

Drift analysis and average time complexity of evolutionary algorithms

Jun He^a, Xin Yao^{b,*}

^a *Department of Computer Science, Northern Jiaotong University, Beijing 100044, PR China*

^b *School of Computer Science, The University of Birmingham, Birmingham B15 2TT, UK*

Received 19 July 1999; received in revised form 15 August 2000

Abstract

The computational time complexity is an important topic in the theory of evolutionary algorithms (EAs). This paper reports some new results on the average time complexity of EAs. Based on drift analysis, some useful drift conditions for deriving the time complexity of EAs are studied, including conditions under which an EA will take no more than polynomial time (in problem size) to solve a problem and conditions under which an EA will take at least exponential time (in problem size) to solve a problem. The paper first presents the general results, and then uses several problems as examples to illustrate how these general results can be applied to concrete problems in analyzing the average time complexity of EAs. While previous work only considered $(1 + 1)$ EAs without any crossover, the EAs considered in this paper are fairly general, which use a finite population, crossover, mutation, and selection. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Evolutionary algorithms; Time complexity; Random sequences; Drift analysis; Stochastic inequalities

1. Introduction

Evolutionary algorithms (EAs) are a powerful class of adaptive search algorithms [6,11,14]. They have been used to solve many combinatorial problems with success in recent years. However, theories on explaining why and how EAs work are still relatively few in spite of recent efforts [3]. The computational time complexity of EAs is largely unknown, except for a few simple cases [1,2,5,17,18]. Ambati et al. [1] and Fogel [5] estimated the computational time complexity of their EAs on the traveling salesman problem. No

* Corresponding author.

E-mail address: x.yao@cs.bham.ac.uk (X. Yao).

theoretical results were given. Rudolph [17] proved that $(1 + 1)$ EAs with mutation probability $p_m = 1/n$, where n is the number of bits in a binary string (i.e., individual) and p_m is the mutation probability, converge in average time $O(n \log n)$ for the ONE-MAX problem. Droste et al. [2] carried out a rigorous complexity analysis of $(1 + 1)$ EAs for linear functions with Boolean inputs. However, all of these results were based on EAs with a population size of 1 and without any crossover operators. Nimwegen et al. [22,23] developed a theory which predicts the total number of fitness function evaluations needed to reach a global optimum by epochal dynamics as a function of mutation rate and population size. However, no relationship to the problem size was studied. He et al. [8,9] showed that genetic algorithms (GAs) may take exponential average time to solve some deceptive problems.

This paper presents a more general theory about the average time complexity of EAs. The motivation of this study is to establish a general theory for a class of EAs, rather than a particular EA. The theory can then be used to derive specific complexity results for different EAs on different problems. The theory has been developed using drift analysis [7, 21]—a very useful technique in analyzing random sequences. It can be used to estimate the first hitting time by estimating the drift of a random sequence. To our best knowledge, this is the first attempt that drift analysis is introduced into the theoretical study of evolutionary computation. One of the major advantages of using drift analysis is that it is often easier to estimate the drift than to estimate the first hitting time directly. The techniques of drift analysis can also be applied to random sequences which are not Markovian [7].

The basic idea of this paper is as follows. We first model the evolution of an EA population as a random sequence, e.g., a Markov chain. A population of multiple individuals will be considered. Both crossover and mutation are included in the EA. Then we analyzed the drift of this sequence to and from the optimal solution (assuming we are solving an optimization problem). Various bounds on the first hitting time will be derived under different drift conditions. Some drift conditions cause the random sequence to drift away from the optimal solution, while other drift conditions enable the sequence to drift towards the optimal solution. We will study the conditions which are used to determine the time complexity of an EA to solve a problem, whether in polynomial time (in problem size) or in exponential time.

To illustrate the application of the above general theory, we will apply the theoretical results to several well-known problems, including a classical combinatorial optimization problem—the subset sum problem. It is shown in this paper that a certain family of subset sum problems can be solved by an EA within polynomial time, while other families of subset sum problems will need at least exponential time to solve. Although the EAs used in our study do not include all possible variations of EAs, they do represent a fairly large class of EAs which have multiple individuals and use both crossover and mutation.

The rest of this paper is organized as follows: Section 2 introduces briefly EAs and drift analysis. Section 3 studies the conditions under which EAs can solve a problem within polynomial time on average. A general theorem is first presented. Then examples, including the subset sum problem, are studied to show the application of the theorem. Section 4 studies the conditions under which EAs need at least exponential computation time to solve a problem. Both a general theorem and an application of the theorem are

given. Section 5 discusses some weaker drift conditions for the subset sum problem. Finally, Section 6 concludes with a brief summary of the paper and some future work.

2. Evolutionary algorithms and drift analysis

2.1. Evolutionary algorithms

The combinatorial optimization problem considered in this paper can be described as follows: Given a finite state space S and a function $f(x)$, $x \in S$, find

$$\max\{f(x); x \in S\}. \quad (1)$$

Assume x^* is one state with the maximum function value, and $f_{\max} = f(x^*)$.

The EA for solving the combinatorial optimization problem can be described as follows:

- (1) *Initialization*: generate, either randomly or heuristically, an initial population of $2N$ individuals, denoted by $\xi_0 = (x_1, \dots, x_{2N})$, and let $k \leftarrow 0$, where $N > 0$ is an integer. For any population ξ_k , define $f(\xi_k) = \max\{f(x_i); x_i \in \xi_k\}$.
- (2) *Generation*: generate a new (intermediate) population by crossover and mutation (or any other operators for generating offspring), and denote it as $\xi_{k+1/2}$.
- (3) *Selection*: select and reproduce $2N$ individuals from populations $\xi_{k+1/2}$ and ξ_k , and obtain another (new intermediate) population ξ_{k+S} .
- (4) If $f(\xi_{k+S}) = f_{\max}$, then stop; otherwise let $\xi_{k+1} = \xi_{k+S}$ and $k \leftarrow k + 1$, and go to step (2).

Obviously the above description includes a wide range of EAs using crossover, mutation and selection. The description does not set any restrictions on the type of crossover, mutation or selection schemes used. It includes EAs which use crossover or mutation alone. The EA framework given above is closer to evolution strategies [19] and evolutionary programming [4] than to GAs [6] in the sense that selection is applied after crossover and/or mutation. However, the main results given in this paper, i.e., Theorems 1 and 10 are independent of any such implementation details. In fact, they hold for virtually any stochastic search algorithms.

2.2. Drift analysis

Assume x^* is an optimal point, and let $d(x, x^*)$ be the distance between a point x and x^* . If there are more than one optimal point (that is, a set S^*), we use $d(x, S^*) = \min\{d(x, x^*); x^* \in S^*\}$ as the distance between individual x and the optimal set S^* . In short we denote the distance by $d(x)$. Usually $d(x)$ satisfies $d(x^*) = 0$ and $d(x) > 0$ for any $x \notin S^*$. However, in some parts of this paper, we will consider a pseudo-distance $d(x)$ which allows $d(x) = 0$ for some $x \notin S^*$.

Given a population $X = \{x_1, \dots, x_{2N}\}$, let

$$d(X) = \min\{d(x); x \in X\}, \quad (2)$$

which is used to measure the distance of the population to the optimal solution.

The sequence $\{d(\xi_k); k = 0, 1, 2, \dots\}$ generated by the EA is a random sequence. The sequence can be modeled by a homogeneous Markov chain if no self-adaptation is used [10].

The drift of the random sequence $\{d(\xi_k); k = 0, 1, \dots\}$ at time k is defined by

$$\Delta(d(\xi_k)) = d(\xi_{k+1}) - d(\xi_k).$$

Define the stopping time of an EA as $\tau = \min\{k: d(\xi_k) = 0\}$, which is the first hitting time on the optimal solution. The task now is to investigate the relationship between the expect first hitting time τ and the problem size n . In this paper, we focus on the following question: under what conditions of the drift $\Delta(d(\xi_k))$ can we estimate the expect first hitting time $E[\tau]$? In particular, we study the conditions under which an EA is guaranteed to find the optimal solution in polynomial time on average and conditions under which an EA takes at least exponential time on average to find the optimal solution.

The idea behind drift analysis is quite straightforward. It can be explained (by sacrificing mathematical rigor) using a deterministic algorithm as an example. Assume the distance between the starting solution and the optimal solution is d , and a deterministic algorithm is used to solve an optimization problem. If the drift towards the optimal solution is greater than Δ at each time step (i.e., iteration), we would need at most d/Δ time steps to find the optimal solution. Hence the key issue here is to estimate Δ and d . Sasaki and Hajek [20] have successfully used this method to estimate the time complexity of simulated annealing for the maximum matching problem.

3. Conditions for polynomial average computation time

3.1. Drift conditions

In this section, we study under which drift conditions an EA can solve an optimization problem in polynomial average time.

Condition 1. There exists a polynomial of problem size n , $h_0(n) > 0$, such that

$$d(X) \leq h_0(n)$$

for any given population X .

This condition says that the distance from any population to the optimal solution is bounded by a polynomial function of the problem size.

Condition 2. At any time $k \geq 0$, if population ξ_k satisfies $d(\xi_k) > 0$, then there exists a polynomial of problem size n , $h_1(n) > 0$, such that

$$E[d(\xi_k) - d(\xi_{k+1}) \mid d(\xi_k) > 0] \geq \frac{1}{h_1(n)}.$$

This condition indicates that the drift of the random sequence $\{d(\xi_k); k = 0, 1, 2, \dots\}$ toward the optimal solution is always positive and bounded by an inverse polynomial.

Now we give the following main result in the section.

Theorem 1. *If $\{d(\xi_k); k = 0, 1, 2, \dots\}$ satisfies Conditions 1 and 2, then starting from any initial population X with $d(X) > 0$,*

$$E[\tau \mid d(\xi_0) > 0] \leq h(n),$$

where $h(n)$ is a polynomial of problem size n .

Proof. According to Condition 2, we know that $\{d(\xi_k); k = 0, 1, 2, \dots\}$ in fact is a supermartingale [15]. Since $0 \leq d(\xi_k) \leq h_0(n)$, it converges almost everywhere [15], and

$$\lim_{k \rightarrow \infty} E[d(\xi_k) \mid d(\xi_0) > 0] = 0.$$

According to the definition of stopping time τ , we have $d(\xi_\tau) = 0$. Hence,

$$E[d(\xi_\tau) \mid d(\xi_0) > 0] = 0.$$

For any time $k \geq 1$,

$$E[d(\xi_k) \mid d(\xi_0) > 0] = E[E[d(\xi_{k-1}) + \Delta(d(\xi_{k-1})) \mid \xi_{k-1}] \mid d(\xi_0) > 0].$$

According to Condition 2, we have for $k - 1 < \tau$,

$$E[d(\xi_{k-1}) + \Delta(d(\xi_{k-1})) \mid \xi_{k-1}] \leq d(\xi_{k-1}) - \frac{1}{h_1(n)}.$$

Therefore

$$E[d(\xi_k) \mid d(\xi_0) > 0] \leq E\left[d(\xi_{k-1}) - \frac{1}{h_1(n)} \mid d(\xi_0) > 0\right].$$

By induction on k , we can get

$$E[d(\xi_k) \mid d(\xi_0) > 0] \leq E\left[d(\xi_0) - \frac{k}{h_1(n)} \mid d(\xi_0) > 0\right].$$

Hence we have

$$\begin{aligned} 0 = E[d(\xi_\tau) \mid d(\xi_0) > 0] &\leq E\left[d(\xi_0) - \frac{\tau}{h_1(n)} \mid d(\xi_0) > 0\right] \\ &\leq E[d(\xi_0)] - \frac{1}{h_1(n)} E[\tau \mid d(\xi_0) > 0]. \end{aligned}$$

According to the above inequality and Condition 1,

$$E[\tau \mid d(\xi_0) > 0] \leq E[d(\xi_0)]h_1(n) \leq h_0(n)h_1(n).$$

Let $h(n) = h_0(n)h_1(n)$. We arrive at

$$E[\tau \mid d(\xi_0) > 0] \leq h(n). \quad \square$$

Under certain stronger conditions, we can get some stronger results.

Condition 3. Let $d_{\max} = \{d(x): x \in S\}$, and the interval $[0, d_{\max}]$ be divided into $L + 1$ sub-intervals: $d_0 \equiv 0 < d_1 < \dots < d_L < d_{L+1} \equiv d_{\max}$, where $L > 0$ is an integer.

- (a) For any l ($0 \leq l \leq L$), if at time k , the population ξ_k enters the interval $[0, d_l]$, i.e., $d(\xi_k) \leq d_l$, then after that time, the population will not return to the interval $(d_l, d_{L+1}]$ again, i.e., for any $t \geq k$: $d(\xi_t) \leq d_l$;
- (b) At any time k , if the population ξ_k is in the interval $(d_l, d_{l+1}]$, then the drift satisfies:

$$E[d(\xi_k) - d(\xi_{k+1}) \mid d_l < d(\xi_k) \leq d_{l+1}] \geq \frac{1}{h_l(n)},$$

where $h_l(n) > 0$.

Theorem 2. If $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies Condition 3, then starting from any initial population ξ_0 with $d(\xi_0) > 0$,

$$E[\tau \mid d(\xi_0) > 0] \leq \sum_{l=0}^L h_l(n)(d_{l+1} - d_l).$$

Proof. Let's consider the worst case with the initial population $d(\xi_0) = d_{L+1}$.

For any l with $0 \leq l \leq L$, define $\tau_{L+1} = 0$ and

$$\tau_l = \min\{t: d(\xi_t) \leq d_l\}.$$

It is easy to see that $\tau = \tau_0 = (\tau_L - \tau_{L+1}) + (\tau_{L-1} - \tau_L) + \dots + (\tau_0 - \tau_1)$.

Given any l with $0 \leq l \leq L$, according to Condition 3(b), we know

$$E[d(\xi_k) - d(\xi_{k+1}) \mid d_l < d(\xi_k) \leq d_{l+1}] \geq \frac{1}{h_l(n)}.$$

Then according to Condition 3(a) and Theorem 1, we have

$$E[\tau_l - \tau_{l+1} \mid d_l < d(\xi_k) \leq d_{l+1}] \leq h_l(n)(d_{l+1} - d_l).$$

Hence

$$E[\tau \mid \xi_0] \leq \sum_{l=0}^L h_l(n)(d_{l+1} - d_l). \quad \square$$

Condition 4. Let $d_{\max} = \{d(x): x \in S\}$, and the interval $[0, d_{\max}]$ be divided into $L + 1$ sub-intervals: $d_0 \equiv 0 < d_1 < \dots < d_L < d_{L+1} \equiv d_{\max}$, where $L > 0$ is an integer. At any time $k \geq 0$, if the population ξ_k is in the interval $(d_l, d_{\max}]$, then the drift satisfies:

$$E[d(\xi_k) - d(\xi_{k+1}) \mid d(\xi_k) \geq d_l] \geq \frac{1}{h_l(n)},$$

where $h_l(n) > 0$.

Theorem 3. If $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies Condition 4, then starting from any initial population ξ_0 with $d(\xi_0) > 0$,

$$E[\tau \mid d(\xi_0) > 0] \leq \sum_{l=0}^L h_l(n)(d_{l+1} - d_l).$$

Proof. The proof is similar to that of Theorem 2. □

Using the same analytical technique as those used in Theorem 1, we can obtain easily the following results.

Condition 5. For some population X ,

$$d(X) \geq h_0(n),$$

where $h_0(n) > 0$ is a function of problem size n .

Condition 6. There exists a polynomial function, $h_1(n) > 0$, of problem size n such that

$$E[d(\xi_k) - d(\xi_{k+1}) \mid \xi_k = X] \leq \frac{1}{h_1(n)}$$

for any time k and population X with $d(X) > 0$.

Theorem 4. *If $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies Conditions 5 and 6, then starting from the initial population with $d(X) \geq h_0(n)$,*

$$E[\tau \mid \xi_0 = X] \geq h(n),$$

where $h(n) = h_0(n)h_1(n)$ is a function of problem size n .

Proof. Similar to the proof of Theorem 1. □

3.2. The subset sum problem

EAs have been applied to the subset sum problem in practice [13]. The problem can be described as follows: Given a set $W_n = \{w_1, \dots, w_n\}$ of n integers and a large integer C , find a subset S of W such that the sum of the elements in S are closest to but not exceeding C . The subset sum problem is NP-complete. The partition problem can be polynomially transformed to it [12].

A solution S to the subset sum problem can be represented by a string $x = (s_1 \cdots s_n)$ where $s_i \in \{0, 1\}$. The presence of w_i in S means that $s_i = 1$ while its absence is represented by $s_i = 0$. A feasible solution to the subset sum problem is a string $x = (s_1 \cdots s_n)$, $s_i \in \{0, 1\}$, such that

$$\sum_{i=1}^n w_i s_i \leq C,$$

where $F(x) = \sum_{j=1}^n w_j s_j$ is called the objective function. The optimal solution is the string that maximizes the objective function (without exceeding C).

The fitness function can be defined as $f(x) = -(C - F(x) + (1 - \theta)F(x))$, where $\theta = 1$ when x is feasible and $\theta = 0$ when x otherwise. Notice that $(0 \cdots 0)$ is a feasible solution.

In this subsection, we are interesting in a particular family of subset sum problems $\{W_n, n = 1, 2, \dots\}$, n is an integer, for which an EA can find the optimal solution within polynomial average time.

The family of problems we focus on is $\{W_1, W_2, \dots, W_n, \dots\}$, where

$$W_n = \{w_1, \dots, w_n\}, \quad (3)$$

$$w_1, w_2, \dots, w_n \text{ are positives}, \quad (4)$$

$$C = \sum_{i=1}^n w_i. \quad (5)$$

It is obvious that $x = (1 \cdots 1)$ is the unique optimal solution and any subset of W_n is a feasible solution. This problem is, in fact, the linear function problem [2].

The EA for solving the family of subset sum problems follows the structure given in Section 2. The crossover, mutation and selection are implemented as follows.

One-point crossover is used. Given two individuals $x = (s_1^{(x)} \cdots s_n^{(x)})$ and $y = (s_1^{(y)} \cdots s_n^{(y)})$ from the population ξ_k , choose a crossover point $m \in \{1, \dots, n-1\}$ at random and exchange all bits after the m th bit between two individuals to form two new individuals x' and y' :

$$x' = (s_1^{(x)} \cdots s_{m-1}^{(x)} s_m^{(y)} s_{m+1}^{(y)} \cdots s_n^{(y)}),$$

$$y' = (s_1^{(y)} \cdots s_{m-1}^{(y)} s_m^{(x)} s_{m+1}^{(x)} \cdots s_n^{(x)}).$$

A new intermediate population ξ_{k+C} of $2N$ individuals will be formed after crossover.

The mutation operator is the bit mutation. Given an individual $x = (s_1 \cdots s_n)$ in ξ_{k+C} , choose a single bit s_i at random from it and flip the bit. Another new intermediate population ξ_{k+M} of $2N$ individuals is formed after mutation.

Selection used implements a kind of probabilistic elitism. $2N$ individuals are selected from ξ_k and ξ_{k+M} as follows: the best individual with the highest fitness is copied with probability at least $1 - e^{-n}$ to the new population ξ_{k+S} , and other individuals are assigned a survival probability according to their fitness. Any selection scheme can be used as long as fitter individuals were assigned higher probabilities.

Theorem 5. *Given the family of subset sum problems and the EA to solve them. For any initial population X with $d(X) > 0$,*

$$E[\tau \mid \xi_0 = X] \leq h(n)$$

where $h(n)$ is a polynomial of n .

Proof. Define the distance function $d(x)$ as:

$$d(x) = \sum_{i=1}^n |s_i - 1|.$$

According to Theorem 1, we need to verify that the random sequence, $\{d(\xi_k); k = 0, 1, \dots\}$, satisfies Conditions 1 and 2.

From the definition of the above distance and Eq. (2), we know that for any population X :

$$d(X) \leq n.$$

Hence the random sequence $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies Condition 1.

For any time $k \geq 0$, and population ξ_k with $d(\xi_k) > 0$, we now investigate the impact of crossover on the drift. One of the three events may happen after crossover:

- (1) event $I\{d(\xi_{k+C}) < d(\xi_k)\}$,
- (2) event $I\{d(\xi_{k+C}) = d(\xi_k)\}$, or
- (3) event $I\{d(\xi_{k+C}) > d(\xi_k)\}$.

We first show that event $I\{d(\xi_{k+C}) > d(\xi_k)\}$ cannot happen. In other words, crossover does not produce a worse intermediate population. Assume x_1 and x_2 are two individuals in population ξ_k , and y_1 and y_2 are their offspring. Since the crossover does not increase or decrease the amount of ones in individuals x_1 and x_2 , we have

$$d(y_1) + d(y_2) = d(x_1) + d(x_2).$$

That is,

$$d(y_1) - d(x_1) = -(d(y_2) - d(x_2)).$$

This means that the increase of one individual's drift will be the decrease of another individual's drift. So the crossover will not make the intermediate population ξ_{k+C} worse. Event $I\{d(\xi_{k+C}) > d(\xi_k)\}$ cannot happen.

Assume that event $I\{d(\xi_{k+C}) = d(\xi_k)\}$ has happened. Then one of the following three events may happen subsequently:

- (a) event $I\{d(\xi_{k+M}) < d(\xi_{k+C})\}$,
- (b) event $I\{d(\xi_{k+M}) = d(\xi_{k+C})\}$, and
- (c) event $I\{d(\xi_{k+M}) > d(\xi_{k+C})\}$.

Event $I\{d(\xi_{k+M}) = d(\xi_{k+C})\}$ cannot happen because the mutation always happens.

The probability of event $I\{d(\xi_{k+M}) < d(\xi_{k+C})\}$ is not less than $1/n$ (if $d(\xi_{k+C}) > 0$), and the probability of event $I\{d(\xi_{k+M}) > d(\xi_{k+C})\}$ is not greater than $(n - 1)/n$. If $d(\xi_{k+C}) = 0$, then the population ξ_{k+C} has one individual with the maximum fitness.

Assume that event $I\{d(\xi_{k+C}) < d(\xi_k)\}$ has happened. Then one of the following three events may happen subsequently:

- (a') event $I\{d(\xi_{k+M}) < d(\xi_{k+C})\}$,
- (b') event $I\{d(\xi_{k+M}) = d(\xi_{k+C})\}$, and
- (c') event $I\{d(\xi_{k+M}) > d(\xi_{k+C})\}$.

The probabilities of the three events are similar to those analysed in the case of $I\{d(\xi_{k+C}) = d(\xi_k)\}$.

Now let's examine the role of selection: the individual with the best fitness will appear in the next population ξ_{k+S} with probability $1 - e^{-n}$, so the probability $P(d(\xi_{k+S}) < d(\xi_k) \mid d(\xi_{k+M}) < d(\xi_k))$ is not less than $1 - e^{-n}$ and the probability $P(d(\xi_{k+S}) > d(\xi_k) \mid d(\xi_{k+M}) < d(\xi_k))$ is not more than e^{-n} . The probability $P(d(\xi_{k+S}) > d(\xi_k) \mid d(\xi_{k+M}) > d(\xi_k))$ is not more than e^{-n} . And the event $I\{d(\xi_{k+S}) < d(\xi_k) \mid d(\xi_{k+M}) > d(\xi_k)\}$ cannot happen.

Considering all the different cases discussed above, we have

$$\begin{aligned} & E[d(\xi_{k+1}) - d(\xi_k) \mid d(\xi_k) > 0] \\ &= E[(d(\xi_{k+1}) - d(\xi_k)) I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}), \\ & \quad d(\xi_{k+S}) < d(\xi_k)\} \mid d(\xi_k) > 0] \end{aligned}$$

$$\begin{aligned}
& + E[(d(\xi_{k+1}) - d(\xi_k))I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}), \\
& \quad d(\xi_{k+S}) > d(\xi_k)\} | d(\xi_k) > 0] \\
& + E[(d(\xi_{k+1}) - d(\xi_k))I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}), \\
& \quad d(\xi_{k+S}) < d(\xi_k)\} | d(\xi_k) > 0] \\
& + E[(d(\xi_{k+1}) - d(\xi_k))I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}), \\
& \quad d(\xi_{k+S}) > d(\xi_k)\} | d(\xi_k) > 0] \\
& + E[(d(\xi_{k+1}) - d(\xi_k))I\{d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}), \\
& \quad d(\xi_{k+S}) < d(\xi_k)\} | d(\xi_k) > 0] \\
& + E[(d(\xi_{k+1}) - d(\xi_k))I\{d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}), \\
& \quad d(\xi_{k+S}) > d(\xi_k)\} | d(\xi_k) > 0] \\
& + E[(d(\xi_{k+1}) - d(\xi_k))I\{d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}), \\
& \quad d(\xi_{k+S}) > d(\xi_k)\} | d(\xi_k) > 0].
\end{aligned}$$

In other words,

$$\begin{aligned}
& E[d(\xi_{k+1}) - d(\xi_k) | d(\xi_k) > 0] \\
& \leq (-1)P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) | d(\xi_k) > 0)(1 - e^{-n}) \\
& \quad + (n - 1)P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) | d(\xi_k) > 0)e^{-n} \\
& \quad + (-1)P(d(\xi_{k+C}) < d(\xi_k), d(\xi_k) > d(\xi_{k+M}) > d(\xi_{k+C}) | d(\xi_k) > 0)(1 - e^{-n}) \\
& \quad + (n - 1)P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}) | d(\xi_k) > 0)e^{-n} \\
& \quad + (-1)P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) | d(\xi_k) > 0)(1 - e^{-n}) \\
& \quad + (n - 1)P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) | d(\xi_k) > 0)e^{-n} \\
& \quad + (n - 1)P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}) | d(\xi_k) > 0)e^{-n}.
\end{aligned}$$

In arriving at the above inequality, we have used the fact that $|d(\xi_{k+1}) - d(\xi_k)| \leq n - 1$.

Since

$$(-1)P(d(\xi_{k+C}) < d(\xi_k), d(\xi_k) > d(\xi_{k+M}) > d(\xi_{k+C}) | d(\xi_k) > 0)(1 - e^{-n}) < 0$$

and

$$P(d(\xi_{k+C}) < d(\xi_k) | d(\xi_k) > 0) + P(d(\xi_{k+C}) = d(\xi_k) | d(\xi_k) > 0) = 1,$$

we have

$$\begin{aligned}
& E[d(\xi_{k+1}) - d(\xi_k) | d(\xi_k) > 0] \\
& \leq (-1)P(d(\xi_{k+C}) < d(\xi_k) | d(\xi_k) > 0)\frac{1}{n}(1 - e^{-n}) \\
& \quad + (n - 1)P(d(\xi_{k+C}) < d(\xi_k) | d(\xi_k) > 0)\frac{n - 1}{n}e^{-n} \\
& \quad + (n - 1)P(d(\xi_{k+C}) < d(\xi_k) | d(\xi_k) > 0)\frac{n - 1}{n}e^{-n}
\end{aligned}$$

$$\begin{aligned}
 &+ (-1)P(d(\xi_{k+C}) = d(\xi_k) \mid d(\xi_k) > 0) \frac{1}{n} (1 - e^{-n}) \\
 &+ (n - 1)P(d(\xi_{k+C}) = d(\xi_k) \mid d(\xi_k) > 0) \frac{n - 1}{n} e^{-n} \\
 &+ (n - 1)P(d(\xi_{k+C}) = d(\xi_k) \mid d(\xi_k) > 0) \frac{n - 1}{n} e^{-n} \\
 \leq &(-1) \frac{1}{n} (1 - e^{-n}) + 2(n - 1) \frac{n - 1}{n} e^{-n} \\
 \leq &-\frac{1 - e^{-n} - 2(n - 1)^2 e^{-n}}{n}.
 \end{aligned}$$

Let

$$h_1(n) = \frac{n}{1 - e^{-n} - 2(n - 1)^2 e^{-n}},$$

then when $n \rightarrow +\infty$, $h_1(n) = O(n)$. Hence,

$$E[d(\xi_{k+1}) - d(\xi_k) \mid d(\xi_k) > 0] \leq -\frac{1}{h_1(n)}$$

and

$$\lim_{n \rightarrow \infty} \frac{-1}{h_1(n)} < 0,$$

where

$$h_1(n) = O(n).$$

So we have proven that the random sequence $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies Condition 2. According to Theorem 1, we know

$$E[\tau \mid \xi_0 = X] \leq h(n)$$

where $h(n) = O(n^2)$. \square

3.3. Other problems

In order to show the power and generality of our main results, we will use some of the problems given in [18] as examples to derive EA's computation time by verifying drift conditions given previously. Rudolph's survey [18] is probably the most comprehensive overview of recent results on the finite time behavior of EAs in finite space and discrete time. It is worth noting that the EA used in this section is more general than the (1+1) EA used in [18]. A $(2N + 2N)$ EA without crossover is used in this section.

The results shown in this section illustrate that drift analysis can be used to derive EA's average computation time for a variety of different problems.

Let $S = \{(s_1 \cdots s_n), s_i \in \{0, 1\}\}$ be the chromosome representation. The mutation and selection are implemented as follows.

A kind of uniform bit mutation is used. For any individual $x = (s_1 \cdots s_n)$ in ξ_k , each bit s_i will flip with a mutation rate $p_m > 0$. A new intermediate population ξ_{k+M} of $2N$ individuals is formed after mutation.

$(2N + 2N)$ elitism is implemented as the selection scheme. In other words, the $2N$ individuals with the highest fitness from populations ξ_k and ξ_{k+M} are copied to the new population ξ_{k+S} .

3.3.1. Linear functions

A function $f : S \rightarrow \mathbb{R}$ is linear if $f(x) = c_0 + \sum_{i=1}^n c_i s_i$ where coefficients $c_i \in \mathbb{R}$ [18]. If $c_i \geq 0$ for all c_i , then it is clear that $(1 \dots 1)$ is the only optimal solution.

Theorem 6. For the linear function with positive coefficients $c_1 > c_2 > \dots > c_n > 0$, the EA with mutation probability $p_m = 1/n$ needs an average $O(n \ln n)$ steps to reach the optimal solution.

Proof. Define the distance function $d(x) = \sum_{i=1}^n (1 - s_i)$. Since $0 \leq d(x) \leq n$, we can divide $[0, d_{\max}]$ into n intervals $d_0 < d_1 < d_2 < \dots < d_n$ where $d_l = l$ for $0 \leq l \leq n$. We will use Theorem 3 to prove the result.

Assume at time $k \geq 0$, the population ξ_k satisfy $d(\xi_k) > d_{l-1}$ where $l \in \{1, \dots, n\}$. Without the loss of generality, assume $d(\xi_k) = d_l$ (other cases can be proven in the same way). Then

$$E[d(\xi_k) - d(\xi_{k+1})] = E[(d(\xi_k) - d(\xi_{k+1}))I\{d(\xi_k) > d(\xi_{k+1})\}] + E[(d(\xi_k) - d(\xi_{k+1}))I\{d(\xi_k) < d(\xi_{k+1})\}].$$

First let's consider $E[(d(\xi_k) - d(\xi_{k+1}))I\{d(\xi_k) > d(\xi_{k+1})\}]$. Let x be the best individual in population ξ_k . The probability of its flipping one of its l "0" bits while keeping its $n - l$ "1" bits unchanged is $C_l \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-l}$. Hence

$$E[(d(\xi_k) - d(\xi_{k+1}))I\{d(\xi_k) > d(\xi_{k+1})\}] \geq \frac{l}{n} \left(1 - \frac{1}{n}\right)^{n-l}.$$

Secondly let's consider $E[(d(\xi_k) - d(\xi_{k+1}))I\{d(\xi_k) < d(\xi_{k+1})\}]$. Let x be the best individual in ξ_k . Assume that event $I\{d(\xi_k) < d(\xi_{k+1})\}$ happens, then it implies event I' : i.e., one of its l "0" bits must flip (In fact the leftmost bit among all flipping bits must flip from "0" to "1" because of $c_1 > c_2 > \dots > c_n$ and elitist selection), and at least two of its $n - l$ "1" bits must also flip. So the probability of event $I\{d(\xi_k) < d(\xi_{k+1})\}$ happening is no more than that of event I' . Event I' can be further divided into the following cases:

- (1) One of the "0" bits in x becomes 1, and two of the "1" bits become 0. The probability of this happening is

$$\frac{l}{n} C_{n-l}^2 \left(\frac{1}{n}\right)^2 \left(1 - \frac{1}{n}\right)^{n-l-2} < \frac{l}{n} \frac{1}{2!} \left(1 - \frac{1}{n}\right)^{n-l}.$$

- (2) One of the "0" bits in x becomes 1, and three of the "1" bits become 0. The probability of this happening is

$$\frac{l}{n} C_{n-l}^3 \left(\frac{1}{n}\right)^3 \left(1 - \frac{1}{n}\right)^{n-l-3} < \frac{l}{n} \frac{1}{3!} \left(1 - \frac{1}{n}\right)^{n-l}.$$

(3) ...

Hence we get

$$E[(d(\xi_k) - d(\xi_{k+1}))I\{d(\xi_k) < d(\xi_{k+1})\}] > \frac{l}{n} \left(\frac{-1}{2!} + \frac{-2}{3!} + \dots \right) \left(1 - \frac{1}{n} \right)^{n-l}.$$

So

$$E[d(\xi_k) - d(\xi_{k+1})] \geq \frac{l}{n} \left(1 + \frac{-1}{2!} + \frac{-2}{3!} + \dots \right) \left(1 - \frac{1}{n} \right)^{n-l} \geq c \frac{l}{n},$$

where $c > 0$ is a constant. In other words, Condition 4 holds.

According to Theorem 3,

$$E[\tau] \leq c \sum_{l=n}^1 \frac{n}{l} = O(n \ln n). \quad \square$$

3.3.2. Pseudo-modular functions

A function $f : S \rightarrow \mathbb{R}$ is *pseudo-modular* if

$$\begin{aligned} \min\{f(x), f(y)\} &\leq \max\{f(x \wedge y), f(x \vee y)\}, \quad \text{and} \\ \max\{f(x), f(y)\} &\leq \min\{f(x \wedge y), f(x \vee y)\}, \end{aligned}$$

for all $x, y \in S$ [18].

An example of the pseudo-modular function is the function

$$f(x) = \sum_{i=1}^n \prod_{j=1}^i s_j. \tag{6}$$

Theorem 7. *The expected first-hitting time of the EA for the fitness function (6) is $E[\tau] \leq n^2(e - 1)$ when mutation rate $p_m = 1/n$.*

Proof. For fitness function (6), the optimal solution is $(1 \dots 1)$. Define the distance function $d(x)$ as follows:

$$d(x) = n - \sum_{i=1}^n \prod_{j=1}^i s_j.$$

We can divide $[0, n]$ into n subintervals $d_0 < d_1 < \dots < d_n$ where $d_l = l$ for $0 \leq l \leq n$.

According to Theorem 2, we need to verify that $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies Condition 3.

First let's verify Condition 3(a). Since the EA adopts an elitist selection strategy, Condition 3(a) holds automatically.

Second let's verify Condition 3(b). At any time $k \geq 0$, if population ξ_k is in the interval $(d_l, d_{l+1}]$ where $0 \leq l \leq n - 1$, then there exists at least one individual x in ξ_k such that $d(x) = l + 1$. The probability of x becoming better is no less than $\frac{1}{n} \left(1 - \frac{1}{n} \right)^{n-l-1}$. Hence,

$$E[d(\xi_k) - d(\xi_{k+1})] \geq \frac{1}{n} \left(1 - \frac{1}{n} \right)^{n-l-1}.$$

That is, Condition 3(b) holds.

According to Theorem 2, we have

$$E[\tau \mid \xi_0] \leq \sum_{l=0}^{n-1} n \left(1 - \frac{1}{n}\right)^{-n+l+1} \leq n^2(e - 1). \quad \square$$

3.3.3. Unimax functions

A function $f : S \rightarrow \mathbb{R}$ is *unimax* if there is a unique locally maximal point $x^* \in S$ [18]. The long path problem is a well known unimax problem.

A long path P_n (where the length n of string is odd) is defined by a recursion with $P_1 = \{0, 1\}$ as the base path [16]. Given long path P_n , create a subpath S_{00} by prepending “00” to each point in P_n and another subpath S_{11} by prepending “11” to each point in the reverse order in P_n . The bridge point is build from the last point in P_n prepending by “01”. Finally, concatenate subpath S_{00} , the bridge point and subpath S_{11} to obtain long path P_{n+2} . The length of the paths is described by the recurrence equations

$$|P_1| = 2, \quad |P_{n+2}| = 2|P_n| + 1,$$

whose solution is $|P_n| = 3 \cdot 2^{(n-1)/2} - 1$ for odd $n \geq 1$. Table 1 shows long path P_5 .

Given a point x on a path P_n , define $Pos(x)$ to be the position of x on the path which is numbered from 0 to $3 \cdot 2^{(n-1)/2} - 2$. For a point not in the path P_n , define $Pos(x)$ to be -1 . Then define the objective function $f(x)$ as:

$$f(x) = -\left(3 \cdot 2^{(n-1)/2} - 2\right) + \begin{cases} Pos(x), & \text{if } Pos(x) \geq 0, \\ -\sum_{i=1}^n s_i, & \text{otherwise.} \end{cases} \quad (7)$$

Theorem 8. For the unimax function (7), starting from the bottom of the increasing path, the expected first hitting time of the EA is $E[\tau] = O(n^3)$ when mutation rate $p_m = 1/n$.

Proof. Decompose space S into a family of sets $\{S_{-1}, S_0, \dots, S_{(n-1)/2}\}$ as follows [16]:

$$\begin{aligned} S_0 &= \{(00 * \dots *) \in P_n\} \cup \{(01 * \dots *) \in P_n\}, \\ S_1 &= \{(1111 * \dots *) \in P_n\} \cup \{(1101 * \dots *) \in P_n\}, \\ S_2 &= \{(110011 * \dots *) \in P_n\} \cup \{(110001 * \dots *) \in P_n\}, \end{aligned}$$

Table 1
Long path P_5

| $Pos(x)$ | x | $Pos(x)$ | x |
|----------|-------|----------|-------|
| 0 | 00000 | 6 | 11110 |
| 1 | 00001 | 7 | 11111 |
| 2 | 00011 | 8 | 11011 |
| 3 | 00111 | 9 | 11001 |
| 4 | 00110 | 10 | 11000 |
| 5 | 10110 | | |

...

$$S_{(l-3)/2} = \{(1100 \cdots 0011*) \in P_n\} \cup \{(1100 * * \cdots 0001*) \in P_n\},$$

$$S_{(l-1)/2} = \{(1100 \cdots 0000*) \in P_n\},$$

and S_{-1} includes all remained points.

Define the distance function $d(x)$ as follows:

$$d(x) = 0, \quad \forall x \in S_{(n-1)/2},$$

$$d(x) = 1, \quad \forall x \in S_{(n-3)/2},$$

...

$$d(x) = (n - 3)/2, \quad \forall x \in S_1,$$

$$d(x) = (n - 1)/2, \quad \forall x \in S_0,$$

$$d(x) = (n - 1)/2 + \min \left\{ \sum_{i=1}^n |s_i^{(x)} - s_i^{(y)}|; y \in P_n \right\}, \quad \forall x \in S_{-1}.$$

In the following, we will prove that $\{d(\xi_k), k = 0, 1, \dots\}$ satisfies Condition 3.

Let $d_l = l$, we can divide $[0, d_{\max}]$ into a finite number of subintervals $d_0 < d_1 < \dots < d_{\max}$.

First, Condition 3(a) holds because the EA adopts an elitism selection strategy.

Secondly, let's verify Condition 3(b). At any time $k \geq 0$, if population ξ_k satisfies $d(\xi_k) > (n - 1)/2$, then no individual x in the population is on path P_n . Let x is the best individual in ξ_k , then the probability of x having a drift is at least $(1 - 1/n)^{n-1}/n$. The drift is at least 1. Hence,

$$E[d(\xi_{k+1}) - d(\xi_k) \mid d(\xi_k) > (n - 1)/2] = \Omega(n^{-1}).$$

Now let's estimate the drift along the path. At any time $k \geq 0$, if population ξ_k is in the interval $(d_l, d_{l+1}]$ where $0 \leq l < (n - 1)/2$, then at least one individual x satisfies $d(x) = d_{l+1}$. If x is a bridge point, the probability of a drift happening is at least $(1 - 1/n)^{n-1}/n$ and the drift length is at least 1. If x is not a bridge point, the probability of a drift happening is at least $(1 - 1/n)^{n-2}/n^2$ and the drift length is at least 1. Summarising both cases, the expected drift on the path is:

$$E[d(\xi_{k+1}) - d(\xi_k) \mid d(\xi_k) = l + 1] = \Omega(n^{-2}).$$

In other words, Condition 2(b) holds.

According to Theorem 2, we come to the conclusion

$$E[\tau] \leq \sum_{l=n}^0 O(n) + \sum_{l=(n-1)/2}^0 O(n^2) = O(n^3),$$

where the first part is the average time for points not on the path to reach the path, and the second part is the average time for points on the path to reach the optimal solution. \square

3.3.4. Almost positive functions

A function $f : S \rightarrow \mathbb{R}$ is *almost-positive* if the coefficients of all nonlinear terms are non-negative [18].

An example of almost-positive function is

$$f(x) = n - \sum_{i=1}^n s_i + (n+1) \prod_{i=1}^n s_i. \quad (8)$$

We can define the distance function as $d(x) = \sum_{i=1}^n |s_i - 1|$.

Theorem 9. *The expected first-hitting time of the EA for the almost positive function (8) is $E[\tau] = \Omega(n^n)$ when mutation rate $p_m = 1/n$ and the EA starts from $d(\xi_0) = n$.*

Proof. For the fitness function (8), the optimal solution is $(1 \cdots 1)$. Individual $x = (0 \cdots 0)$ is the second best (maximum) point because $f(x) = n$, but it is farthest from the optimal solution with $d(x) = n$.

According to Theorem 4, we need to verify that $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies Conditions 5 and 6.

First, let's assume that $d(\xi_0) = n$, i.e., the initial population is composed of individuals $(0 \cdots 0)$ only. Let $h_0(n) = n$, then Condition 5 holds.

Secondly, at any time $k \geq 0$, if ξ_k satisfies $d(\xi_k) = n$, there are only two possible events which may happen after mutation and elitist selection: event $I\{d(\xi_{k+1}) = n\}$ or event $I\{d(\xi_{k+1}) = 0\}$, because of elitist selection. Hence,

$$E[d(\xi_k) - d(\xi_{k+1}) \mid d(\xi_k) = n] \leq n \left(\frac{1}{n}\right)^n.$$

Let $h_1(n) = n^{n-1}$, then Condition 6 holds.

According to Theorem 4, we have $E[\tau] = \Omega(n^n)$. \square

4. Drift conditions for exponential average computation time

4.1. Drift conditions

In this subsection, we investigate the drift conditions under which EAs will take the average time exponential in the problem size n to find the optimal solution. Our analysis is based on Hajek's earlier work [7].

In order to consider the case where an EA might not be able to find the exact optimal solution, but only an approximate solution, define the stopping time τ of an EA as: $\tau = \min\{k: d(\xi_k) \leq d_b\}$ where $d_b \geq 0$.

Condition 7. For any population X with $d_b < d(X) < d_a$, where $d_b \geq 0$ and $d_a > 0$,

$$E[e^{-(d(\xi_{k+1}) - d(\xi_k))} \mid \xi_k = X, d_b < d(\xi_k) < d_a] \leq \rho < 1, \quad (9)$$

where $\rho > 0$ is a constant.

This condition indicates that (d_b, d_a) is a very difficult interval to search. When the condition is satisfied, $d(\xi_{k+1}) > d(\xi_k)$. In other words, the offspring population is on average drifting away from the optimal solution, rather than getting closer to it.

Condition 8. For any population X with $d(X) \geq d_a, d_a > 0$,

$$E[e^{-d(\xi_{k+1}-d_a)} \mid \xi_k = X, d(\xi_k) \geq d_a] \leq D, \tag{10}$$

where $D \geq 1$ is a constant.

The above condition indicates that a population in the interval $[d_a, +\infty)$ will not, on average, drift towards the optimal solution too much because $d(\xi_{k+1}) \geq d_a - \ln D$.

Given the above two conditions, the following lemma and theorem can be shown by following Hajek’s work on drift analysis [7].

Lemma 1. *If $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies Conditions 7 and 8, then for any initial population ξ_0 ,*

$$E[e^{-d(\xi_k)} \mid \tau > k - 1, d(\xi_0)] \leq \rho^k e^{-d(\xi_0)} + \frac{1 - \rho^k}{1 - \rho} D e^{-d_a}, \tag{11}$$

and

$$P[d(\xi_k) \leq d_b \mid \tau > k - 1, d(\xi_0)] \leq \rho^k e^{-(d(\xi_0)-d_b)} + \frac{1 - \rho^k}{1 - \rho} D e^{-(d_a-d_b)}. \tag{12}$$

Proof. Inequality (11) is clearly true for $k = 0$.

For $k \geq 0$ and $\tau > k$,

$$E[e^{-d(\xi_{k+1})} \mid \tau > k, d(\xi_0)] = E[E[e^{-d(\xi_{k+1})} \mid \tau > k, d(\xi_k)] \mid d(\xi_0)],$$

where

$$\begin{aligned} & E[e^{-d(\xi_{k+1})} \mid \tau > k, d(\xi_k)] \\ &= E[e^{-d(\xi_{k+1})} \mid \tau > k, d(\xi_k) \geq d_a] + E[e^{-d(\xi_{k+1})} \mid \tau > k, d(\xi_k) < d_a]. \end{aligned} \tag{13}$$

The first term on the right-hand side of inequality (13) is upper-bounded by $D e^{-d_a}$ according to Condition 8, and the second term is upper-bounded by $\rho e^{-d(\xi_k)}$ according to Condition 7. Using these bounds we can arrive at

$$E[e^{-d(\xi_{k+1})} \mid \tau > k, d(\xi_0)] \leq \rho E[e^{-d(\xi_k)} \mid \tau > k - 1, d(\xi_0)] + D e^{-d_a}.$$

By induction on k , it is easy to show that the above inequality implies inequality (11) for all $k \geq 0$.

Inequality (12) follows from inequality (11):

$$\begin{aligned} & E[e^{-(d(\xi_k)-d_b)} \mid \tau > k - 1, d(\xi_0)] \\ &= E[E[e^{-(d(\xi_k)-d_b)} \mid \tau > k - 1, d(\xi_k) \leq d_b] \mid d(\xi_0)] \\ &\quad + E[E[e^{-(d(\xi_k)-d_b)} \mid \tau > k - 1, d(\xi_k) > d_b] \mid d(\xi_0)] \\ &\geq E[E[e^{-(d(\xi_k)-d_b)} \mid \tau > k - 1, d(\xi_k) \leq d_b] \mid d(\xi_0)] \\ &\geq e^0 P(d(\xi_k) \leq d_b \mid \tau > k - 1, d(\xi_0)). \end{aligned}$$

That is,

$$P(d(\xi_k) \leq d_b \mid \tau > k - 1, d(\xi_0)) \leq E[e^{-(d(\xi_k)-d_b)} \mid \tau > k - 1, d(\xi_0)].$$

According to inequality (11) we have

$$P(d(\xi_k) \leq d_b \mid d(\xi_0) \geq d_a) \leq \rho^k e^{-(d(\xi_0)-d_b)} + \frac{1-\rho^k}{1-\rho} D e^{-(d_a-d_b)}. \quad \square$$

The following theorem is the main result of this section.

Theorem 10. Assume Conditions 7 and 8 hold. If $d(\xi_0) \geq d_a$, $D \geq 1$ and $\rho < 1$, then there exist some $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$E[\tau \mid d(\xi_0) \geq d_a] \geq \delta_1 e^{\delta_2(d_a-d_b)} \quad (14)$$

Proof. Because $d(\xi_0) \geq d_a$, we have

$$e^{-(d(\xi_0)-d_b)} \leq e^{-(d_a-d_b)}.$$

Since $D \geq 1$ and $\rho < 1$, we can obtain

$$\rho^k e^{-(d(\xi_0)-d_b)} \leq \frac{\rho^k}{1-\rho} D e^{-(d_a-d_b)}.$$

According to inequality (12) and the above inequality,

$$P(d(\xi_k) \leq d_b \mid \tau > k - 1, d(\xi_0)) \leq \frac{1}{1-\rho} D e^{-(d_a-d_b)}.$$

By using the fact that $P(\tau = k \mid d(\xi_0)) = P(d(\xi_k) \leq d_b, \tau > k - 1 \mid d(\xi_0))$, we have

$$P(\tau > k \mid d(\xi_0)) = 1 - \sum_{j=1}^k P(\tau = j \mid d(\xi_0)) \geq \max\left(0, 1 - k \frac{D e^{-(d_a-d_b)}}{1-\rho}\right).$$

Therefore

$$\begin{aligned} E[\tau \mid d(\xi_0)] &= \sum_{j=1}^{+\infty} P(\tau > j \mid d(\xi_0)) \geq \sum_{k=0}^{+\infty} \max\left(0, 1 - k \frac{D e^{-(d_a-d_b)}}{1-\rho}\right) \\ &\geq \frac{1-\rho}{2D} e^{d_a-d_b}. \end{aligned}$$

Let $\delta_1 = \frac{1-\rho}{2D}$ and $\delta_2 = 1$, then

$$E[\tau \mid d(\xi_0) \geq d_a] \geq \delta_1 e^{\delta_2(d_a-d_b)}. \quad \square$$

4.2. The subset sum problem revisited

In this subsection, we consider another family of subset sum problems. We will show that some EAs described in this subsection take at least an exponential time on average to find the optimal solution.

The family of subset sum problems that we focus on in this subsection is $\{W_1, W_2, \dots, W_n, \dots\}$, where

$$W_n = \{w_1, \dots, w_n\},$$

$$w_1, w_2, \dots, w_{n-1} \text{ are positives greater than } 2, w_n = \sum_{i=1}^{n-1} w_i - 1,$$

and

$$C = w_n.$$

It is easy to see that the subset $\{w_n\}$ is the unique optimal solution and any subset of $W_n - \{w_n\}$ is a feasible solution. This is a deceptive problem.

The EA used to solve the above family of problems follows the framework given in section 2. The crossover, mutation and selection are implemented as follows.

One-point crossover is used in the EA. Given two individuals $x = (s_1^{(x)} \dots s_n^{(x)})$ and $y = (s_1^{(y)} \dots s_n^{(y)})$ from population ξ_k , choose a crossover point $m \in \{1, \dots, n - 1\}$ at random and exchange all bits from the m th bit between two individuals to form two new individuals x' and y' :

$$\begin{aligned} x' &= (s_1^{(x)} \dots s_{m-1}^{(x)} s_m^{(y)} s_{m+1}^{(y)} \dots s_n^{(y)}), \\ y' &= (s_1^{(y)} \dots s_{m-1}^{(y)} s_m^{(x)} s_{m+1}^{(x)} \dots s_n^{(x)}). \end{aligned}$$

If an offspring is infeasible, one of the parents will be retained. A new intermediate population ξ_{k+C} of $2N$ individuals will be formed after crossover.

Bit mutation is used in the EA. Given an individual $x = (s_1 \dots s_n)$ in ξ_{k+C} , choose a single bit s_i at random from it and flip the bit. If the offspring is infeasible, the parent will be retained. Another new intermediate population ξ_{k+M} of $2N$ individuals is formed after mutation.

Selection in the EA can be regarded as a simple form of ranking. $2N$ individuals are selected from ξ_k and ξ_{k+M} as follows: the best $2N$ individuals are assigned a survival probability of $(1 - e^{-n})/2N$ each, and the worst $2N$ individuals are assigned a survival probability of $e^{-n}/2N$ each. Yet another new intermediate population ξ_{k+S} of $2N$ individuals is formed after selection. It should be noted that the selection used here is similar to but not the same as that used in Section 3.2. The chromosome representation used is the same as that described in Section 3.2.

Define

$$d(x) = \left| \sum_{i=1}^n s_i - 1 \right| \quad \text{and} \quad d_{\max} = \max\{d(x): x \text{ is a feasible solution}\}.$$

Note that the distance is a pseudo-distance.

Given two individuals x_1 and x_2 with $d(x_1) > d(x_2) > 2$, the fitness of x_1 will be higher than that of x_2 , that is $f(x_1) > f(x_2)$.

Let $\tau = \min\{k: d(\xi_k) = 0\}$, $d_a = d_{\max} = n - 2$, $d_b = 3n/4$, and $\tau' = \min\{k: d(\xi_k) \leq d_b\}$. Obviously $E[\tau] \geq E[\tau']$. The following theorem gives the main result of this section.

Theorem 11. For the random sequence, $\{d(\xi_k); k = 0, 1, \dots\}$, defined by the family of subset sum problems and the EA in this section, if $d(\xi_0) \geq d_a$, then there exist two constants $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$E[\tau \mid d(\xi_0) \geq d_a] \geq \delta_1 \exp(\delta_2 n)$$

for sufficiently large n .

Proof. According to Theorem 10, all we need to show is to verify that Conditions 7 and 8 are satisfied.

First we show that Condition 7 can be satisfied, i.e.,

$$E[e^{-(d(\xi_{k+1})-d(\xi_k))} \mid d_b < d(\xi_k) < d_a] \leq \rho < 1.$$

After the crossover, one of the following three events may happen: event $I\{d(\xi_{k+C}) < d(\xi_k)\}$, event $I\{d(\xi_{k+C}) = d(\xi_k)\}$ or event $I\{d(\xi_{k+C}) > d(\xi_k)\}$.

Let x_1 and x_2 be two individuals in population ξ_k . Since crossover does not increase or decrease the amount of ones in individuals x_1 and x_2 , so event $I\{d(\xi_{k+C}) > d(\xi_k)\}$ cannot happen.

Because $d(\xi_k) > d_b = 3n/4$, x_1 and x_2 cannot be the optimal solution. Both of their n th bits will be 0. The bits will still be 0 after crossover. Let y_1 and y_2 be offspring of x_1 and x_2 , we have $d(y_1) > n/4$ and $d(y_2) > n/4$. Therefore, $d(\xi_{k+C}) > n/4$,

$$|d(\xi_{k+C}) - d(\xi_k)| \leq \frac{3n}{4} - \frac{n}{4} + 1 = \frac{n}{2} + 1. \quad (15)$$

Assume that either event $I\{d(\xi_{k+C}) < d(\xi_k)\}$ or event $I\{d(\xi_{k+C}) = d(\xi_k)\}$ has happened. In either case, one of the following three events may happen after mutation: event $I\{d(\xi_{k+M}) < d(\xi_{k+C})\}$, event $I\{d(\xi_{k+M}) = d(\xi_{k+C})\}$, or event $I\{d(\xi_{k+M}) > d(\xi_{k+C})\}$. Since $d(\xi_{k+C}) > n/4$, $d(\xi_{k+M}) > n/4 - 1 > 2$.

As regard to the impact of selection on the drift, it is easy to see that

$$\begin{aligned} P(d(\xi_{k+S}) > d(\xi_k) \mid 2 < d(\xi_k) < d(\xi_{k+M})) &\geq \left(\frac{1}{2N}(1 - e^{-n})\right)^{2N} \\ &> \left(\frac{1 - e^{-1}}{2N}\right)^{2N}, \end{aligned}$$

and

$$P(d(\xi_{k+S}) < d(\xi_k) \mid d(\xi_k) > d(\xi_{k+M}) > 2) \leq \frac{1}{2N} e^{-n}.$$

Summarizing all the above events, we have

$$\begin{aligned} &E[e^{-(d(\xi_{k+1})-d(\xi_k))} \mid d(\xi_k) > d_b] \\ &= E[e^{-(d(\xi_{k+1})-d(\xi_k))} I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}), \\ &\quad d(\xi_{k+S}) < d(\xi_{k+M})\} \mid d(\xi_k) > d_b] \\ &\quad + E[e^{-(d(\xi_{k+1})-d(\xi_k))} I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}), \\ &\quad d(\xi_{k+S}) = d(\xi_{k+M})\} \mid d(\xi_k) > d_b] \end{aligned}$$

$$\begin{aligned}
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) > d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) < d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) = d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) > d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) < d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) = d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) > d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) < d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) = d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) > d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) < d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) = d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) > d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) < d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) = d(\xi_{k+M})\} | d(\xi_k) > d_b] \\
 &+ E[e^{-d(\xi_{k+1})-d(\xi_k)} I\{d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}), \\
 &\quad d(\xi_{k+S}) > d(\xi_{k+M})\} | d(\xi_k) > d_b].
 \end{aligned}$$

Let

$$\begin{aligned}
 p_1 &= P(d(\xi_{k+S}) > d(\xi_{k+M}) \mid d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}), d(\xi_k) > d_b), \\
 p_2 &= P(d(\xi_{k+S}) > d(\xi_{k+M}) \mid d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}), d(\xi_k) > d_b), \\
 p_3 &= P(d(\xi_{k+S}) > d(\xi_{k+M}) \mid d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}), d(\xi_k) > d_b), \\
 p_4 &= P(d(\xi_{k+S}) > d(\xi_{k+M}) \mid d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}), d(\xi_k) > d_b), \\
 p_5 &= P(d(\xi_{k+S}) > d(\xi_{k+M}) \mid d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}), d(\xi_k) > d_b), \\
 p_6 &= P(d(\xi_{k+S}) > d(\xi_{k+M}) \mid d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}), d(\xi_k) > d_b),
 \end{aligned}$$

where p_1, \dots, p_6 are all greater than $(1 - e^{-1}/(2N))^{2N}$. They represent the probabilities of a population drifting away from the optimal solution after selection.

According to Eq. (15) and the above analysis, we have

$$\begin{aligned}
 &E[e^{-(d(\xi_{k+1})-d(\xi_k))} \mid d(\xi_k) > d_b] \\
 &\leq e^{n/2+1} P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) \mid d(\xi_k) > d_b) O(e^{-n}) \\
 &\quad + e^0 P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) \mid d(\xi_k) > d_b) (1 - p_1) \\
 &\quad + e^{-1} P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) \mid d(\xi_k) > d_b) p_1 \\
 &\quad + e^{n/2+1} P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}) \mid d(\xi_k) > d_b) O(e^{-n}) \\
 &\quad + e^0 P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}) \mid d(\xi_k) > d_b) (1 - p_2) \\
 &\quad + e^{-1} P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}) \mid d(\xi_k) > d_b) p_2 \\
 &\quad + e^{n/2+1} P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}) \mid d(\xi_k) > d_b) O(e^{-n}) \\
 &\quad + e^0 P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}) \mid d(\xi_k) > d_b) (1 - p_3) \\
 &\quad + e^{-1} P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}) \mid d(\xi_k) > d_b) p_3 \\
 &\quad + e^{n/2+1} P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) \mid d(\xi_k) > d_b) O(e^{-n}) \\
 &\quad + e^0 P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) \mid d(\xi_k) > d_b) (1 - p_4) \\
 &\quad + e^{-1} P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) \mid d(\xi_k) > d_b) p_4 \\
 &\quad + e^{n/2+1} P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}) \mid d(\xi_k) > d_b) O(e^{-n}) \\
 &\quad + e^0 P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}) \mid d(\xi_k) > d_b) (1 - p_5) \\
 &\quad + e^{-1} P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}) \mid d(\xi_k) > d_b) p_5 \\
 &\quad + e^{n/2+1} P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}) \mid d(\xi_k) > d_b) O(e^{-n}) \\
 &\quad + e^0 P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}) \mid d(\xi_k) > d_b) (1 - p_6) \\
 &\quad + e^{-1} P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}) \mid d(\xi_k) > d_b) p_6.
 \end{aligned}$$

Let

$$\begin{aligned}
 q_1 &= P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) \mid d(\xi_k) > d_b), \\
 q_2 &= P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}) \mid d(\xi_k) > d_b), \\
 q_3 &= P(d(\xi_{k+C}) < d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}) \mid d(\xi_k) > d_b),
 \end{aligned}$$

$$\begin{aligned} q_4 &= P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) \mid d(\xi_k) > d_b), \\ q_5 &= P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}) \mid d(\xi_k) > d_b), \\ q_6 &= P(d(\xi_{k+C}) = d(\xi_k), d(\xi_{k+M}) > d(\xi_{k+C}) \mid d(\xi_k) > d_b). \end{aligned}$$

Since event $I\{d(\xi_{k+C}) > d(\xi_k)\}$ cannot happen, we have

$$\sum_{i=1}^6 q_i = 1.$$

Now we arrive at

$$\begin{aligned} E[e^{-d(\xi_{k+1})-d(\xi_k)} \mid d(\xi_k) > d_b] &\leq q_1(O(e^{-n/2+1}) + 1 - p_1 + e^{-1} p_1) \\ &\quad + q_2(O(e^{-n/2+1}) + 1 - p_2 + e^{-1} p_2) \\ &\quad + q_3(O(e^{-n/2+1}) + 1 - p_3 + e^{-1} p_3) \\ &\quad + q_4(O(e^{-n/2+1}) + 1 - p_4 + e^{-1} p_4) \\ &\quad + q_5(O(e^{-n/2+1}) + 1 - p_5 + e^{-1} p_5) \\ &\quad + q_6(O(e^{-n/2+1}) + 1 - p_6 + e^{-1} p_6). \end{aligned}$$

Since $1 - p_i + e^{-1} p_i < 1$, $\sum_{i=1}^6 q_i = 1$, and $\lim_{n \rightarrow \infty} O(e^{-n/2+1}) = 0$, for sufficiently large n , there exists a positive constant $\rho < 1$ such that

$$\begin{aligned} E[e^{-d(\xi_{k+1})-d(\xi_k)} \mid d(\xi_k) > d_b] &\leq \sum_{i=1}^6 q_i (O(e^{-n/2+1}) + 1 - p_i + e^{-1} p_i) \\ &= \sum_{i=1}^6 q_i O(e^{-n/2+1}) + \sum_{i=1}^6 q_i (1 - p_i + e^{-1} p_i) < \rho < 1. \end{aligned}$$

This shows that $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies Condition 7.

The following analysis shows that Condition 8 can also be satisfied.

Let population ξ_k have the property of $d(\xi_k) = d_a = d_{\max}$, which implies that all individuals in the population are the same, i.e., $x_i = (1 \cdots 10)$ where x_i is an individual. Then the crossover has no influence on the drift, i.e., $d(\xi_{k+C}) = d(\xi_k)$.

There are only two events which may happen after mutation:

- (1) event $I\{d(\xi_{k+M}) < d(\xi_{k+C})\}$, and
- (2) event $I\{d(\xi_{k+M}) = d(\xi_{k+C})\}$.

Event $I\{d(\xi_{k+M}) > d(\xi_{k+C})\}$ cannot happen as $d(\xi_{k+C}) = d_{\max}$.

Similarly, there are following two cases after selection:

- (1) event $I\{d(\xi_{k+S}) < d(\xi_k)\}$, and
- (2) event $I\{d(\xi_{k+S}) = d(\xi_k)\}$.

Let the probability of the first event's happening be p_0 and that of the second be $1 - p_0$.

By summarizing all the above events, we have

$$\begin{aligned}
& E[e^{-(d(\xi_{k+1})-d_a)} \mid d(\xi_k) = d_a] \\
&= E[e^0 I\{d(\xi_{k+S}) < d(\xi_k), d(\xi_{k+M}) = d(\xi_{k+C}) \mid d(\xi_k) = d_a\}] \\
&\quad + E[e^0 I\{d(\xi_{k+S}) < d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) \mid d(\xi_k) = d_a\}] \\
&\quad + E[e^1 I\{d(\xi_{k+S}) = d(\xi_k), d(\xi_{k+M}) < d(\xi_{k+C}) \mid d(\xi_k) = d_a\}] \\
&\leq e^0 + e^0 p_0 + e^1(1 - p_0) \\
&\leq 2 + e.
\end{aligned}$$

Let $D = 2 + e$, we arrive at Condition 8.

According to Theorem 10 and the fact that

$$d_a - d_b = n - 2 - 3n/4 = n/4 - 2,$$

there exist two positive numbers, δ_1 and δ_2 , such that

$$E[\tau \mid d(\xi_0) \geq d_a] \geq E[\tau' \mid d(\xi_0) \geq d_a] \geq \delta_1 e^{\delta_2(n/4-2)},$$

where δ_1 and δ_2 are independent of n . \square

5. Discussion on weaker drift conditions

Conditions 7 and 8 may not be easy to understand and verify for some applications, e.g., the subset sum problem. The question now is whether we could come up with some weaker and more intuitive drift conditions. This section discusses such conditions.

Assume that a distance function $d(x)$ has been defined. Let $d_{\max} = \max\{d(x) : x \in X\}$, and d_b and d_a be positive numbers such that $d_b < d_a \leq d_{\max}$.

Condition 9. Let X be a population such that $d_b < d(X) < d_a$. Then

$$E[d(\xi_{k+1}) - d(\xi_k) \mid \xi_k = X] \geq C_1,$$

where $C_1 > 0$.

This condition is a simplified version of Condition 7. It implies that when a population X is in the area (d_a, d_b) , its offspring tends to drift away from, rather than move closer to, the optimal solution.

Condition 10. Let X be a population such that $d(X) \geq d_a$. Then

$$E[d(\xi_{k+1}) - d_a \mid \xi_k = X] \geq -C_2,$$

where $C_2 > 0$.

This condition is a simplified version of Condition 8. It implies that for a population X in the area $[d_a, +\infty)$, its offspring will not drift too far away from $[d_a, +\infty)$. The drift is bounded by $d_a - C_2$.

An interesting question now is: Given a random sequence $\{d(\xi_k); k = 0, 1, \dots\}$, generated by the EA for solving the subset sum problem, which satisfies Conditions 9 and 10,

whether the average time $E[\tau \mid \xi_0 = X]$ is still at least exponential in the problem size n starting from an initial population X with $d(X) \geq d_a$?

The answer to this question is negative. The two conditions are not sufficient to derive a positive answer although Condition 9 appears to imply a positive answer. We will show in the following using an example that the random sequence $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies both conditions, but EA's average computation time is polynomial in the problem size n .

Consider a family of subset sum problems, $\{W_1, W_2, \dots, W_n, \dots\}$, where

$$\begin{aligned} W_n &= \{w_1, \dots, w_n\}, \\ w_1 &= w_2 = \dots = w_{n-1} = 2n, w_n = 2nC_0 + 1, \\ C &= 2nC_0 + 1, \end{aligned}$$

where $C_0 \leq n/2$ is an integer greater than 2.

It is obvious that the subset $\{w_n\}$ is the unique optimal solution. Any subset of $W_n - \{w_n\}$ with no more than C_0 elements is a feasible solution.

The EA for solving the above family of subset sum problems follows the framework given in Section 2, but its implementation is much simpler than the EAs used in previous sections. The same chromosome representation is used as before. The EA used here is a $(1 + 1)$ EA. It uses bit mutation. A single bit s_i is chosen at random from an individual and flipped. If the offspring is infeasible, the parent will be retained. A new intermediate population $\xi_{k+M} = \{y\}$ is formed after mutation.

In terms of selection, the individual with better fitness between ξ_k and ξ_{k+M} has a higher survival probability of p , and that with worse fitness has a survival probability of $q = 1 - p$, where $p > q$. The new intermediate population after selection will be $\xi_{k+S} = \{z\}$.

Since $s^* = (0 \dots 01)$ is the optimal solution, define

$$d(x) = \sum_{i=1}^n |s_i - s_i^*| \quad \text{and} \quad d_{\max} = \max\{d(x): x \text{ is a feasible solution}\}.$$

It is easy to see that if $d(x_1) > d(x_2) > 1$, then $f(x_1) > f(x_2)$.

Let $d_b = 2$ and $d_a = d_{\max}$. We first verify that $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies Condition 9.

Let $\xi_k = \{x\}$ be the current population with $d_b < d(x) < d_a$, and $\xi_{k+M} = \{y\}$ be the intermediate population after mutation. Then one of the following three events may happen: event $I\{d(y) > d(x)\}$, event $I\{d(y) = d(x)\}$ or event $I\{d(y) < d(x)\}$.

Since the mutation is a bit mutation, it is easy to show that the probability of event $I\{d(y) > d(x)\}$ happening is

$$P(d(y) > d(x) \mid d_b < d(x) < d_a) = (n - d(x) - 1)/n.$$

The probability of event $I\{d(y) = d(x)\}$ happening is

$$P(d(y) = d(x) \mid d_b < d(x) < d_a) = 1/n.$$

The probability of event $I\{d(y) < d(x)\}$ happening is

$$P(d(y) < d(x) \mid d_b < d(x) < d_a) = d(x)/n.$$

If event $I\{d(y) < d(x)\}$ has happened, one of the following two events may happen: event $I\{d(z) < d(x)\}$ or event $I\{d(z) = d(x)\}$. According to our selection scheme, the

probability of event $I\{d(z) = d(x)\}$ happening is p and the probability of event $I\{d(z) < d(x)\}$ happening is q .

If event $I\{d(y) > d(x)\}$ has happened, one of the following two events may happen: event $I\{d(z) > d(x)\}$ or event $I\{d(z) = d(x)\}$. It is easy to see that the probability of event $I\{d(z) > d(x)\}$ happening is p and the probability of event $I\{d(z) = d(x)\}$ happening is q .

Summarizing all the above events, we have

$$\begin{aligned} & E[d(\xi_{k+1}) - d(\xi_k) \mid d_b < d(\xi_k) < d_a] \\ &= E[(d(z) - d(x))I\{d(y) > d(x), d(z) = d(y)\} \mid d_b < d(\xi_k) < d_a] \\ &\quad + E[(d(z) - d(x))I\{d(y) > d(y), d(z) = d(x)\} \mid d_b < d(\xi_k) < d_a] \\ &\quad + E[(d(z) - d(x))I\{d(y) < d(x), d(z) = d(y)\} \mid d_b < d(\xi_k) < d_a] \\ &\quad + E[(d(z) - d(x))I\{d(y) < d(x), d(z) = d(x)\} \mid d_b < d(\xi_k) < d_a] \\ &= p \frac{n - d(x) - 1}{n} - q \frac{d(x)}{n} \\ &\geq p \frac{n - d_a - 1}{n} - q \frac{d_a}{n}. \end{aligned}$$

Since $d_a = d_{\max} < C_0 \leq n/2$ and $p > q$,

$$E[d(\xi_{k+1}) - d(\xi_k) \mid d_b < d(\xi_k) < d_a] \geq p \frac{n - n/2 - 1}{n} - q \frac{n/2}{n} \geq \frac{1}{4}(p - q)$$

when n is sufficiently large. Let $C_1 = (p - q)/4 > 0$. Then Condition 9 is satisfied.

Now we verify that $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies Condition 10.

From the definition of feasible solutions, we know that any subset with a cardinality greater than C_0 is an infeasible solution. For any given individual x with $d(x) = d_a = d_{\max}$, let y be its offspring after mutation. Then one of the following events may happen: event $I\{d(y) = d_a\}$ or event $I\{d(y) < d_a\}$.

The probability of event $I\{d(y) = d_a\}$ happening is

$$P(d(y) = d(x) \mid d(x) = d_a = d_{\max}) = (n - d_a)/n.$$

The probability of event $I\{d(y) < d_a\}$ happening is

$$P(d(y) < d(x) \mid d(x) = d_a = d_{\max}) = d_a/n.$$

If event $I\{d(y) = d_a\}$ happens, then event $I\{d(z) = d_a\}$ happens with probability 1. If event $I\{d(y) < d_a\}$ happens, one of the following two may happen: event $I\{d(z) = d_a\}$ or event $I\{d(z) < d_a\}$. It is clear that the probability of event $I\{d(z) = d_a\}$ happening is p and the probability of event $I\{d(z) < d_a\}$ happening is q .

Summarizing all the above events, we obtain

$$\begin{aligned} & E[d(\xi_{k+1}) - d_a \mid d(\xi_k) \geq d_a] \\ &= E[(d(z) - d_a)I\{d(z) = d(y), d(y) < d_a\} \mid d(\xi_k) \geq d_a] \\ &\quad + E[(d(z) - d_a)I\{d(z) = d_a\} \mid d(\xi_k) \geq d_a] \\ &= -q \frac{d_a}{n}. \end{aligned}$$

Let $C_2 = qd_a/n$. Then Condition 10 is satisfied.

From the above analysis, we know that $\{d(\xi_k); k = 0, 1, \dots\}$ satisfies Conditions 9 and 10. It appears that it would take at least an exponential time to reach the optimal solution. However, this is not the case. We can show that the EA's average computation time for solving the family of subset sum problems is polynomial in n in spite of the fact that Conditions 9 and 10 can be satisfied.

Let ξ_0 be any individual (feasible solution) with $d_b < d(\xi_0) \leq d_{\max}$, we will prove that

$$P(\xi_{t=C_0} = x^* \mid d_b < d(\xi_0) \leq d_{\max}) \geq \left(\frac{q}{n}\right)^{C_0} q,$$

where x^* represents the optimal population (i.e., solution).

According to the mutation defined in our EA, the probability of event $I\{d(\xi_{k+M}) < d(\xi_k)\}$ happening is

$$P(d(\xi_{k+M}) < d(\xi_k) \mid d_b < d(\xi_k) \leq d_{\max}) \geq \frac{d(x)}{n} \geq \frac{1}{n}.$$

According to the selection, the probability of event $I\{d(\xi_{k+1}) = d(\xi_{k+M})\}$ happening is q . So

$$P(d(\xi_{k+S}) < d(\xi_k) \mid d(\xi_{k+M}) < d(\xi_k), d_b < d(\xi_k) \leq d_{\max}) \geq \frac{q}{n}.$$

Therefore

$$\begin{aligned} P(d(\xi_{t=C_0}) = 0 \mid d_b < d(\xi_0) \leq d_{\max}) \\ &\geq P(\xi_{t=C_0} = x^*, I\{d(\xi_0) > d(\xi_1) > d(\xi_2) > \dots > d(\xi_t)\} \mid \xi_0) \\ &\geq \left(\frac{q}{n}\right)^{C_0} q > 0. \end{aligned}$$

Since $\{\xi_k; k = 0, 1, \dots\}$ is homogeneous, we can obtain the following result:

$$E[\tau \mid \xi_0] = O(n^{C_0}).$$

In other words, the average computation time to solve this family of problems is polynomial in the problem size.

6. Conclusions and further work

This paper presents a number of new results on the computational time complexity of EAs. It has established several general conditions under which EAs will have polynomial or exponential average time complexity. It also shows successfully how these general results can be applied to different problems.

This paper also introduces a new approach to analyzing EAs, i.e., by drift analysis. Instead of estimating the first hitting time directly (which may be difficult in some cases), we can estimate the drift (which may be easier) first and then use the result to derive the upper or lower bounds of the first hitting time. Using drift analysis, we have shown several important theorems. For example, Theorem 1 gives some general conditions under which an EA can solve a problem in polynomial time on average. Theorem 10 gives some general

conditions under which an EA needs at least exponential computation time on average to solve a problem.

There are, however, much further work to be done to fully understand the computational time complexity of various EAs on different classes of problems.

- (1) This paper has assumed implicitly that the number of generations (or equivalently the number of fitness evaluations) is the most important factor in determining the order of EA's computation time. While this is true for a vast majority of EA applications where fitness evaluation is the most time consuming part of EA's execution time, it is possible in theory that the computation time of a single crossover, mutation and/or selection might be significant. In all EAs we considered in this paper, crossover and mutation only takes $O(n)$ time and selection needs $O(n \log n)$, where n is the problem size.
- (2) Algorithm parameters, e.g., mutation rate, may have a significant impact on the time complexity of an EA. It would be interesting to investigate how significant such impact is for different problems. In general, it would be quite interesting to investigate the average computation time of different EAs on a problem in order to gain some insights into the effectiveness and efficiency of different operators and parameter settings.
- (3) More strict drift conditions need to be studied in order to obtain tighter upper-bounds of the average computation time.
- (4) More analysis on well-known combinatorial problems, such as graph matching, should be carried out to gain more insights into the question where the real power of EAs is and when they are efficient.

Acknowledgement

This work was partially supported by the State Key Laboratory of Software Engineering at Wuhan University, National Nature Science Foundation of China, the Australian Research Council, and the School of Computer Science, the University of Birmingham. Part of the work was done while the first author was visiting the School of Computer Science, University College, UNSW, ADFA, and the School of Computer Science, the University of Birmingham.

References

- [1] B.K. Ambati, J. Ambati, M.M. Mokhtar, Heuristic combinatorial optimization by simulated Darwinian evolution: A polynomial time algorithm for the traveling salesman problem, *Biological Cybernetics* 65 (1991) 31–35.
- [2] S. Droste, T. Jansen, I. Wegener, A rigorous complexity analysis of the $(1 + 1)$ evolutionary algorithm for linear functions with boolean inputs, *Evolutionary Computation* 6 (2) (1998) 185–196.
- [3] A.E. Eiben, G. Rudolph, Theory of evolutionary algorithms: A bird eye view, *Theoret. Comput. Sci.* 229 (1999) 3–9.
- [4] D.B. Fogel, *System Identification Through Simulated Evolution: A Machine Learning Approach to Modeling*, Ginn Press, Needham Heights, MA, 1991.

- [5] D.B. Fogel, Empirical estimation of the computation required to discover approximate solutions to the traveling salesman problem using evolutionary programming, in: D.B. Fogel, W. Atmar (Eds.), Proc. Second Ann. Conf. on Evol. Prog., Evolutionary Programming Society, La Jolla, CA, 1993, pp. 56–61.
- [6] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, MA, 1989.
- [7] B. Hajek, Hitting time and occupation time bounds implied by drift analysis with applications, *Adv. Appl. Probab.* 14 (1982) 502–525.
- [8] J. He, H. Huang, The computational time analysis of genetic algorithms, in: Proc. Fifth Chinese Joint Conference on Artificial Intelligence, Xi'an Jiaotong University Press, Xi'an, 1998, pp. 440–443.
- [9] J. He, H. Huang, L. Kang, The computational time of genetic algorithms for fully deceptive problem, *Chinese J. Comput.* 21 (9) (1999) 999–1003.
- [10] J. He, L. Kang, On the convergence rate of genetic algorithms, *Theoret. Comput. Sci.* 229 (1999) 23–39.
- [11] J.H. Holland, *Adaptation in Natural and Artificial Systems*, 1st edn., MIT Press, Cambridge, MA, 1992.
- [12] R.M. Karp, Reducibility among combinatorial problems, in: R. Miller, W. Thatcher (Eds.), *Complexity of Computer Computation*, Plenum, New York, 1972, pp. 85–103.
- [13] S. Khuri, T. Bäck, J. Heitkötter, An evolutionary approach to combinatorial optimization problems, in: D. Cizmar (Ed.), Proc. 22nd Annual ACM Computer Science Conference, ACM Press, New York, 1994, pp. 66–73.
- [14] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd edn., Springer-Verlag, Berlin, 1996.
- [15] J. Neveu, *Discrete-Parameter Martingales*, North Holland, Amsterdam, 1975.
- [16] G. Rudolph, How mutation and selection solve long path problem in polynomial expected time, *Evolutionary Computation* 4 (2) (1996) 195–205.
- [17] G. Rudolph, *Convergence Properties of Evolutionary Algorithms*, Verlag Dr. Kovač, Hamburg, 1997.
- [18] G. Rudolph, Finite Markov chain results in evolutionary computation: A tour d'horizon, *Fundamenta Informaticae* 35 (1–4) (1998) 67–89.
- [19] H.-P. Schwefel, *Evolution and Optimum Seeking*, Wiley, New York, 1995.
- [20] G.H. Sasaki, B. Hajek, The time complexity of maximum matching by simulated annealing, *J. ACM* 35 (2) (1988) 387–403.
- [21] R.L. Tweedie, Criteria for classifying general Markov chains, *Adv. Appl. Probab.* 8 (1976) 737–771.
- [22] E. van Nimwegen, J.P. Crutchfield, M. Mitchell, Statistical dynamics of the royal road genetic algorithm, *Theoret. Comput. Sci.* 229 (1999) 41–102.
- [23] E. van Nimwegen, J.P. Crutchfield, Optimizing epochal evolutionary search: Population-size dependent theory, *Machine Learning J.*, to appear.