}(i)$ is known [9] as,

$$a_{l,k}(i) = \begin{cases} \sum_{v=0}^{(n-l)\wedge k} \binom{l}{l-k+v}\binom{n-l}{v}(p_m)^{l-k+2v}(1-p_m)^{n-l+k-2v}, & k = 0, \ldots, l-1, \\ 1 - \sum_{k=0}^{l-1} a_{l,k}(i), & k = l, \\ 0, & \text{otherwise,} \end{cases}$$

where $u \wedge v = \min\{u, v\}$.

Since $a_{l,k}(i)$ is only dependent on l and k, we denoted it in short as $a_{l,k}$. Now we can give the explicit expression of the mean first hitting times using Theorem 4,

$$\begin{aligned} m_0 &= 0, \\ m_l &= \frac{1 + a_{l,1}m_1 + a_{l,2}m_2 + \cdots + a_{l,l-1}m_{l-1}}{a_{l,0} + a_{l,1} + a_{l,2} + \cdots + a_{l,l-1}}, \quad l = 1, \ldots, n. \end{aligned}$$

and $m_i = m_l$ for any $i \in S_l$ ($l = 0, \ldots, n$).

It now becomes straightforward, albeit still complex, to compare first hitting times using different mutation rates and determine which mutation rate p_m is optimal.

From a more practical point of view, it is interesting to study the conditions under which the mean first hitting time of an EA is polynomial in its input size. Two immediate corollaries are of interest to us.

Corollary 5. *If $\|(\mathbf{I} - \mathbf{T})^{-1}\|_\infty$ is polynomial in n, then $\|\mathbf{m}\|_\infty$ is also polynomial in n.*

Corollary 6. *If $\|(\mathbf{I} - \mathbf{T})^{-1}\|_1$ is polynomial in n, then $\|\mathbf{m}\|_1$ is also polynomial in n.*

Theorem 5. *For the absorbing Markov chain associated with an EA using Mutation I and Selection I, if L is polynomial in n, then $\|\mathbf{m}\|_\infty$ is polynomial in n if and only if for any l ($l = 1, \ldots, L$) and for any $i \in S_l$,*

$$\frac{1}{\sum_{k=0}^{l-1} \sum_{j \in S_k} p(i, j)}$$

is polynomial in n.

Proof. First we prove the necessary condition.

Since for $l = 1, \ldots, L$ and $i \in S_l$

$$m_i = \frac{1 + \sum_{k=1}^{l-1} \sum_{j \in S_k} p(i, j) m_j}{\sum_{k=0}^{l-1} \sum_{j \in S_k} p(i, j)}$$

$$\geqslant \frac{1}{\sum_{k=0}^{l-1} \sum_{j \in S_k} p(i, j)},$$

and m_i is polynomial in n, we arrive at the necessary condition.

Secondly we prove the sufficient condition.

For $l = 1$ and any $i \in S_1$, since $1 / \sum_{j \in S_0} p(i, j)$ is polynomial in n, we know that for any $i \in S_1$, m_i is polynomial in n.

Now assume for an $l \geqslant 1$ and for any $i \in S_l$, m_i is polynomial in n. Since L is polynomial in n, from (6) we get for $l + 1$ and for any $i \in S_{l+1}$, m_i is polynomial in n too. By induction, we have shown for $l = 0, \ldots, L$ and $i \in S_l$, m_i is polynomial in n. $\quad \square$

Now we investigate the conditions under which $\|\mathbf{m}\|_\infty$ is exponential in n. From Theorem 4, it is easy to see that there are only two cases which can lead to an exponential $\|\mathbf{m}\|_\infty$.

Theorem 6. *For the absorbing Markov chain associated with an EA using Mutation* I *and Selection* I, $\|\mathbf{m}\|_\infty$ *is exponential in* n *if and only if*

(1) *for some l $(1 \leqslant l \leqslant L)$ and some state i: $i \in S_l$,*

$$\frac{1}{\sum_{k=0}^{l-1} \sum_{j \in S_k} p(i, j)}$$

is exponential in n, or

(2) *for all l $(l = 1, \ldots, L)$ and all states $i \in S_l$,*

$$\frac{1}{\sum_{k=0}^{l-1} \sum_{j \in S_k} p(i, j)}$$

is polynomial in n. L is exponential in n. There is a metric function $\mathbf{d} = [d_i]$, where $d_i = 0$ for all $i \in S_0$, $d_i \geqslant 1$ for all $i \in S_l$ $(l = 1, \ldots, L)$, such that

$$\sum_{k=0}^{l-1} \sum_{j \in S_k} (d_i - d_j) p(i, j) \leqslant 1,$$

and for some state i: $i \in S$, d_i is exponential in n.

Proof. First we prove the sufficient conditions.

(1) Assume that condition (1) holds. Then for some l $(1 \leqslant l \leqslant L)$ and some i: $i \in S_l$,

$$\frac{1}{\sum_{k=0}^{l-1} \sum_{j \in S_k} p(i, j)}$$

is exponential in n. According to Theorem 5, m_i is exponential in n.

(2) Assume that condition (2) holds. We use induction to prove that for all $l = 0, \ldots, L$, $m_i \geqslant d_i$.

When $l = 0$, for any $i \in S_0$, we know $m_i = d_i = 0$. Hence, for any $i \in S_l$: $m_i \geqslant d_i$.

Now assume that given some l ($l \geqslant 0$), and for all $k = 0, \ldots, l$, we have, for any $i \in S_k$: $m_i \geqslant d_i$. In the following, we will prove that for $l + 1$, we also have, for any $i \in S_{l+1}$: $m_i \geqslant d_i$.

For any $i \in S_{l+1}$, we know from Theorem 4 that

$$
\begin{aligned}
m_i - d_i &= \frac{1 + \sum_{k=1}^{l} \sum_{j \in S_k} p(i, j) m_j}{\sum_{k=0}^{l} \sum_{j \in S_k} p(i, j)} - d_i \\
&\geqslant \frac{1 - \sum_{k=0}^{l} \sum_{j \in S_k} (d_i - d_j) p(i, j)}{\sum_{k=0}^{l} \sum_{j \in S_k} p(i, j)} \\
&\geqslant 0.
\end{aligned}
$$

So we have proven that for $l + 1$, $m_i \geqslant d_i$ for any $i \in S_{l+1}$.

By induction, we get for $l = 0, 1, \ldots, L$ and for any $i \in S_l$, we have $m_i \geqslant d_i$.

Since for some state i, d_i is exponential in n, so $\|\mathbf{m}\|_\infty$ is also exponential in n.

We now prove the necessary conditions of the theorem. Assume $\|\mathbf{m}\|_\infty$ is exponential in n. There exist only two cases.

(1) For some l: $1 \leqslant l \leqslant L$ and some state i: $i \in S_l$,

$$
\frac{1}{\sum_{k=0}^{l-1} \sum_{j \in S_k} p(i, j)}
$$

is exponential in n. This is condition (1).

(2) For all $l = 0, 1, \ldots, L$ and all states $i \in S_l$,

$$
\frac{1}{\sum_{k=0}^{l-1} \sum_{j \in S_k} p(i, j)}
$$

is polynomial in n. This is the first part of condition (2). Now we prove the remaining part of condition (2).

According to Theorem 5, L must be exponential in n.

Let $d_i = m_i$ for any $i \in S$, it is easy to see that $d_i = 0$ for any state $i \in S_0$ and $d_i \geqslant 1$ for any state $i \in S_l$ ($l = 1, \ldots, L$). As a result of Theorem 4,

$$
\sum_{k=0}^{l-1} \sum_{j \in S_k} (d_i - d_j) p(i, j) = 1.
$$

Since $\|\mathbf{m}\|_\infty$ is exponential in n, we know that for some $i \in S$, $d_i = m_i$ is exponential in n. $\quad \square$

The above theorem implies an classification of hard functions with respect to the worst-case analysis of EAs. Intuitively, we can understand the above two conditions from a "geometric" point of view.

Condition (1) means that starting from some state in a subspace S_l, the probability for the EA to move to upper subspaces with higher fitness is very small. The EA is "trapped" in the subspace S_l, because there is a "wide gap" between the current subspace and upper subspaces with higher fitness.

Condition (2) means that starting from each subspace S_l, the probability for the EA to move to upper subspaces with higher fitness is reasonably large, but the length L (i.e., the number of subspaces) is exponential in n. The EA has to take a "long path" to reach the optimum.

A similar classification according to the average-case analysis is much more difficult to achieve. Some initial results are given here, which we hope will lead to a complete classification in the future.

Corollary 7. *For the absorbing Markov chain associated with an EA using Mutation* I *and Selection* I*, denote $S_1(NP)$ as the set consisting of all following states i: $i \in S_l$ ($l = 1, \ldots, L$), such that*

$$\frac{1}{\sum_{k=0}^{l-1} \sum_{j \in S_k} p(i, j)}$$

is exponential in n. If $S_1(NP)$ is reasonably large, that is, $|S_1(NP)|/|S|$ is an inverse polynomial of n, then $\|\mathbf{m}\|_1$ is exponential in n.

Proof. From Theorem 6, we know that for each state i in $S_1(NP)$, m_i is exponential in n. Since

$$\|\mathbf{m}\|_1 \geqslant \frac{1}{|S|} \sum_{i \in S_1(NP)} m_i,$$

and $|S_1(NP)|/|S|$ is an inverse polynomial of n, we know that

$$\frac{1}{|S|} \sum_{i \in S_1(NP)} m_i$$

is exponential in n, and thus $\|\mathbf{m}\|_1$ is exponential in n. $\square$

Corollary 8. *For the absorbing Markov chain associated with an EA using Mutation* I *and Selection* I*, assume that for all l ($l = 1, \ldots, L$) and any $i \in S_l$,*

$$\frac{1}{\sum_{k=0}^{l-1} \sum_{j \in S_k} p(i, j)}$$

is polynomial in n, where L is exponential in n. There is a metric function $\mathbf{d} = [d_i]$, where $d_i = 0$ for all $i \in S_0$ and $d_i \geqslant 1$ for all $i \in S_l$ ($l = 1, \ldots, L$), such that

$$\sum_{k=0}^{l-1} \sum_{j \in S_k} (d_i - d_j) p(i, j) \leqslant 1,$$

and for some state i: $i \in S$, d_i is exponential in n. Denote $S_2(NP)$ as the set of all these states i. If $S_2(NP)$ is reasonably large, that is, $|S_2(NP)|/|S|$ is an inverse polynomial of n, then $\|\mathbf{m}\|_1$ is exponential in n.

Theorem 6 gives us a powerful and useful tool to characterise problems that are hard for the $(1 + 1)$ EA using Mutation II with mutation rate $p_m = 1/n$ and Selection I under the maximum norm. In the following, some examples are given to illustrate how the theorem can be applied in practice.

First we consider a class of hard functions related to Condition 1 in Theorem 6. This class of hard functions, $f(x)$, can be described as follows.

Hard Function I. State space S is decomposed into $L + 1$ subspaces $S_0, S_1, \ldots, S_L$ based on fitness, and $f_0 > f_1 > \cdots > f_L$. The distance between S_0 and S_1 satisfies

$$\min\{D(x, y); i \in S_1, j \in S_0\} \geqslant k,$$

where k is a linear function of n, and thus n^k is exponential in n.

It is easy to verify that the above functions satisfy Condition 1 of Theorem 6. For any $i \in S_1$, the transition probability $\sum_{j \in S_0} p(i, j) \leqslant O(n^{-k})$, and then

$$\frac{1}{\sum_{j \in S_0} p(i, j)} = \Omega(n^k)$$

will be exponential. This implies that condition (1) of Theorem 6 holds.

Such functions are what we called "wide gap" problems because the "gap" between S_0 and S_1 is wide. If $|S_1|/|S|$ is reasonably large, i.e., an inverse polynomial of n, then Hard Function I is also hard in terms of the average norm. An example of Hard Function I is the deceptive function.

Second, according to condition (2) of Theorem 6, we consider another class of hard functions, which is hard for the $(1 + 1)$ EA using Mutation II and Selection I in the maximum norm.

Hard Function II. Space S is decomposed into $L + 1$ subspaces $S_0, S_1, \ldots, S_L$ based on fitness, where L is exponential in n and $f_0 > f_1 > \cdots > f_L$. There exist an integer $L' < L$, where L' is exponential in n, and a polynomial $g(n) > 0$. For all $l (l = 0, 1, \ldots, L')$ and for any $i \in S_l$,

$$\sum_{k=0}^{l-1} (l - k) \sum_{j \in S_k} \left(\frac{1}{n - 1} \right)^{D(i, j)} \leqslant \frac{1}{g(n)}.$$

We can verify that Hard Function II satisfies condition (2) of Theorem 6. For $l = 0, \ldots, L'$, let

$$d_l = \lg(n).$$

Here we do not need to give any definition of the metric function to states i in S_l $(l = L' + 1, \ldots, L)$. We can ignore them. It is easy to see that for all $l = 0, 1, \ldots, L'$,

$$\sum_{k=0}^{l-1} \sum_{j \in S_k} (d_i - d_j) p(i, j) \leqslant 1.$$

Hence, condition (2) of Theorem 6 holds.

An example of Hard Function II is the long k-path problem given in Definition 25 in [5] with $k = \sqrt{n-1}$.

Finally, we consider two concrete examples, i.e., the bit-sum function and the linear function, to illustrate the application of our general results.

Example 2. Assume $\{c_l; l = 0, 1, \ldots, n\}$ are $n+1$ different positives. For $x = (s_1 \ldots s_n)$, define

$$f(x) = c_l, \quad \text{if } \sum_{k=1}^{n} s_k = l, \ l = 0, 1, \ldots, n. \tag{8}$$

Consider the $(1+1)$ EA with Mutation II and Selection I for solving the problem.

First we give an explicit expression of the first hitting time of the EA. The analysis is similar to that of Example 1.

We sort $\{c_k; k = 0, \ldots, n\}$ based on their values from large to small. Without loss of generality, assume $c_{\beta_0} > c_{\beta_1} > \cdots > c_{\beta_n}$, where $\{\beta_0, \ldots, \beta_n\}$ is a permutation of $\{0, \ldots, n\}$.

Define subspace $S_{\beta_k} = \{x; f(x) = c_{\beta_k}\}, k = 0, 1, \ldots, n$. Let $i \in S_{\beta_l}$ and $j \in S_{\beta_k}$, where $l, k = 0, 1, \ldots, n$, and define

$$a_{\beta_l, \beta_k}(i) = \sum_{j \in S_{\beta_k}} p(i, j),$$

then the transition probability $a_{\beta_l, \beta_k}(i)$ is given by

$$a_{\beta_l, \beta_k}(i) = \begin{cases} 1, & \text{if } l = k = 0, \\ \sum_{j \in S_{\beta_k}} (p_m)^{D(i,j)}(1 - p_m)^{n-D(i,j)}, & \text{if } l \geqslant 1 \text{ and } k < l, \\ 1 - \sum_{v=0}^{l-1} a_{\beta_l, \beta_v}(i), & \text{if } l \geqslant 1 \text{ and } k = l, \\ 0, & \text{otherwise.} \end{cases}$$

Because $a_{\beta_l, \beta_k}(i)$ is only dependent on β_l and β_k, we denote them in short by a_{β_l, β_k}. Now we can give the explicit expression of the mean first hitting time according to Theorem 4,

$$m_{\beta_0} = 0,$$
$$m_{\beta_l} = \frac{1 + a_{\beta_l, \beta_1} m_{\beta_1} + a_{\beta_l, \beta_2} m_{\beta_2} + \cdots + a_{\beta_l, \beta_{l-1}} m_{\beta_{l-1}}}{a_{\beta_l, \beta_0} + a_{\beta_l, \beta_1} + a_{\beta_l, \beta_2} + \cdots + a_{\beta_l, \beta_{l-1}}}, \quad l = 1, \ldots, n.$$

and $m_i = m_{\beta_l}$ for any $i \in S_{\beta_l}$ $(l = 0, \ldots, n)$.

Since the function takes only $n+1$ values, we can apply Theorem 5 to it. According to Theorem 5, $\|\mathbf{m}\|_{\infty}$ is polynomial in n if and only if for all l $(l = 1, \ldots, n)$ and $i \in S_{\beta_l}$,

$$\left(\sum_{k=0}^{l-1} \sum_{j \in S_{\beta k}} (p_m)^{D(i,j)}(1 - p_m)^{n-D(i,j)} \right)^{-1}$$

is polynomial in n.

Let us examine the case of mutation rate $p_m = 1/n$. In this case, $\|\mathbf{m}\|_\infty$ is polynomial in n if and only if for all l ($l = 1, \ldots, n$) and $i \in S_{\beta_l}$,

$$\left(\sum_{k=0}^{l-1} \sum_{j \in S_{\beta_k}} (n-1)^{-D(i,j)} \right)^{-1}$$

is polynomial in n.

Let

$$\Delta(\beta) = \min\{\beta_l - \beta_k; l = 1, \ldots, L, k = 0, 1, \ldots, l-1\}.$$

If $\Delta(\beta)$ is constant, then $\|\mathbf{m}\|_\infty$ is polynomial in n. If $\Delta(\beta)$ is a linear function of n, then $\|\mathbf{m}\|_\infty$ is exponential in n.

Now we consider another example. We can generalise the $O(n \log n)$ bound for the linear function [5] to a more general non-negative coefficient function as described below.

Example 3. For $x = (s_1 \ldots s_n)$, define

$$f(x) = c_0 + \sum_{k=1}^{n} c_k s_k + \sum_{I \subseteq \{1, \ldots, n\}} c_I \prod_{k \in I} s_k, \tag{9}$$

where $c_k > 0$, $k = 1, \ldots, n$, $c_0 \geqslant 0$, and $c_I \geqslant 0$.

Here we give a new and shorter proof than that used in [5] based on Corollary 3.

Decompose the space S into $n + 1$ subspaces: let $x = (s_1 \cdots s_n)$,

$$S_l = \left\{ x; n - \sum_{k=1}^{n} s_k = l \right\}, \quad l = 0, \ldots, n.$$

Define $\mathbf{d} = [d_l]$ to be

$$d_l = \begin{cases} 0, & \text{if } l = 0, \\ 3e(1 + 1/2 + \cdots + 1/l)n, & \text{if } l > 0, \end{cases} \tag{10}$$

where e is the base of the natural logarithm. Let $d_i = d_l$ for any $i \in S_l$ ($l = 0, 1, \ldots, n$).

First we estimate transition probabilities among different subspaces. For $l = 1, \ldots, n$ and for any $i \in S_l$, let j be its offspring.

Event $j \in S_{l-k}$ ($k = 1, \ldots, l$) happens if k of the zero-valued bits in i are flipped and other bits in i are kept unchanged. Since $f(j) > f(i)$, j is always selected. Then we have, for the given k,

$$\sum_{j \in S_{l-k}} p(i, j) \geqslant \binom{l}{k} \left(\frac{1}{n} \right)^k \left(1 - \frac{1}{n} \right)^{n-k}.$$

Event $j \in S_{l+v}$ ($v = 1, \ldots, n - l$) happens only if the following event happens: k ($k = 1, \ldots, l$) of the zero-valued bits in i are flipped; $k + v$ of the one-valued bits in i

are also flipped; other bits in i are kept unchanged. Since $f(j)$ may be smaller than $f(i)$, j is not always selected to survive. Then we have, for the given v,

$$\sum_{j \in S_{l+v}} p(i, j) \leqslant \sum_{k=1}^{l} \binom{l}{k} \binom{n-l}{k+v} \left(\frac{1}{n}\right)^{2k+v} \left(1 - \frac{1}{n}\right)^{n-2k-v}.$$

Hence we get

$$\sum_{k=0}^{n} \sum_{j \in S_k} (d_l - d_k) p(i, j)$$

$$= \sum_{k=1}^{l} (d_l - d_{l-k}) \sum_{j \in S_{l-k}} p(i, j) + \sum_{v=1}^{n-l} (d_l - d_{l+v}) \sum_{j \in S_{l+v}} p(i, j)$$

$$\geqslant \sum_{k=1}^{l} (d_l - d_{l-k}) \binom{l}{k} \left(\frac{1}{n}\right)^{k} \left(1 - \frac{1}{n}\right)^{n-k}$$

$$+ \sum_{k=1}^{l} \sum_{v=1}^{n-l-k} (d_l - d_{l+v}) \binom{l}{k} \binom{n-l}{k+v} \left(\frac{1}{n}\right)^{2k+v} \left(1 - \frac{1}{n}\right)^{n-2k-v}$$

$$= \sum_{k=1}^{l} \binom{l}{k} \left(\frac{1}{n}\right)^{k} \left(1 - \frac{1}{n}\right)^{n-k}$$

$$\times \left(d_l - d_{l-k} + \sum_{v=1}^{n-l-k} (d_l - d_{l+v}) \binom{n-l}{k+v} \left(\frac{1}{n-1}\right)^{k+v}\right).$$

Since for any $l \geqslant 1$, $n - 1 \geqslant n - l$, we have

$$d_l - d_{l-k} + \sum_{v=1}^{n-l-k} (d_l - d_{l+v}) \binom{n-l}{k+v} \left(\frac{1}{n-1}\right)^{k+v}$$

$$\geqslant d_l - d_{l-k} + \sum_{v=1}^{n-l-k} (d_l - d_{l+v}) \frac{1}{(v+1)!}$$

$$\geqslant 3en \left(\frac{1}{l-k+1} + \cdots + \frac{1}{l} - \sum_{v=1}^{\infty} \left(\frac{1}{l+1} + \cdots + \frac{1}{l+v}\right) \frac{1}{(v+1)!}\right)$$

$$\geqslant 3en \left(\frac{1}{l-k+1} + \cdots + \frac{1}{l} - \frac{5}{6(l+1)} - \frac{1}{l+1} \sum_{v=3}^{\infty} \frac{1}{v!}\right)$$

$$\geqslant \begin{cases} 1.5e(3.5-e)n, & \text{if } l = 1, \\ 3e(3.5-e)n/l, & \text{if } l > 1, \end{cases}$$

$$\geqslant en/l.$$

In the above, we use the infinite sum: $e = \sum_{v=0}^{\infty} 1/v!$.

Finally, we obtain

$$\sum_{k=0}^{n} \sum_{j \in S_k} (d_l - d_k) p(i, j) \geqslant \sum_{k=1}^{l} \binom{l}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k} \frac{en}{l} \geqslant 1.$$

According to Corollary 3, we know that $m_i \leqslant d_i = d_l$ for any $i \in S_l$ where $l = 0, \ldots, n$.

4.2. Elitist selection: Accepting equal-valued individuals

We consider a general $(1 + 1)$ EA using Mutation I and the following selection:

Selection II. If $f(X^{(m)}) \geqslant f(X_t)$, then $X_{t+1} = X^{(m)}$; otherwise, $X_{t+1} = X_t$.

For the absorbing Markov chain (X_t) associated with the EA, its transition probability between $i \in S_l$ and $j \in S_k$ $(l, k = 0, \ldots, L)$ is

$$p(i, j) = \begin{cases} 1, & i \in S_0, \ j = i, \\ 0, & i \in S_0, \ j \neq i, \\ \geqslant 0, & i \in S_l, \ j \in S_k, \ k \leqslant l, \\ 0, & \text{otherwise.} \end{cases}$$

In matrix form, we have

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

However, $\mathbf{T}$ is not a lower triangular matrix in this case.

Theorem 7. *For an EA using Mutation* I *and Selection* II*, its first hitting time is given by*

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

We can obtain the explicit expression of $\mathbf{m}$ only for some matrices $\mathbf{T}$ of very simple forms. In the following we give an example to show that when $\mathbf{T}$ is a tridiagonal matrix, we can get the explicit expression of $\mathbf{m}$.

Example 4.

$$f(x) = \begin{cases} \sum s_i, & \text{if } \sum_{i=1}^{n-1} s_i s_{i+1} = 0, \\ \text{infeasible,} & \text{if } \sum_{i=1}^{n-1} s_i s_{i+1} > 0, \end{cases}$$

where $x = \{s_1 \ldots s_n\}$ and n is assumed to be an odd integer.

This is a constraint optimisation problem, which can be regarded as a simple maximum matching problem [23]: Let $G = (V, E)$ be a graph, where node set $V = \{v_0, \ldots, v_n\}$ and edge set $E = \{(v_0, v_1), \ldots, (v_{n-1}, v_n)\}$. Then $s_i = 1$ indicates that edge (v_{i-1}, v_i) appears in the matching and $s_i = 0$ means that (v_{i-1}, v_i) is not in the matching.

Consider an EA for solving this problem. If the EA employs Mutation I, it will generate more infeasible than feasible solutions. Instead, we use the following mutation, which can

generate feasible solutions more efficiently: Choose a position k from $\{1, \ldots, n\}$ at random and generate an offspring $x' = (s'_1 \cdots s'_n)$ as follows:

Case 1. If $1 < k < n - 1$, then

(1) if $s_k = 1$, then let $x' = x$,
(2) if $s_k = 0, s_{k-1} = 1, s_{k+1} = 1$, then let $x' = x$,
(3) if $s_k = 0$, either $s_{k-1} = 0$ or $s_{k+1} = 0$, then let $s'_k = 1, s'_{k-1} = s'_{k+1} = 0$, and other bits unchanged.

Case 2. If $k = 1$, then

(1) if $s_1 = 1$, then let $x' = x$,
(2) if $s_1 = 0$, then let $s'_1 = 1, s'_2 = 0$, and other bits unchanged.

Case 3. If $k = n$, then

(1) if $s_n = 1$, then let $x' = x$,
(2) if $s_n = 0$, then let $s'_n = 1, s'_{n-1} = 0$, and other bits unchanged.

Assume that the initial individual $\xi_0 = (s_1 \ldots s_n)$ satisfies $s_1 = s_3 = \cdots = s_n = 0$ and $s_2 = s_4 = \cdots = s_{n-1} = 1$. In the following, we compute the expected first hitting time $\mathbb{E}[\tau \mid X_0]$.

Since the EA adopts an elitist selection strategy and $f(X_0) \geqslant (n-1)/2$, the behaviour of Markov chain $(Xi_t, t = 0, 1, \cdots)$ is restricted in the following state space:

$$S = \{x; f(x) \geqslant (n-1)/2\}.$$

S can be divided into two subspaces:

$$S_0 = \{x; f(x) = (n+1)/2\}, \quad \text{and} \quad S_1 = \{x; f(x) = (n-1)/2\},$$

where S_1 can be decomposed further based on the length of augmenting paths [23]:

$$SH_k = \{x; x \in S_1 \text{ and its minimum length of augmenting path is } k\},$$

$$k = 1, 3, \ldots, n.$$

The transition probabilities among different subspaces can be computed exactly:

(1) If $i \in SH_1$, then

$$p(i, j) = \begin{cases} 1/n, & j \in S_0, \\ 2/n, & j \in SH_3, \\ 1 - 3/n, & j \in SH_1. \end{cases}$$

(2) If $i \in SH_l$ $(l = 2, \ldots, n - 1)$, then

$$p(i, j) = \begin{cases} 2/n, & j \in SH_{l-2}, \\ 2/n, & j \in SH_{l+2}, \\ 1 - 4/n, & j \in SH_l. \end{cases}$$

(3) If $i \in SH_n$, then

$$p(i, j) = \begin{cases} 2/n, & j \in SH_{n-2}, \\ 1 - 2/n, & j \in SH_n. \end{cases}$$

Obviously the behaviour of the EA could be regarded as a random walk. Since the transition matrix $\mathbf{P}$ is a tridiagonal matrix, it is easy to obtain an explicit solution to the linear system $(\mathbf{I} - \mathbf{T})\mathbf{m} = \mathbf{1}$ and obtain the expected first hitting time:

$$\mathbb{E}[\tau \mid X_0] = n^2 + (n^3 - n^2)/4.$$

In the following, we will prove that for any fitness function, given any subspace S_l $(l > 0)$, if the initial individual is taken from S_l, then the EA using Selection II will arrive at a subspace with higher fitness not slower than the EA using Selection I, both in terms of worse case and average case.

Given a subspace S_l $(l = 1, \ldots, L)$ for the Markov chain (X_t) associated with an EA, let $S_l^+ = \bigcup_0^{l-1} S_l$ be the subspaces with higher fitness. Denote

$$\tau_l^+(i) = \min\{t; X_t \in S_l^+, X_0 = i \in S_1\}, \tag{11}$$

to be the first hitting time of the EA to arrive at subspace S_l^+ starting from state i and its expectation is

$$m_l^+(i) = \mathbb{E}[\tau_l^+(i); \tau_l^+(i) < \infty]. \tag{12}$$

For simplicity, we ignore the subscript l in the following discussion.

Theorem 8. *Given any fitness function, for the Markov chain (X_{H_t}) associated with the EA using Mutation* I *and Selection* I *and the Markov chain X_{E_t} associated with the EA using Mutation* I *and Selection* II*, let their mean first hitting times be $\mathbf{m}_H^+ = [m_H^+(i)]_{i \in S_l}$ and $\mathbf{m}_E^+ = [m_E^+(i)]_{i \in S_l}$, where $m^+(i)$ is defined by (12). Then*

$$\|\mathbf{m}_H^+\|_\infty \geqslant \|\mathbf{m}_E^+\|_\infty.$$

If $\mathbf{T}_E$ is symmetric, we also have

$$\|\mathbf{m}_H^+\|_1 \geqslant \|\mathbf{m}_E^+\|_1.$$

Proof. In the following we use the subscript H to represent the EA using Mutation I and Selection I, and the subscript E to represent the EA using Mutation I and Selection II.

For the Markov chain (X_{H_t}), the EA uses elitist selection strategy. If the initial individual is taken from S_l, then its offspring always belongs to $S^+ \cup S_l$. So we only need to consider its behaviours in the space $\{S^+, S_l\}$. Since we only consider the first hitting time m^+, the transition probability is given by

$$p_H(i, j) = \begin{cases} 1, & i \in S^+, \ j = i, \\ 0, & i \in S^+, \ j \neq i, \\ \geqslant 0, & i \in S_l, \ j \in S^+, \\ \geqslant 0, & i \in S_l, \ j = i, \\ = 0, & i \in S_l, \ j \neq i, \\ 0, & \text{otherwise.} \end{cases}$$

In matrix form,

$$\mathbf{P}_H = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R}_H & \mathbf{T}_H \end{pmatrix},$$

where $\mathbf{T}_H$ is a diagonal matrix.

For the Markov chain (X_{E_t}), we only need to consider its behaviours in the space $\{S^+, S_l\}$. The transition probability is given by

$$p_E(i, j) = \begin{cases} 1, & i \in S^+, \ j = i, \\ 0, & i \in S^+, \ j \neq i, \\ \geqslant 0, & i \in S_l, \ j \in S^+, \\ \geqslant 0, & i \in S_l, \ j \in S_l, \\ 0, & \text{otherwise.} \end{cases}$$

In matrix form,

$$\mathbf{P}_E = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R}_E & \mathbf{T}_E \end{pmatrix}.$$

Since these two EAs use the same mutation, for any $i \in S_l$ and $j \in S^+$, $p_H(i, j) = p_E(i, j)$, or $\mathbf{R}_E = \mathbf{R}_H$. Hence, $\|\mathbf{T}_H\|_\infty = \|\mathbf{T}_E\|_\infty$.

(1) Since the entries of $(\mathbf{I} - \mathbf{T}_H)^{-1}$ are non-negative, we have

$$\|\mathbf{m}_H^+\|_\infty = \left\|(\mathbf{I} - \mathbf{T}_H)^{-1}\mathbf{1}\right\|_\infty = \left\|(\mathbf{I} - \mathbf{T}_H)^{-1}\right\|_\infty = \left\|\sum_{k=0}^{\infty}(\mathbf{T}_H)^k\right\|_\infty,$$

and since $\mathbf{T}_H$ is a diagonal matrix, we have

$$\|\mathbf{m}_H^+\|_\infty = \sum_{k=0}^{\infty}\left\|(\mathbf{T}_H)^k\right\|_\infty = \sum_{k=0}^{\infty}\|\mathbf{T}_H\|_\infty^k.$$

Since the entries of $(\mathbf{I} - \mathbf{T}_E)^{-1}$ are non-negative, we have

$$\|\mathbf{m}_E^+\|_\infty = \left\|(\mathbf{I} - \mathbf{T}_E)^{-1}\mathbf{1}\right\|_\infty = \left\|(\mathbf{I} - \mathbf{T}_E)^{-1}\right\|_\infty = \left\|\sum_{k=0}^{\infty}(\mathbf{T}_E)^k\right\|_\infty \leqslant \sum_{k=0}^{\infty}\|\mathbf{T}_E\|_\infty^k.$$

Because $\|\mathbf{T}_H\|_\infty = \|\mathbf{T}_E\|_\infty$,

$$\|\mathbf{m}_E^+\|_\infty \leqslant \sum_{k=0}^{\infty}\|\mathbf{T}_E\|_\infty^k = \sum_{k=0}^{\infty}\|\mathbf{T}_H\|_\infty^k.$$

Therefore, $\|\mathbf{m}_H^+\|_\infty \geqslant \|\mathbf{m}_E^+\|_\infty$.

(2) Since $\mathbf{T}_H$ is a diagonal matrix and $\mathbf{T}_E$ is symmetric, we have

$$\begin{aligned}
&\|\mathbf{m}_H^+\|_1 - \|\mathbf{m}_E^+\|_1 \\
&= \mathbf{1}^t(\mathbf{I} - \mathbf{T}_H)^{-1}\mathbf{1} - \mathbf{1}^t(\mathbf{I} - \mathbf{T}_E)^{-1}\mathbf{1} \\
&= \mathbf{1}^t(\mathbf{I} - \mathbf{T}_H)^{-1/2}(\mathbf{I} - (\mathbf{I} - \mathbf{T}_H)^{1/2}(\mathbf{I} - \mathbf{T}_E)^{-1}(\mathbf{I} - \mathbf{T}_H)^{1/2})(\mathbf{I} - \mathbf{T}_H)^{-1/2}\mathbf{1}.
\end{aligned}$$

If we can prove that the eigenvalue of matrix $(\mathbf{I} - \mathbf{T}_H)^{1/2}(\mathbf{I} - \mathbf{T}_E)^{-1}(\mathbf{I} - \mathbf{T}_H)^{1/2}$ or matrix $(\mathbf{I} - \mathbf{T}_H)(\mathbf{I} - \mathbf{T}_E)^{-1}$ is not more than 1, then we will get our conclusion:

$\|\mathbf{m}_H^+\|_1 \geqslant \|\mathbf{m}_E^+\|_1$. In the following we only need to prove that the eigenvalue of matrix $(\mathbf{I} - \mathbf{T}_E)(\mathbf{I} - \mathbf{T}_H)^{-1}$ is not less than 1.

Denote

$$
\mathbf{T}_H = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & b_{kk} \end{pmatrix}, \qquad \mathbf{T}_E = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ \vdots & \vdots & \vdots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kk} \end{pmatrix}.
$$

Then

$$
\mathbf{Q} = (\mathbf{I} - \mathbf{T}_E)(\mathbf{I} - \mathbf{T}_H)^{-1} = \begin{pmatrix} \dfrac{1-c_{11}}{1-b_{11}} & \dfrac{-c_{12}}{1-b_{22}} & \cdots & \dfrac{-c_{1k}}{1-b_{kk}} \\ \vdots & \vdots & \vdots & \\ \dfrac{-c_{k1}}{1-b_{11}} & \dfrac{-c_{k2}}{1-b_{22}} & \cdots & \dfrac{1-c_{kk}}{1-b_{kk}} \end{pmatrix}.
$$

Since the two EAs use the same mutation, we have $\mathbf{R}_E = \mathbf{R}_H$, and $b_{ii} = \sum_{j=1}^k c_{ij}$.

Let $\lambda(\mathbf{Q}^t)$ be the eigenvalue of $\mathbf{Q}^t$. According to Geršgorin's theorem [17], we know that λ lies in at least one of the disks:

$$
\left| \lambda - \frac{1-c_{jj}}{1-b_{jj}} \right| \leqslant \sum_{i \neq j} \frac{c_{ji}}{1-b_{jj}}, \quad j = 1, \ldots, k.
$$

Then

$$
\lambda \geqslant \frac{1-c_{jj}}{1-b_{jj}} - \sum_{i \neq j} \frac{c_{ji}}{1-b_{jj}} = 1.
$$

We obtain $\lambda(\mathbf{Q}^t) \geqslant 1$ and then $\lambda(\mathbf{Q}) \geqslant 1$. Hence $\|\mathbf{m}_H^+\|_1 \geqslant \|\mathbf{m}_E^+\|_1$. $\square$

4.3. Non-elitist selection strategies

Consider the following non-elitist selection strategy.

Selection III. Assign the survival probabilities of the better and worse individuals between $X^{(m)}$ and X_t to be p and $q = 1 - p$, respectively.

The fitness proportional (roulette wheel) selection scheme is a well-known example of Selection III: Assume f_1 is higher between $f(X^{(m)})$ and $f(X_t)$, then

$$
p = \frac{f_1}{f_1 + f_2}, \qquad q = \frac{f_2}{f_1 + f_2}.
$$

As we have discussed in the previous section, we only consider the behaviour of the chain up to the first hitting time. The EA can still be modeled by an absorbing Markov chain $(\overline{X}_t)$:

$$
p(i, j) = \begin{cases} 1, & i \in H, \ j = i, \\ 0, & i \in H, \ j \neq i, \\ \mathbb{P}(X_{t+1} = j \mid X_t = i), & i \notin H. \end{cases} \tag{13}
$$

In matrix form,

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

Here the first hitting time is still the number of generations when the EA first encounters an optimal solution. However, the found optimal point may be lost in the future.

The following result can be established from Theorem 1.

Theorem 9. *For the EA using Mutation* I *and Selection* III*, its first hitting time is given by*

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

The first question which we are interested in is whether we can give an explicit solution to the above linear system. Since $\mathbf{T}$ is no longer a lower-triangular matrix, we cannot get an explicit solution to the linear system in most cases. But for some simple functions and mutation operators, we are able to derive an explicit solution to $\mathbf{m}$. Here we give such an example.

Example 5. In the bit-sum function (8), assume c_n is the largest among all c_l where $l = 0, \ldots, n$.

Consider a $(1 + 1)$ EA using Selection III and the following mutation:

Mutation III. Given an individual $X_t = (s_1 \ldots s_n)$ at time t, choose one bit from $(s_1 \cdots s_n)$ uniformly at random and flip it.

We can decompose space S into $n + 1$ subspaces:

$$S_l = \left\{ x; n - \sum_{k=1}^{n} s_k = l, \right\}, \quad l = 0, \ldots, n.$$

The transition probabilities among the subspaces can be summarised as follows.

(1) For any $i \in S_0$,

$$p(i, j) = \begin{cases} 1, & j = i, \\ 0, & \text{otherwise;} \end{cases}$$

(2) For any $i \in S_k$ $(k = 1, \ldots, n - 1)$

$$p(i, j) = \begin{cases} a_k, & j \in S_{k-1}, \ D(j, i) = 1, \\ b_k, & j \in S_{k+1}, \ D(j, i) = 1, \\ 1 - a_k - b_k, & j = i, \\ 0, & \text{otherwise;} \end{cases}$$

(3) For any $i \in S_k$ $(k = n)$

$$p(i, j) = \begin{cases} a_k, & j \in S_{k-1}, \ D(j, i) = 1, \\ 1 - a_k, & j = i, \\ 0, & \text{otherwise;} \end{cases}$$

where a_k and b_k are

$$a_k = \begin{cases} p\frac{k}{n}, & \text{if } c_{n-k} < c_{n-k+1}, \\ q\frac{k}{n}, & \text{if } c_{n-k} > c_{n-k+1}, \end{cases}$$

$$b_k = \begin{cases} p\frac{n-k}{n}, & \text{if } c_{n-k} < c_{n-k-1}, \\ q\frac{n-k}{n}, & \text{if } c_{n-k} > c_{n-k-1}. \end{cases}$$

Construct an auxiliary Markov chain $(Y_t; t = 0, 1, \ldots,)$ which is defined on the state space $\{0, 1, \ldots, n\}$ with the transition matrix

$$\mathbf{P}_Y = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ a_1 & 1 - a_1 - b_1 & b_1 & 0 & \cdots & 0 & 0 \\ 0 & a_2 & 1 - a_1 - b_1 & b_2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & a_n & 1 - a_n \end{pmatrix},$$

where the subscript Y here and below is used to distinguish Markov chain (Y_t) from Markov chain (X_t).

For this absorbing Markov chain, its mean first hitting time to the absorbing state, $\mathbf{m}_Y = (\mathbf{I} - \mathbf{T}_Y)^{-1}\mathbf{1}$, is given by

$$m_{Y0} = 0,$$

$$m_{Yk} = m_{Yk-1} + \frac{1}{a_k} + \sum_{j=0}^{n-k-1} \frac{1}{a_{j+k+1}} \prod_{i=0}^{j} \frac{b_{k+i}}{a_{k+i}}, \quad k = 1, \ldots, n-1,$$

$$m_{Yn} = m_{Yn-1} + \frac{1}{a_n}.$$

For the original Markov chain (X_t), for any state $i \in S_k$, we have $m_i = m_{Yk}$.

For this example, we discuss how to choose p and q. Assume $p = p(n)$ and $q = q(n)$ are functions of n, there are two interesting cases: (i) For all integer n, $q(n)/p(n) \geqslant C$, where $C > 0$ is a constant, and (ii) $\lim_{n \to \infty} q(n)/p(n) = 0$.

The first choice will not be a good selection strategy for some problems. We can prove that for the ONE-MAX problem, which satisfies: $c_0 < c_1 < \cdots < c_n$, $\|\mathbf{m}\|_\infty$ is exponential if $q/p \geqslant C$, where $C > 0$ is a constant.

Substituting a_k and b_k into the solutions given previously, we have

$$m_{Y_1} = \frac{n}{p} + \sum_{j=0}^{n-2} \frac{n}{(j+2)p} \left(\prod_{i=0}^{j} \frac{n-1-i}{1+i} \right) \left(\frac{q}{p} \right)^j.$$

For any small constant $\varepsilon \in (0, 1)$ and for the fixed $q/p = C > 0$, when n is sufficiently large, there exists some integer $j_0 \in (0, n/3]$ such that for any $i \leqslant j_0$,

$$C \frac{n-1-i}{1+i} \geqslant 1 + \varepsilon.$$

In fact, we can choose j_0 to be the integer below closest to (when n is sufficiently large)

$$\min \left\{ \frac{pn}{(1+\varepsilon)q + p} - 1, n \right\}.$$

Then we have

$$m_{Y_1} \geqslant \prod_{i=0}^{j_0} \left(C \frac{n-1-i}{1+i} \right) \geqslant (1+\varepsilon)^{j_0}.$$

Note that j_0 is a linear function of n and ε is a constant in $(0, 1)$. So m_{Y_1} is an exponential function of n.

The second choice, $\lim_{n\to\infty} q(n)/p(n) = 0$, is much better for the ONE-MAX problem. For example, let $p = 1 - 1/n$ and $q = 1/n$, then $\|\mathbf{m}\|_\infty$ is polynomial in n for the ONE-MAX problem.

It is interesting to compare EAs using Selection I (elitist) and those using Selection III (non-elitist). Intuitively a non-elitist selection strategy appears to be worse than an elitist one. However, for some functions belonging to Hard Function I, a non-elitist selection strategy might be better, because it can "climb" out of the absorbing area of a local optimal solution. In the following, we will compare two EAs: one uses Mutation I and Selection I, and the other uses Mutation I and Selection III.

Example 6. Assume L is a linear function of n, the state space S is decomposed into $L+1$ subspaces as follows:

$$\begin{aligned}
S_1 &= \{\bar{x}\}, \\
S_k &= \{x\colon D(x, \bar{x}) = k\}, \quad k = 2, \ldots, L, \\
S_0 &= \{x\colon D(x, \bar{x}) \geqslant L+1\},
\end{aligned}$$

and $f(x)$ satisfies

$$f(x) = c_k, \quad x \in S_k, \tag{14}$$

where $c_0 > c_1 > \cdots > c_L$.

In the following we use subscripts O and H to represent the EA with a non-elitist selection strategy and that with accepting only higher-fitness individuals.

For the EA using Mutation I and Selection I, its transition matrix can be written as

$$\begin{pmatrix}
\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{R}_{H10} & \mathbf{T}_{H11} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{R}_{H20} & \mathbf{T}_{H21} & \mathbf{T}_{H22} & \mathbf{0} & \cdots & \mathbf{0} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\mathbf{R}_{HL0} & \mathbf{T}_{HL1} & \mathbf{T}_{HL2} & \mathbf{T}_{HL3} & \cdots & \mathbf{T}_{HLL}
\end{pmatrix}.$$

Let $i = \bar{x}$, then

$$\|\mathbf{m}_H\|_\infty = \frac{1}{\sum_{j \in S_0} p(i, j)}.$$

Denote $r = \sum_{j \in S_0} p(i, j)$.

For the EA using Mutation I and Selection III, its transition matrix can be written as

$$
\begin{pmatrix}
\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{R}_{O_{10}} & \mathbf{T}_{O_{11}} & \mathbf{T}_{O_{12}} & \mathbf{T}_{O_{13}} & \cdots & \mathbf{T}_{O_{1L}} \\
\mathbf{R}_{O_{20}} & \mathbf{T}_{O_{21}} & \mathbf{T}_{O_{22}} & \mathbf{T}_{O_{23}} & \cdots & \mathbf{T}_{O_{2L}} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\mathbf{R}_{O_{L0}} & \mathbf{T}_{O_{L1}} & \mathbf{T}_{O_{L2}} & \mathbf{T}_{O_{L3}} & \cdots & \mathbf{T}_{O_{LL}}
\end{pmatrix}.
$$

Since the two EAs use the same mutation, we have for any $i \in S_k$ $(k = 1, \ldots, L)$ and $j \in S_0$, $p_H(i, j) = p_O(i, j)$. Then

$$\|\mathbf{T}_O\|_\infty = \|\mathbf{T}_H\|_\infty = r.$$

For all $l = 1, \ldots, L - 1$ and $k = l, \ldots, L$, we know that $\mathbf{T}_{Olk} \geqslant 0$. So we have

$$\left\|\mathbf{T}_O^v\right\|_\infty \leqslant r^v$$

for any integer $v > 0$. The strict inequality "$<$" holds if v is greater than some positive.

From the above we can obtain for any integer $v > 0$:

$$
\begin{aligned}
\|\mathbf{m}_O\|_\infty &= \left\|(\mathbf{I} - \mathbf{T}_O)^{-1}\mathbf{1}\right\|_\infty \\
&= \left\|\sum_{k=1}^{\infty} \mathbf{T}_O^k\right\|_\infty \\
&= \left\|\left(\mathbf{I} + \cdots + \mathbf{T}_O^v\right)\sum_{k=0}^{\infty}(\mathbf{T}_O)^{vk}\right\|_\infty \\
&\leqslant \left\|\mathbf{I} + \cdots + \mathbf{T}_O^v\right\|_\infty \sum_{k=0}^{\infty} r^{vk}.
\end{aligned}
$$

As a result, we have

$$\frac{\|\mathbf{m}_O\|_\infty}{\|\mathbf{m}_H\|_\infty} \leqslant \frac{\|\mathbf{I} + \cdots + \mathbf{T}_O^v\|_\infty}{1 + \cdots + r^v} \leqslant 1.$$

The strict inequality "$<$" holds for v being greater than some positive.

5. Analysis of population-based evolutionary algorithms

We analysed $(1 + 1)$ EAs in previous sections. This section concentrates on EAs with population size greater than 1. EAs with and without crossover will be studied. In this section, we still limit our discussions to the binary representation of individuals.

5.1. Population-based evolutionary algorithms without crossover

Let's consider the following population-based EAs without crossover:

Mutation IV. Any proper mutation operator. After mutation we get an intermediate population $X^{(m)}$.

Selection IV. Any elitist selection strategy which always keeps the best individual among populations X_t and $X^{(m)}$.

For a population $X = \{x_1, \dots, x_N\}$ with N individuals, define its fitness to be $f(X) = \max\{f(x); x \in X\}$. Similar to the analysis of $(1+1)$ EAs in previous sections, we can decompose the population space E into $L+1$ subspaces, i.e.,

$$E_l = \big\{X \in E;\ f(X) = f_l\big\}, \quad l = 0, \dots, L.$$

The transition matrix $\mathbf{P}$ of Markov chain (X_t) associated with an EA has a matrix form:

$$\mathbf{P} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{R}_{10} & \mathbf{T}_{11} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{R}_{20} & \mathbf{T}_{21} & \mathbf{T}_{22} & \mathbf{0} & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{R}_{L0} & \mathbf{T}_{L1} & \mathbf{T}_{L2} & \mathbf{T}_{L3} & \cdots & \mathbf{T}_{LL} \end{pmatrix}.$$

According to Theorem 1, the following can be established easily.

Theorem 10. *For an EA using Mutation* IV *and Selection* IV*, its first hitting time is given by*

$$\begin{aligned} \mathbf{m}_1 &= (\mathbf{I} - \mathbf{T}_{11})^{-1}\mathbf{1}, \\ \mathbf{m}_k &= (\mathbf{I} - \mathbf{T}_{kk})^{-1}(\mathbf{T}_{k1}\mathbf{m}_1 + \cdots + \mathbf{T}_{kk-1}\mathbf{m}_{k-1}), \quad k = 2, \dots, L. \end{aligned}$$

Since the matrix $\mathbf{T}_{kk}$ $(k = 1, \dots, L)$ usually is not in a simple form, we cannot get an explicit solution to the linear system. This is different from the situation for $(1+1)$ EAs.

In the remaining part of this subsection, we will compare population-based EAs and $(1+1)$ EAs. Although it appears to be a common sense that for a population-based EA, its first hitting time (the number of generations) is shorter than that of $(1+1)$ EAs, no formal proof is available so far.

In the following we use subscript H to represent the $(1+1)$ EA and P to indicate the population-based EA.

Since population space E has higher dimensions than individual space S, if we want to compare two EAs as fair as possible, we have to restrict the behaviour of Markov chain (X_{Pt}) on the states of S during the comparison.

Let I be a population, we assume individuals in I are sorted in a descend order (the individual with the same fitness are arranged in a random order among them). We pay our attention only to the first and also the best individual in a population. Denote $E(i\#)$ to be the set of populations in E whose first individual is i. For a metric function d_i in the individual space, we can define a metric function d_I in the population space E as follows: $d_I = d_i$ where i is the first individual in population I.

Theorem 11. *Let* $\mathbf{m}_H$ *be the first hitting time of the* $(1+1)$ *EA using Mutation* I *and Selection* I*, and for any* $i \in S$*,* $d_i = m_H(i)$*, and* $\mathbf{m}_P$ *be the first hitting time of the population-based EA using Mutation* IV *and Selection* IV*, and for any* $I \in E(i\#)$*,* $d_I = d_i$*. If for* $l = 1, \dots, L$ *and* $k = 0, \dots, l-1$*, for any* $i \in S_l$ *and for any* $I \in E(i\#)$*,*

$$\sum_{J \in E_k} (d_I - d_J) p_P(I, J) \geqslant (>) \sum_{j \in S_k} (d_i - d_j) p_H(i, j), \tag{15}$$

then for any individual $i \in S_l$ ($l = 1, \ldots, L$) and any population $I \in E(i\#)$,

$$m_H(i) \geqslant (>) m_P(I).$$

Proof. We only prove the case of "$\geqslant$". The case of "$>$" can be proven in a similar way.

For the absorbing Markov chain (X_{H_t}) associated with the $(1+1)$ EA, since $d_i = m_H(i)$, according to Theorem 4, we know for $l = 1, \ldots, L$, and any $i \in S_l$,

$$\sum_{k=0}^{l-1} \sum_{j \in S_k} (d_i - d_j) p_H(i, j) = 1.$$

For any population $I \in E(i\#)$, from (15), we have for $k = 0, \ldots, l-1$,

$$\sum_{J \in E_k} (d_I - d_J) p_P(I, J) \geqslant \sum_{j \in S_k} (d_i - d_j) p_H(i, j).$$

Then we get

$$\sum_{k=0}^{l-1} \sum_{J \in E_k} (d_I - d_J) p_P(I, J) \geqslant \sum_{k=0}^{l-1} \sum_{j \in S_k} (d_i - d_j) p_H(i, j) = 1.$$

By applying Corollary 3, we get $d_I \geqslant m_H(I)$.

Since $d_I = d_i = m_H(i)$, we have $m_H(i) \geqslant m_P(I)$. $\square$

Condition (15) means that for the population-based EA, its transition probability towards the subspaces with a higher fitness is greater than that of the $(1+1)$ EA using the same mutation. Many population-based EAs satisfy this condition. The simplest example is the population-based EA which runs N copies of $(1+1)$ EA simultaneously.

If a function is easy for $(1+1)$ EAs, it is still easy for the corresponding population-based EA. For both Hard Functions I and II, a population can shorten the mean first hitting times according to Theorem 11. However, for Hard Function I, if the function is hard for the $(1+1)$ EA, it is still hard for the population-based EA. Assume that, for any $i \in S_1$ and for the $(1+1)$ EA using Mutation I and Selection I,

$$m_H(i) = \frac{1}{\sum_{j \in S_0} p_H(i, j)}$$

is exponential in n. For the simplest population-based EA, which runs N $(1+1)$ EAs simultaneously, Hard Function I is still hard under $\|\mathbf{m}\|_\infty$.

Obviously, if the initial population starts from state $i \in S_1$, then

$$m_P(I) = \frac{1}{\sum_{J \in E_0} p_P(I, J)}$$

is still exponential in n if N is a linear function of n. This implies that $m_P(i)$ and $\|\mathbf{m}_P\|_\infty$ are still exponential in n. However, the probability of the population-based EA to find

an optimum in a polynomial time may be improved significantly, because the following decreases significantly as N increases if $|S(P)|/|S|$ is reasonably large:

$$\frac{|E(P)|}{|E|} = 1 - \frac{|E(NP)|}{|E|} = 1 - \left(1 - \frac{|S(P)|}{|S|}\right)^N.$$

Some examples have already been given to show the improvement [5].

Since the population-based EA always allows individuals with a lower fitness to survive in the next generation with a probability greater than 0, an individual with a lower fitness would behave the same as that in a $(1 + 1)$ EA using non-elitist selection. Similar to a $(1 + 1)$ EA using a non-elitist selection strategy, a population-based EA is able to "climb out" of the absorbing area of a local optimum.

5.2. Population-based evolutionary algorithms with crossover

In many EAs, crossover is used before mutation and selection. This subsection analyses EA with crossover, which is applied before Mutation IV and Selection IV. A general form of crossover can be described as follows.

Crossover I. Choose a set of individual pairs, (x, y), from population X_t, and recombine them with probability $p_c > 0$. A new intermediate population $X^{(c)}$ is formed after crossover (i.e., recombination).

The first hitting time of an EA with crossover is defined below, which is slightly different from the previous definition: Given an initial population $X_0 = i$, τ_i is defined as

$$\tau_i = \min\{t; t \geqslant 0, X_t^{(c)} \in H \text{ or } X_t^{(m)} \in H \mid X_0 = i\}.$$

Assume population space E is decomposed into $L + 1$ subspaces based on fitness, then the transition matrix $\mathbf{P}$ of Markov chain (X_t) has the following form:

$$\mathbf{P} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{R}_{10} & \mathbf{T}_{11} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{R}_{20} & \mathbf{T}_{21} & \mathbf{T}_{22} & \mathbf{0} & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{R}_{L0} & \mathbf{T}_{L1} & \mathbf{T}_{L2} & \mathbf{T}_{L3} & \cdots & \mathbf{T}_{LL} \end{pmatrix}.$$

Theorem 12. *For an EA using Crossover* I*, Mutation* IV *and Selection* IV*, its first hitting time is given by*

$$\mathbf{m}_0 = (\mathbf{I} - \mathbf{T}_{11})^{-1} \mathbf{1},$$
$$\mathbf{m}_l = (\mathbf{I} - \mathbf{T}_{ll})^{-1} (\mathbf{T}_{l1} \mathbf{m}_1 + \cdots + \mathbf{T}_{ll-1} \mathbf{m}_{l-1}), \quad l = 1, \ldots, L.$$

Similar to the case of Theorem 11, we can show that a population-based EA with crossover is better than a $(1 + 1)$ EA under some conditions.

Let I be a population whose individuals are sorted in a descend order (the individual with the same fitness are arranged in a random order among them). Denote $E(i\#)$ to be the set of populations in E whose first (and best) individual is i. For a metric function d_i in the

individual space, we can define a metric function d_I in the population space E as follows: $d_I = d_i$ where i is the first individual in population I.

Theorem 13. *Let* $\mathbf{m}_H$ *be the first hitting time of the* $(1+1)$ *EA using Mutation* I *and Selection* I*, and for any* $i \in S$*,* $d_i = m_H(i)$*, and* $\mathbf{m}_C$ *be the first hitting time of the population-based EA using Mutation* IV *and Selection* IV*, and for any* $I \in E(i\#)$*,* $d_I = d_i$*. If for* $l = 1, \ldots, L$ *and* $k = 0, \ldots, l-1$*, for any* $i \in S_l$ *and* $j \in S_k$*, and for any* $I \in E(i\#)$*,*

$$\sum_{J \in E_k} (d_I - d_J) p_C(I, J) \geqslant (>) \sum_{j \in S_k} (d_i - d_j) p_H(i, j),$$

then for any individual $i \in S_l$ $(l = 1, \ldots, L)$ *and any population* $I \in E(i\#)$*,*

$$m_H(i) \geqslant (>) m_C(I).$$

An important question is: When will EAs with crossover perform better than those without for hard functions? While the comparison between $(1+1)$ EAs and population-based EAs with crossover is interesting. It does not reveal a full picture of the exact role of crossover. It is important to compare theoretically EAs with and without crossover. We will use Hard Function I as an example in our comparison. The two EAs we are comparing will have the same mutation and selection. They differ only in crossover.

For Hard Function I, some EAs without crossover spend a long time trying to jump over a "wide gap" between E_1 and E_0. In the following we will focus on analysing the first hitting time from E_1 to E_0.

Now let's examine what will happen when crossover is added to the EA. If an EA uses crossover, then for some pairs of individuals in populations I in E_1, the crossover may bring some shortcuts for the individuals to reach a higher fitness subspace. That is,

$$\sum_{J \in E_0} p_C(I, J) > \sum_{J \in E_0} p_P(I, J),$$

where p_C and p_P are the transition probabilities of the EAs with and without crossover, respectively. For these populations, we would have $m_C(I) > m_P(I)$. However, not all populations have such shortcuts. If we decompose E_1 further into K non-overlapping subspaces EH_k $(k = 1, \ldots, K)$, then in different subspaces there are different number of shortcuts. The following corollary summarises this informal argument formally.

Corollary 9. *Given any instance of Hard Function* I*, let* $\mathbf{m}_P$ *be the first hitting time of the EA using Mutation* IV *and Selection* IV*, and* $\mathbf{m}_C$ *be the first hitting time of the EA using Crossover* I*, Mutation* IV *and Selection* IV*. Assume subspace* E_1 *is decomposed into* K *non-overlapping subspaces,* $EH_1, EH_2, \ldots, EH_K$*, and there exist* K *positives* $d_1, d_2, \ldots, d_K$*. Let* $d_i = d_l$ *for any* $i \in EH_l$ $(l = 1, \ldots, K)$ *and* $d_i = 0$ *for any* $i \in E_0$*. If*

(1) *for any* $l = 1, \ldots, K$*, and for any* $i \in EH_l$*,*

$$\sum_{j \in E_0} (d_i - d_j) p_C(i, j) + \sum_{k=1}^{K} \sum_{j \in EH_k} (d_i - d_j) p_C(i, j) \geqslant 1,$$

(2) *for any* $l = 1, \ldots, K$, *and for any* $i \in EH_l$,

$$\sum_{j \in E_0} (d_i - d_j) p_P(i, j) + \sum_{k=1}^{K} \sum_{j \in EH_k} (d_i - d_j) p_P(i, j) \leqslant 1,$$

then for $l = 1, \ldots, K$, *and any* $i \in EH_l$,

$$m_C(i) \leqslant d_l \leqslant m_P(i). \tag{16}$$

Proof. This is a direct result of Corollaries 3 and 4. $\quad\square$

In the following we will present an example of comparing EAs with and without crossover.

A bit-sum function, called the "jump" problem, was given in [16] to show that crossover was useful in reducing the first hitting times of EAs. However, their discussion only gave a total upper bound for all states based on $\|\mathbf{m}\|_1$. We will show here that a more detailed result can be established to separate upper or lower bounds for different states using our approach.

Example 7. A simple "wide gap" function is given by:

$$f(x) = \begin{cases} 1, & \text{if } \sum_{k=1}^{n} s_k \geqslant 11n/20, \\ 0, & \text{if } \sum_{k=1}^{n} s_k = n/2, \\ \text{infeasible}; & \text{otherwise,} \end{cases} \tag{17}$$

where, for convenience, we assume $n/20$ is an integer.

Consider a population-based EA using the following Mutation V and Selection V but without crossover. We will show that its mean first hitting time is exponential in n.

Mutation V. Given a population X_t, for each individual in it, flip each of its bits with probability $1/n$.

Selection V. Any elitist selection strategy that always abandons infeasible solutions.

The population space can be divided into two subspaces: E_0 which consists of populations with at least one individual being the optimal solution, and E_1 which includes all other populations.

Let

$$a(n) = \sum_{k=n/20}^{n} \binom{n/2}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k}$$

and

$$d_i = \begin{cases} 0, & \text{if } i \in E_0, \\ (1 - (1 - a(n))^N)^{-1}, & \text{if } i \in E_1. \end{cases}$$

Given a population i in E_1, the transition probability for each individual to the subspace E_0 will be less than $a(n)$.

We have

$$\sum_{j \in E_0} p(i, j) = 1 - \left(1 - a(n)\right)^N \quad \text{and} \quad \sum_{j \in E_0} (d_i - d_j) p_p(i, j) = 1.$$

We have now shown that, for any $i \in E_1$: $m_i = d_i$, m_i is exponential in n if population size N is a constant or a linear function of n.

Consider the EA with crossover, which employs only two individuals. It uses the same Mutation V and Selection V as given before and the following Crossover II.

Crossover II. Let x and y be the two individuals in population X_t. Their offspring are given by $x' = (x \text{ AND } y) = (x_k \text{ AND } y_k, k = 1, \ldots, n)$ and $y' = (x \text{ OR } y) = (x_k \text{ OR } y_k, k = 1, \ldots, n)$. If x' or y' is infeasible, then replace it by x or y.

Decompose E_1 further into subspaces based on the Hamming distance between x and y,

$$EH_{2l} = \left\{(x, y) \in E_1 : D(x, y) = 2l\right\}; \quad 2l = 0, 2, \ldots, n.$$

Let d_i be

$$d_i = \begin{cases} 0, & \text{if } i \in E_0; \\ 1, & \text{if } i \in EH_{2l}, \ 2l = n/20, \ldots, n, \\ 40(1 + n/8 - 2l)/(9\mathrm{e}); & \text{if } i \in EH_{2l}, \ 2l = 0, \ldots, n/20 - 1. \end{cases}$$

For any $2l$ ($n/20 \leqslant 2l \leqslant n$) and for any population $i \in EH_{2l}$, individuals x and y satisfy $D(x, y) \geqslant n/20$. Their offspring j will be in E_0 after crossover, i.e.,

$$\sum_{j \in E_0} p(i, j) = 1,$$

So we have

$$\sum_{j \in E_0} (d_i - d_j) p(i, j) + \sum_{2k=0}^{n} \sum_{j \in EH_{2k}} (d_i - d_j) p(i, j) = 1,$$
$$\forall i \in E_{2l}, \ 2l = n/20, \ldots, n. \tag{18}$$

For any $2l$ ($0 \leqslant 2l < n/20$) and for any $i \in E_{2l}$,

$$\sum_{j \in E_0} (d_i - d_j) p(i, j) + \sum_{2k=0}^{n} \sum_{j \in EH_{2k}} (d_i - d_j) p(i, j)$$

$$\geqslant \sum_{2k=2}^{n-2l} \sum_{j \in EH_{2l+2k}} (d_i - d_j) p(i, j) + \sum_{2k=2}^{2l} \sum_{j \in EH_{2l-2k}} (d_i - d_j) p(i, j)$$

$$\geqslant \sum_{2k=2}^{2l} \left(\sum_{j \in EH_{2l+2k}} (d_i - d_j) p(i, j) + \sum_{j \in EH_{2l-2k}} (d_i - d_j) p(i, j) \right).$$

We now need to estimate the following for $2k = 2, \ldots, 2l$:

$$\sum_{j \in EH_{2l+2k}} (d_i - d_j) p(i, j) + \sum_{j \in EH_{2l-2k}} (d_i - d_j) p(i, j). \tag{19}$$

(1) The case of $2k = 2$.

For any population $i = (x, y) \in E_{2l}$, since $D(x, y) = 2l$, there are $2l$ bit positions where x and y have different values and $n - 2l$ bit positions they share the same values. More precisely, the bit position set $\{1, \ldots, n\}$ can be grouped into four classes:

$PC_0(x, y)$: the set of positions where both x and y take the common value of 0;
$PC_1(x, y)$: the set of positions where both x and y take the common value of 1;
$PD_0(x, y)$: the set of positions where x takes the value of 0 but y takes the value of 1;
$PD_1(x, y)$: the set of positions where x takes the value of 1 but y takes the value of 0.

Since $0 \leqslant 2l < n/20$, the cardinalities of these four sets satisfy

$$|PC_0| = |PC_1| = n/2 - l > |PD_0| = |PD_1| = l.$$

First, let's examine the role of crossover.

Assume x' and y' are offspring of x and y after crossover. If $2l = 0$, then $x = y$, the crossover plays no role. If $0 < 2l < n/20$, that is, $0 < D(x, y) < n/20$, then the offspring are two infeasible solutions. We must replace them by their parents. The crossover plays no role either. No matter what $2l$ ($0 \leqslant 2l < n/20$) may be, we always have $x' = x$ and $y = y'$.

Second, let's examine the role of mutation.

Assume x' and y' are offspring of x and y after mutation (mutation is performed after crossover). During mutation, if both x and y change their values at the same bit position, there is no impact on Hamming distance $D(x, y)$. Only when an individual changes its values at some bit positions and another individual keeps its corresponding values unchanged will Hamming distance $D(x, y)$ be changed. So in the discussion below, we will use the term "a bit being flipped" in the following sense: x changes its value at this bit position but y does not, or y changes its value at this bit position but x does not.

We also notice that if a bit in set PC_0 and a bit in PC_1 are flipped, then Hamming distance $D(x', y')$ will increase. But if a bit in set PD_0 and a bit in PD_1 are flipped, then Hamming distance $D(x', y')$ decreases.

The event of $D(x', y') = 2l + 2$ consists of the following events:

(1) One bit in PC_0 and one bit in PC_1 are flipped, and others are kept unchanged. The event's probability is

$$\binom{n/2 - l}{1}^2 \left(\frac{1}{n}\right)^2 \left(1 - \frac{1}{n}\right)^{n-2}. \tag{20}$$

(2) Two bits in PC_1 and two bits in PC_0 are flipped. One bit in PD_0 and one bit in PD_1 are flipped. Others are kept unchanged. This event's probability is

$$\binom{n/2 - l}{2}^4 \binom{l}{1}^2 \left(\frac{1}{n}\right)^6 \left(1 - \frac{1}{n}\right)^{n-6}. \tag{21}$$

(3) Three bits in PC_1 and three bits in PC_0 are flipped, and so on.

The event of $D(x', y') = 2l - 2$ consists of the following events:

(1) One bit in PD_0 and one bit in PD_1 are flipped. Others are kept unchanged. The event's probability is

$$\binom{l}{1}^2 \left(\frac{1}{n}\right)^2 \left(1 - \frac{1}{n}\right)^{n-2}. \tag{22}$$

(2) Two bits in PD_0 and two bits in PD_1 are flipped. One bit in PC_1 and one bits in PC_0 are flipped. Others are kept unchanged. The event's probability is

$$\binom{l}{2}^4 \binom{n/2-l}{1}^2 \left(\frac{1}{n}\right)^6 \left(1 - \frac{1}{n}\right)^{n-6}. \tag{23}$$

(3) Three bits in PD_0 and three bits in PD_1 are flipped, and so on.

From the above inequalities and $2l < n/20$, we can obtain

$$\sum_{j \in EH_{2l+2}} p(i, j) - \sum_{j \in EH_{2l+2}} p(i, j)$$

$$= \left(\binom{n/2-l}{1}^2 - \binom{l}{1}^2 \right) \left(\frac{1}{n}\right)^2 \left(1 - \frac{1}{n}\right)^{n-2}$$

$$+ \left(\binom{n/2-l}{2}^4 \binom{l}{1}^2 - \binom{l}{2}^4 \binom{n/2-l}{1}^2 \right) \left(\frac{1}{n}\right)^6 \left(1 - \frac{1}{n}\right)^{n-6} + \cdots$$

$$\geqslant 9e/40.$$

Hence,

$$\sum_{j \in EH_{2l+2}} (d_i - d_j) p(i, j) + \sum_{j \in EH_{2l-2}} (d_i - d_j) p(i, j)$$

$$= 80/(9e) \left(\sum_{j \in EH_{2l+2k}} p(i, j) - \sum_{j \in EH_{2l-2k}} p(i, j) \right)$$

$$\geqslant 1.$$

(2) The cases of $2k = 4, \ldots, 2l$.
In a similar way, we can prove that for $2k = 4, \ldots, 2l$,

$$\sum_{j \in EH_{2l+2k}} (d_i - d_j) p(i, j) + \sum_{j \in EH_{2l-2k}} (d_i - d_j) p(i, j) \geqslant 0.$$

Combining the above two cases (1) and (2), we have for any $2l$ ($2l = 0, \ldots, n/20 - 1$) and for any population $i \in E_{2l}$,

$$\sum_{j \in E_0} (d_i - d_j) p(i, j) + \sum_{2k=0}^{n} \sum_{j \in E_{2k}} (d_i - d_j) p(i, j) \geqslant 1. \tag{24}$$

According to Corollary 9 and (18) and (24), we arrive at

$$m_i \leqslant d_i = \begin{cases} 0, & \text{if } i \in E_0, \\ 1, & \text{if } i \in EH_{2l}, \ 2l = n/20, \dots, n, \\ 40(1 + n/8 - 2l)/(9e); & \text{if } i \in EH_{2l}, \ 2l = 0, \dots, n/20 - 1, \end{cases}$$

which is polynomial in n.

Consider another EA using the following uniform crossover.

Crossover IV. Let x and y to be the individuals in population X_t, then independently for all positions $i \in \{1, \dots, n\}$, exchange s_{xi} and s_{yi} with probability 0.5. If x' or y' is infeasible, replace it by x or y.

In this example we will estimate the lower bound of **m** using our algebraic approach. We first decompose E_1 in the same way as we did previously for the EA with Crossover II.

Let

$$d_i = \begin{cases} 0, & \text{if } i \in E_0, \\ \dfrac{1}{2}\left(\dfrac{1}{a(n, n/2) + (1 - a(n, n/2))(1 - (1 - b(n))^2)} + \dfrac{n - 2l}{2e} \right), \\ & \text{if } i \in EH_{2l}, \ 2l = 0, \dots, n, \end{cases}$$

where

$$a(n, l) = \begin{cases} 0.5^{2l} \sum_{k, k'=1, |k'-k| \geqslant n/20}^{l} \binom{l}{k}\binom{l}{k'}, & \text{if } 2l = n/20, \dots, n, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$b(n) = \sum_{k=n/20}^{n/2} \binom{n/2}{k}\left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n/2-k} \sum_{v=0}^{k-n/20} \binom{n/2}{v}\left(\frac{1}{n}\right)^v \left(1 - \frac{1}{n}\right)^{n/2-v}.$$

For any $2l$ $(2l = 0, \dots, n)$ and for any population $i = (x, y) \in EH_{2l}$, we analyse the first hitting time in two steps.

(1) First we estimate the probability of event $X^{(c)}$ or $X^{(m)}$ entering E_0.

If $2l$ satisfies $0 \leqslant 2l < n/20$, similar to the previous analysis for Crossover II, crossover has no impact on Hamming distance $D(x, y)$. We only need to analyse mutation.

Let x' and y' be the offspring of x and y after mutation, then event $x' \in E_0$ happens if k of the zero-valued bits in x is flipped (k must be greater than $n/20$) and $v = 0, \dots, k - n/20$ of the one-valued bits in x is flipped. The probability of x' being in E_0 is

$$b(n).$$

The event of either x' or y' being in E_0 after mutation happens with a probability

$$1 - \left(1 - b(n)\right)^2. \tag{25}$$

If $2l$ satisfies $2l \geqslant n/20$, we can divide the bit position set $\{1, \ldots, n\}$ into four classes: $PC_0(x, y)$, $PC_1(x, y)$, $PD_0(x, y)$ and $PD_1(x, y)$, similar to the analysis of the EA with Crossover II.

We first analyse the role of crossover. Let x' and y' be the offspring of x and y after crossover. It is clear that if any bits in PC_0 and PC_1 are exchanged by crossover, x' and y' will not be changed. If the bits in PD_0 or PD_1 are exchanged by crossover, and the number of exchanged bits is greater than $n/20$, then one of x' or y' may enter set E_0. The probability of the event satisfies the bivariate binomial distribution, that is,

$$a(n, l), \quad l = n/20, \ldots, n.$$

We then analyse the role of mutation. If after crossover, one of the offspring enters set E_0, we say that the chain has already first-hit E_0. We only need to consider the case of neither x' nor y' in E_0. In this case, x' and y' are infeasible solutions. They will be replaced by x and y. For population (x, y), we know from (25) that its offspring enter set E_0 with probability

$$1 - \left(1 - b(n)\right)^2.$$

Hence, we can obtain

$$\sum_{j \in E_0} p(i, j) \leqslant a(n, l) + \left(1 - a(n, l)\right)\left(1 - \left(1 - b(n)\right)^2\right). \tag{26}$$

(2) Second we estimate the probability of event $X^{(c)}$ or $X^{(m)}$ entering EH_{2l+2k} ($2k = 2, \ldots, n - 2l$).

We first consider the role of crossover. Since the crossover always generates infeasible solutions in this case, it has no impact on $D(x, y)$. We will examine mutation's role.

Similar to the analysis of the EA with Crossover II, we know that this event happens only if at least k bits in PC_0 and k bits in PC_1 are flipped by mutation. The probability of this event is no more than

$$\sum_{j \in EH_{2l+2k}} p(i, j) \leqslant \binom{n/2 - l}{k}^2 \left(\frac{1}{n}\right)^{2k} \leqslant \frac{1}{2k!}. \tag{27}$$

According to (25), (26) and (27), we get

(1) For $2l = n$, and any population $i \in EH_{2l}$, since for any k ($k < l$) and $j \in S_{2k}$, $d_i < d_j$, then

$$\sum_{j \in E_0} (d_i - d_j) p(i, j) + \sum_{j \in E_1} (d_i - d_j) p(i, j)$$
$$\leqslant \sum_{j \in E_0} (d_i - d_j) p(i, j)$$
$$\leqslant \frac{1}{2} \left(\frac{a(n, n/2) + (1 - a(n, n/2))(1 - (1 - b(n))^2)}{a(n, n/2) + (1 - a(n, n/2))(1 - (1 - b(n))^2)} \right)$$
$$< 1.$$

(2) For any $2l$ ($n/20 \leqslant 2l < n$) and for any population $i \in E_{2l}$,

$$\sum_{j \in E_0} (d_i - d_j) p(i, j) + \sum_{j \in E_1} (d_i - d_j) p(i, j)$$

$$\leqslant \sum_{j \in E_0} (d_i - d_j) p(i, j) + \sum_{k=2}^{n-2l} \sum_{j \in EH_{2l+2k}} (d_i - d_j) p(i, j)$$

$$\leqslant \frac{1}{2} \left(\frac{a(n, l) + (1 - a(n, l))(1 - (1 - b(n))^2)}{a(n, n/2) + (1 - a(n, n/2))(1 - (1 - b(n))^2)} \right) + \sum_{k=1}^{n/2-l} \frac{1}{4e(k - 1)!}$$

$$< 1.$$

(3) For any $2l$ ($0 \leqslant 2l < n/20$) and for any population $i \in E_{2l}$, when n is sufficiently large,

$$\sum_{j \in E_0} (d_i - d_j) p(i, j) + \sum_{j \in E_1} (d_i - d_j) p(i, j)$$

$$\leqslant \sum_{j \in E_0} (d_i - d_j) p(i, j) + \sum_{k=2}^{n-2l} \sum_{j \in EH_{2l+2k}} (d_i - d_j) p(i, j)$$

$$\leqslant \frac{1}{2} \left(\frac{(1 - (1 - b(n))^2}{a(n, n/2) + (1 - a(n, n/2))(1 - (1 - b(n))^2)} \right) + \sum_{k=1}^{n/2-l} \frac{1}{4e(k - 1)!}$$

$$< 1.$$

Hence, from Corollary 4 we can obtain an estimation of the lower bound of the first hitting time:

$$m_i \geqslant d_i = \begin{cases} 0, & \text{if } i \in E_0, \\ \frac{1}{2} \left(\frac{1}{a(n, n/2) + (1 - a(n, n/2))(1 - (1 - b(n))^2)} + \frac{n - 2l}{2e} \right), \\ & \text{if } i \in EH_{2l}. \end{cases}$$

6. Conclusions and discussions

In spite of some successful applications of EAs in combinatorial optimisation, theoretical analyses of EA's computation time and time complexity are few. The existing approaches used to analysed EA's computation time appear to be ad hoc and do not seem to be easily generalisable to different problems. This paper gives a general framework for analysing EA's computation time on different problems. It enables us to derive both new and old results about EA's computation time under a unified framework.

Under this framework, conditions under which an EA will need polynomial (or exponential) mean computation time to solve a problem are studied. A number of case studies are given to illustrate how different results can be established by verifying these conditions. The general framework also helps us to understand better what causes an EA to have a polynomial (or exponential) time behaviour.

The general framework we gave is based on the linear equations (2) and its bounds of the first hitting time of an EA's Markov chain model. Once the transition matrix of an Markov chain is known, the first hitting time for the Markov chain will satisfy Eq. (2). In particular, we can obtain explicit solutions to these equations when the transition matrix is a lower triangular or a tridiagonal matrix. $(1 + 1)$ EAs using elitist selection which accepts only higher fitness and some $(1 + 1)$ EAs using non-elitist selection can be analysed in this way.

However, it may not be practical (or even possible) sometimes to solve these equations explicitly due to the huge number of them. Fortunately, Theorems 2 and 3 enable us to derive bounds of the first hitting time without solving the linear system.

In our framework, we have systematically investigated $(1 + 1)$ EAs using different selection strategies and population-based EAs with and without crossover. We have proved that hard problems to a simple $(1 + 1)$ EA can be classified into two classes: "wide gap" problems (e.g., Hard Function I) and "long path" problems (e.g., Hard Function II). We have shown that how some $(1 + 1)$ EAs and population-based EAs could deal with these hard problems better than others. These results give some interesting insights into the fundamental question: what makes a problem hard for an EA?

Within the same framework, we have also investigated the application of the analytic method to some special cases: we have generalised the $O(n \log n)$ bound on the linear function to a class of general positive coefficient functions and provided a new and shorter proof. An example that an EA with crossover is much better than that without one was also given.

This paper represents the first step toward analysing EAs in a greater depth and detail. One of our future work is to study EAs on some well-known combinatorial optimisation problems, rather than just some artificial binary problems.

Acknowledgements

References

[1] T. Bäck, The interaction of mutation rate, selection and self-adaptation within a genetic algorithm, in: Parallel Problem Solving from Nature 2, North-Holland, Amsterdam, 1992, pp. 85–94.

[2] T. Bäck, D.B. Fogel, Z. Michalewicz (Eds.), Handbook of Evolutionary Computation, Oxford University Press, Oxford, 1997.

[3] E. Cantú-Paz, Markov chain models of parallel genetic algorithms, IEEE Trans. Evolutionary Comput. 4 (3) (2000) 216–226.

[4] S. Droste, T. Jansen, I. Wegener, A rigorous complexity analysis of the $(1 + 1)$ evolutionary algorithm for linear functions with Boolean inputs, Evolutionary Comput. 6 (2) (1998) 185–196.

[5] S. Droste, T. Jansen, I. Wegener, On the analysis of the $(1 + 1)$ evolutionary algorithms, Theoret. Comput. Sci. 276 (1–2) (2002) 51–81.

[6] A.E. Eiben, G. Rudolph, Theory of evolutionary algorithms: A bird's eye view, Theoret. Comput. Sci. 229 (1/2) (1999) 3–9.

[7] D. Fogel, Evolution Computation: Toward a New Philosophy of Machine Intelligence, IEEE Press, Piscataway, NJ, 1995.

[8] J. Garnier, L. Kallel, Statistical distribution of the convergence time of evolutionary algorithms for long path problems, IEEE Trans. Evolutionary Comput. 4 (1) (2000) 16–30.

[9] J. Garnier, L. Kallel, M. Schoenauer, Rigorous hitting times for binary mutations, Evolutionary Comput. 7 (2) (1999) 167–203.

[10] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, MA, 1989.

[11] J. He, L. Kang, On the convergence rate of genetic algorithms, Theoret. Comput. Sci. 229 (1/2) (1999) 23–39.

[12] J. He, X. Yao, Drift analysis and average time complexity of evolutionary algorithms, Artificial Intelligence 127 (1) (2001) 57–85.

[13] J. He, X. Yao, From an individual to a population: An analysis of the first hitting time of population-based evolutionary algorithms, IEEE Trans. Evolutionary Comput. 6 (5) (2002) 495–511.

[14] J. He, X. Yao, Erratum to: Drift analysis and average time complexity of evolutionary algorithms [Artificial Intelligence 127 (2001) 57–85], Artificial Intelligence 140 (1) (2002) 245–248.

[15] M. Iosifescu, Finite Markov Processes and Their Applications, Wiley, Chichester, 1980.

[16] T. Jansen, I. Wegener, On the analysis of evolutionary algorithms—A proof that crossover really can help, in: J. Nesetril (Ed.), Proc. of the 7th Annual European Symposium on Algorithms (ES'99), Springer, Berlin, 1999, pp. 184–193.

[17] P. Lancaster, M. Tismenetsky, The Theory of Matrix, 2nd Edition, Academic Press, Orlando, FL, 1985.

[18] K.-H. Liang, X. Yao, C. Newton, D. Hoffman, A new evolutionary approach to cutting stock problems with and without contiguity, Comput. Oper. Res. 29 (12) (2002) 1641–1659.

[19] S.P. Meyn, R.L. Tweedie, Markov Chains and Stochastic Stability, Springer-Verlag, New York, 1993.

[20] Z. Michalewicz, Genetic Algorithms + Data Structure = Evolution Program, 3rd Edition, Springer-Verlag, New York, 1996.

[21] H. Mühlenbein, How genetic algorithms really works I: Mutation and hill-climbing, in: Parallel Problem Solving from Nature 2, North-Holland, Amsterdam, 1992, pp. 15–25.

[22] E.A. Nix, M.D. Vose, Modeling genetic algorithms with Markov chains, Ann. Math. Artificial Intelligence 5 (1992) 79–88.

[23] C.H. Papadimitriou, K. Steiglitz, Combinatorial Optimization: Algorithms and Complexity, Dover, Mineola, NY, 1998.

[24] G. Rudolph, How mutation and selection solve long path problems in polynomial expected time, Evolutionary Comput. 4 (2) (1996) 194–205.

[25] G. Rudolph, Convergence Properties of Evolutionary Algorithms, Kovač, Hamburg, Germany, 1997.

[26] G. Rudolph, Finite Markov chain results in evolutionary computation: A tour d'Horizon, Fundamenta Informaticae 35 (1–4) (1998) 67–89.

[27] H.-P. Schwefel, Evolution and Optimum Seeking, Wiley, New York, 1995.

[28] W.M. Spears, Evolutionary Algorithms: The Role of Mutation and Recombination, Springer, Berlin, 2000.

[29] J. Suzuki, A Markov chain analysis on simple genetic algorithms, IEEE Trans. Systems Man Cybernet. 25 (1995) 655–659.

[30] R. Syski, Passage Times for Markov Chains, IOS Press, Amsterdam, 1992.

[31] M.D. Vose, The Simple Genetic Algorithm: Foundations and Theory, MIT Press, Cambridge, MA, 1999.

[32] I. Wegener, Methods for the analysis of evolutionary algorithms on pseudo-Boolean functions, in: R. Sarker, M. Mohammadian, X. Yao (Eds.), Evolutionary Optimization, Chapter 14, Kluwer Academic, Boston, 2002, pp. 349–370.

[33] A.H. Wright, Y. Zhao, Markov chain models of genetic algorithms, in: W. Banzhaf, J. Daida, A.E. Eiben, M.H. Garzon, V. Honavar, M. Jakiela, R.E. Smith (Eds.), Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-99), Morgan Kauffman, San Francisco, CA, 1999, pp. 734–741.

[34] X. Yao (Ed.), Evolutionary Computation: Theory and Applications, World Scientific Publishing, Singapore, 1999.

[35] C. Zhang, X. Yao, J. Yang, An evolutionary approach to materialized views selection in a data warehouse environment, IEEE Trans. Systems Man Cybernet. Part C 31 (3) (2001) 282–294.