From an Individual to a Population: An Analysis of the First Hitting Time of Population-Based Evolutionary Algorithms

Jun He and Xin Yao, *Senior Member, IEEE*

***Abstract*—Almost all analyses of time complexity of evolutionary algorithms (EAs) have been conducted for $(1+1)$ EAs only. Theoretical results on the average computation time of population-based EAs are few. However, the vast majority of applications of EAs use a population size that is greater than one. The use of population has been regarded as one of the key features of EAs. It is important to understand in depth what the real utility of population is in terms of the time complexity of EAs, when EAs are applied to combinatorial optimization problems. This paper compares $(1+1)$ EAs and $(N+N)$ EAs theoretically by deriving their first hitting time on the same problems. It is shown that a population can have a drastic impact on an EA's average computation time, changing an exponential time to a polynomial time (in the input size) in some cases. It is also shown that the first hitting probability can be improved by introducing a population. However, the results presented in this paper do not imply that population-based EAs will always be better than $(1+1)$ EAs for all possible problems.**

***Index Terms*—Evolutionary algorithms, first hitting time, population, time complexity.**

I. Introduction

EVOLUTIONARY algorithms (EAs) have been used to solve many combinatorial optimization problems [1]–[4]. Time complexity is a key issue in the analysis of various optimization algorithms [5]. It shows how efficient an algorithm is for a large problem. However, relatively few results on the time complexity of EAs on combinatorial optimization problems are available [6]–[9], which makes theoretical comparison between EAs and other optimization algorithms difficult. It is necessary to gain a deeper understanding of the time complexity of EAs in order to understand whether an EA is expected to scale well with the input size and when an EA is expected to provide the most benefits to a given problem.

There has been some work on the analysis of time complexity of $(1+1)$ EAs for certain simple functions [10], e.g., the ONE-MAX function [11]–[14], the linear function [15], and the unimodal function [16],[17]. Few results were obtained using EAs with a population size greater than one [8]. Because $(1+1)$ EAs do not include recombination and population-based selection, the results on $(1+1)$ EAs cannot be generalized to EAs with population size greater than one. It is important to understand the impact a population may have on an EA's average computation time. Such an understanding is expected to shed some light on the real utility of population-based EAs in combinatorial optimization [1], [2].

In this paper, we compare $(1+1)$ and $(N+N)$ EAs theoretically on two families of problems. We derive the first hitting time for $(1+1)$ and $(N+N)$ EAs, respectively. Such results enable us to observe when the time would be polynomial or exponential in input size. The mathematical techniques used in this paper follow those in the analytical approach to the passage time of Markov chains [18]. Unlike drift analysis [8], which estimates the first hitting time from the drift of a Markov chain, these techniques calculate the first hitting time of a Markov chain directly from the transition matrix. The advantage over the drift analysis is that an exact expression of the first hitting time can be obtained for some EAs. The disadvantage is that such exact expressions are difficult, if not impossible, to derive from transition matrices if they are too complex.

The rest of this paper is organized as follows. Section II introduces some notations, definitions, and theorems about the first hitting time of a Markov chain. Section III contains our main results. Given typical problems, we derive the first hitting time of $(1+1)$ and $(N+N)$ EAs, respectively, and try to answer the following questions.

1) Will an $(N+N)$ EA change the time complexity of a $(1+1)$ EA for a given problem, e.g., from exponential to polynomial time or vice verse?
2) How much will an $(N+N)$ EA change the mean first hitting probability of a $(1+1)$ EA for a given problem?
3) How much will an $(N+N)$ EA change the mean first hitting time of a $(1+1)$ EA for a given problem?

Even partial answers to the above questions will deepen our understanding of population's roles in EAs. Finally, Section IV concludes the paper with a brief summary of the paper and a few remarks.

II. Evolutionary Algorithms, Markov Chain Models, and First Hitting Time

A combinatorial optimization problem can be described as follows [5]. Given a problem instance, i.e., a pair (S, c), where S is the set of feasible points and c is an objective function $c: S \to R^1$, the problem is to find an $x \in S$ such that

$$c(x) \geq c(y), \quad \forall y \in S$$

Manuscript received July 23, 2001; revised February 4, 2002. This work was supported in part by an EPSRC Grant (GR/R52541/01) and by the State Key Lab of Software Engineering, Wuhan University.

The authors are with the School of Computer Science, The University of Birmingham, Birmingham B15 2TT, U.K. (e-mail: x.yao@cs.bham.ac.uk).

Digital Object Identifier 10.1109/TEVC.2002.800886.

where x is called a global optimal solution to the given instance, or when no confusion can arise, simply an optimal solution. An optimization problem consists of a set I of problem instances. If S is discrete, the problem is known as a combinatorial one. Although not a requirement, the objective function will (or can be made to) take on only nonnegative values in many cases.

A. Evolutionary Algorithms and Markov Chain Models

In EAs, a point x in S is represented by an individual. A population is a collection of individuals. We use $(x_1, \ldots, x_N)$ to indicate a population with N individuals. The population space consists of all possible populations with N individuals. Because a population usually does not depend on the order of its individuals, we take the space $E = S^N/eqp$, where $eqp \in S^N \times S^N$ is the equivalence: $(x_1, \ldots, x_N) eqp (y_1, \ldots, y_N)$ if there exists the permutation δ such that $(x_1, \ldots, x_N) = (y_{\delta(1)}, \ldots, y_{\delta(N)})$.

An EA for solving a combinatorial optimization 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 produced as follows.

Recombination: Individuals in population X_t are recombined, denoted by

$$\text{rec: } E \to E$$

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

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

$$\text{mut: } E \to E$$

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

$$\text{sel: } E \times E \to E.$$

For most EAs, the state of population X_{t+1} depends only on the population X_t. In such a case, the process $(X_t; t = 0, 1, \ldots)$ can be modeled by a Markov chain [6], [19], [20], whose state space is the population space E, and the transition probability is

$$p_{ij}(t) = \mathbb{P}(X_{t+1} = j \mid X_0 = i), \quad i, j \in E.$$

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.

B. First Hitting Time and Time Complexity of an EA

Let $(X_t; t = 0, 1, \ldots)$ be the Markov chain associated with an EA. Its first hitting time is defined as follows.

Definition 1: Let E_{opt} be the set of populations that contain at least one optimal solution for a given instance of a combinatorial optimization problem. Then $T = \min\{t \geq 0; X_t \in E_{\text{opt}} \mid X_0\}$ is defined as the *first hitting time*.

For simplicity, we use the notation $H = E_{\text{opt}}$ from now on.

Definition 2: Given a combinatorial optimization problem,

$$\mathbb{E}[T \mid X_0 = i] = \sum_{t=0}^{\infty} t \mathbb{P}(T = t \mid X_0 = i)$$

is called the *mean first hitting time*, conditional on initial state i. Let μ_0 be the distribution of initial population X_0, then

$$\mathbb{E}[T] = \sum_i \mu_0(i) \mathbb{E}[T \mid X_0 = i]$$

is called the *mean first hitting time*.

It is worth noting that $\mathbb{E}[T]$ measures the average number of generations rather than the worst one for an EA before producing the final solution.

In the analysis of algorithms, we express the time requirements of algorithms in terms of the number of elementary steps [5]: arithmetic operations, comparisons, branching instructions, and so on, that is required for the execution of the algorithm on a hypothetical computer. The number of steps required by an algorithm is not the same for different inputs. We consider a distribution of initial inputs with given size n, and define the complexity of the algorithm for that input size to be the average-case behavior of the algorithm.

The input of an EA is represented as a string of symbols in this paper. The size of the input is the length of the string, which is n in our case. The time complexity of an EA is a function of the input size n for the given input, defined as follows.

Definition 3: Let n be the input size of an instance of a combinatorial optimization problem, and N the population size. Assume that during one generation of an EA, the numbers of operations in recombination, mutation, selection, and fitness evaluation are $f_r(n)$, $f_m(n)$, $f_s(n)$, and $f_o(n)$, respectively, then the *time complexity* of the EA for the given problem instance is

$$N\mathbb{E}[T](f_r(n) + f_m(n) + f_s(n) + f_o(n)) \tag{1}$$

which is a function of the input size n.

If one of $f_r(n), f_m(n), f_s(n)$, and $f_o(n)$ is an exponential function of n, the time complexity of the EA will be exponential, even for just one generation of the EA. Such EAs will be of little practical use and should be avoided in the algorithm design. The rest of this paper assumes that $f_r(n)$, $f_m(n)$, $f_s(n)$, and $f_o(n)$ are all polynomials of n.

In many EAs, most of the computation is spent on fitness evaluation in each generation, so the complexity of an EA can be simplified as

$$N\mathbb{E}[T] f_o(n) \tag{2}$$

and in a more simplified version $N\mathbb{E}[T]$.

C. Preliminary Theorems for First Hitting Time

We now provide some results on how to compute the first hitting time. The notations and theorems introduced below follow those given by Syski [18]. That discussion is for a general Markov process with a continuous-time parameter. A Markov chain with a discrete parameter can be regarded as its special case.

Let $(X_t; 0 \le t < \infty)$ be a homogeneous Markov process defined on a probability space with a discrete (countable) state–space E and a continuous-time parameter $t(0 \le t < \infty)$ and a standard stochastic matrix of transition probabilities, for $i, j \in E$,

$$\mathbf{P}(t) = (p_{ij}(t)) = (\mathbb{P}(X_t = j \mid X_0 = i))$$

with a conservative intensity matrix

$$\mathbf{Q} = (q_{ij})$$

where the associated intensities q_{ij} are defined by

$$q_{ij} = \lim_{t \to 0} \frac{p_{ij}(t)}{t}, \quad i \ne j$$

and

$$q_{ii} = \lim_{t \to +\infty} \frac{p_{ii}(t) - 1}{t} = -\sum_{j \ne i} q_{ij}.$$

For a Markov chain with a discrete-time parameter, let $\mathbf{Q} = \mathbf{P} - \mathbf{I}$.

Let H be a set in E in which we are interested (i.e., the set of global optima), and $T = \min\{t \ge 0; X_t \in H \mid X_0\}$.

The first hitting probability to a state $j \in H$, restricted to a finite T and conditional on initial state i, is defined by

$$D_{ij} = \mathbb{P}(X_T = j, T < \infty \mid X_0 = i).$$

Similarly

$$D_i = \mathbb{P}(T < \infty, X_T \in H \mid X_0 = i) = \sum_{j \in H} D_{ij}.$$

The mean first hitting time to a state $j \in H$, restricted to a finite T and conditional on initial state i, is defined by

$$m_{ij} = \mathbb{E}(T, T < \infty, X_T = j \mid X_0 = i).$$

Similarly

$$m_i = \mathbb{E}(T, T < \infty, X_T \in H \mid X_0 = i).$$

The following theorem by Syski [18, Th. 1, Ch. II, Sec. 3.1.3] gives the equations for the first hitting probability.

Theorem 1: The first hitting probability D_i satisfies the following:

$$\sum_k q_{ik} D_k = 0, \quad i \in H^c = E - H$$
$$D_i = 1, \quad i \in H. \tag{3}$$

The following theorem establishes the results on D_{ij} [18, Th. 3, Ch. II, Sec. 3.1.3].

Theorem 2: For $j \in H$

$$\sum_k q_{ik} D_{kj} = 0, \quad i \in H^c$$
$$D_{ij} = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \ne j, \end{cases} \quad i \in H. \tag{4}$$

The mean first hitting times m_i and m_{ij} are given by the following two theorems, respectively: [18, Th. 10, Ch. II, Sec. 3.1.1 and Th. 8, Ch. II, Sec. 3.1.2].

Theorem 3: For $i \in H^c$, the conditional means satisfy

$$q_i m_i = D_i + \sum_{j \in H^c, j \ne i} q_{ij} m_j \tag{5}$$

where $q_i = -q_{ii}$.

Theorem 4: For $i \in H^c, j \in H$, the conditional means satisfy

$$q_i m_{ij} = D_{ij} + \sum_{k \in H^c, k \ne i} q_{ik} m_{kj}. \tag{6}$$

For a Markov chain with a discrete-time parameter, (3)–(6) still hold.

III. From an Individual to a Population

A. Problems Considered in This Paper

The first family of objective functions is used in our studies in this paper:

$$c(x) = \begin{cases} n + 1 + \frac{\lambda n - n - 1}{\lambda n} \left(\sum_{i=1}^n s_i\right), & \forall 0 \le \sum_{i=1}^n s_i \le \lambda n \\ \sum_{i=1}^n s_i, & \forall \lambda n < \sum_{i=1}^n s_i \le n \end{cases} \tag{7}$$

where $x = (s_1 \dots s_n)$ is a binary string with length n and the parameter $\lambda \in (0, 1]$.

This is a simple but typical objective function, with one global optimal point: $(0 \dots 0)$. We construct the family of objective functions based only on the sum effect of s_i, without any interaction among s_i, which will be more difficult for certain EAs [21].

When $\lambda = 1/n$, the objective function becomes a deceptive problem for classical simple EAs [22], i.e.,

$$c(x) = \begin{cases} n + 1 - n \left(\sum_{i=1}^n s_i\right), & \forall 0 \le \sum_{i=1}^n s_i \le 1 \\ \sum_{i=1}^n s_i, & \forall 1 < \sum_{i=1}^n s_i \le n. \end{cases} \tag{8}$$

When $\lambda = 1$, the objective function becomes an easy unimodal problem for classical simple EAs, i.e.,

$$c(x) = n + 1 - \frac{1}{n} \left(\sum_{i=1}^n s_i\right), \quad \forall 0 \le \sum_{i=1}^n s_i \le n. \tag{9}$$

It is clear from the above cases that when λ changes from $1/n$ to 1, the difficulty of the problem changes from hard to easy for classical simple EAs.

The second family of objective functions is based on the first one, but more complex. A distributing term $\varepsilon(x)$ is added to the first function, i.e., $c'(x) = c(x) + \varepsilon(x)$. An example of such functions is given in (10), shown at the bottom of the next page.

Another example of such functions is to add function (9) with a distribution term -2 if the sum of bits $\sum_{i=1}^n s_i$ is odd. That is

$$c'(x) = \begin{cases} n + 1 - \frac{1}{n} \left(\sum_{i=1}^n s_i\right), & \text{if } \sum_{i=1}^n s_i \text{ is even} \\ n + 1 - \frac{1}{n} \left(\sum_{i=1}^n s_i\right) - 2, & \text{if } \sum_{i=1}^n s_i \text{ is odd} \end{cases} \tag{11}$$

Fig. 1 shows the above objective functions.

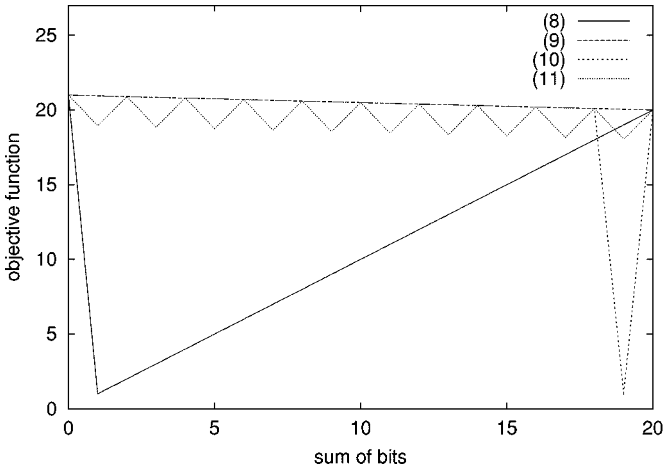

Fig. 1. Objective functions (8), (9), (10), and (11).

In this paper, we use the following function $d(Z)$ to measure how far a population is away from the optimal point. For an individual $x = (s_1 \ldots s_n)$, define the Hamming distance

$$d(x) = \sum_{i=1}^{n} (s_i - 0) \tag{12}$$

and, for a population Z, define

$$d(Z) = \min\{d(x); x \in Z\}. \tag{13}$$

B. Impact of Population on the Time Complexity of EAs

Despite the common wisdom in the evolutionary computation community that a population ought to benefit evolutionary search, few theoretical results are available on the existence of such benefit and how much benefit there is if it exists. There are some interesting open questions to be answered. For example, will the time complexity change if we introduce a population into EAs? Could an exponential-time $(1+1)$ EA be turned into a polynomial-time $(N+N)$ EA $(N > 1)$ by introducing a population? Would a polynomial-time $(1+1)$ EA still be polynomial in time after the introduction of a population?

We answer these questions by case studies using the problems given in Section III-A. We keep the mutation operator used in $(1+1)$ EAs and $(N+N)$ EAs the same all the time. However, population-based selection will be used for $(N+N)$ EAs, since $(1+1)$ EAs do not have population-based selection.

1) Population Can be Beneficial: An Example: First, we examine a case where an $(1+1)$ EA for the problem (8) is exponential in time, but its corresponding $(N+N)$ EA is polynomial in time. The framework of all EAs used in this paper is the same as that described in Section II-A. The details of the $(1+1)$ EA under consideration are given below.

Mutation I: Given a population (only a single individual in the case of $(1+1)$ EA) X_t at generation t. For each individual $(s_1 \ldots s_n)$ in the population, choose one bit from the individual and flip it. The mutated population is denoted as $X_t^{(m)}$.

Selection I: Assign the survival probability of the better individual between $X_t^{(m)}$ and X_t to be p $(0 \le p \le 1)$, and that of the worse individual to be $q = 1 - p$. Generate the next generation X_{t+1} using these survival probabilities.

The selection procedure given above includes a wide range of different selection schemes, depending on the choice of p. If $p = 1$, it is the elitist selection.

The corresponding $(N + N)$ EA uses Mutation I and the following population-based selection:

Selection II: Retain the best and worst individuals in the combined population of $X_t^{(m)}$ and X_t and assign other $2N - 2$ individuals positive survival probabilities $p_2, \ldots, p_{2N-1}$, respectively, based on their fitness. Select the next generation from $X_t^{(m)}$ and X_t using these survival probabilities.

Recombination is not used in the algorithm.

Proposition 1: Given the objective function (8), we have the following.

1) For the $(1+1)$ EA with Mutation I and Selection I, define $S_k = \{x \,|\, d(x) = k\}, k = 0, \ldots, n$ and $m_{S_k} = \mathbb{E}[T; T < \infty \,|\, X_0 \in S_k]$. If $p < 1$, the mean first hitting time m_{S_k} satisfies the equation shown at the bottom of the page, where if p/q is a constant greater than 1, then $m_{S_1}, \ldots m_{S_n}$ are exponential functions of n.

   If $p = 1$, the mean first hitting time m_{S_k} starting from an individual in S_k will be

$$\begin{cases} m_{S_0} = 0, \\ m_{S_1} = \frac{1}{n} \quad \text{(with probability} 1/n), \\ m_{S_k} = 0, \quad \mathbb{E}[T \,|\, X_0 \in S_k] = \infty, \quad k = 2, \ldots, n. \end{cases}$$

2) For the $(N + N)$ EA with Mutation I and Selection II, define $S_k = \{Z \,|\, d(Z) = k\}, k = 0, \ldots, n$, and define

$$c'(x) = \begin{cases} n + 1 - \frac{1}{n}\left(\sum_{i=1}^{n} s_i\right) - (n-1), & \text{if } \sum_{i=1}^{n} s_i = n - 1 \\ n + 1 - \frac{1}{n}\left(\sum_{i=1}^{n} s_i\right), & \text{else.} \end{cases} \tag{10}$$

$$\begin{cases} m_{S_0} = 0, \\ m_{S_1} = \frac{n}{p} + \sum_{j=0}^{n-2} \frac{n}{q(j+2)} \left(\frac{p}{q}\right)^j \left(\prod_{i=0}^{j} \frac{n-1-i}{1+i}\right) \\ m_{S_k} = m_{S_{k-1}} + \frac{n}{kq} + \sum_{j=0}^{n-k-1} \frac{n}{q(j+k+1)} \left(\frac{p}{q}\right)^{j+1} \left(\prod_{i=0}^{j} \frac{n-k-i}{k+i}\right), \quad k = 2, \ldots, n-1 \\ m_{S_n} = m_{S_{n-1}} + q^{-1}. \end{cases}$$

the mean first hitting time $m_{S_k} = \mathbb{E}[T; T < \infty \mid X_0 \in S_k]$. Then

$$\begin{cases} m_{S_0} = 0, \\ m_{S_k} \le \sum_{i=1}^{k} \frac{n}{i}, \quad k = 1, \dots, n \end{cases}$$

where $Nm_{S_1}, \dots Nm_{S_n}$ are polynomial functions of n.

Proof—Part (1): From the given $(1+1)$ EA, transition probabilities among S_k can be derived.

When $k = 1$, for any $i \in S_1$

$$\mathbb{P}(X_{t+1} \in S_0 \mid X_t = i) = \frac{1}{n}p$$
$$\mathbb{P}(X_{t+1} \in S_1 \mid X_t = i) = 1 - \frac{1}{n}p - \frac{(n-1)}{n}p$$
$$\mathbb{P}(X_{t+1} \in S_2 \mid X_t = i) = \frac{(n-1)}{n}p.$$

When $1 < k < n$, for any $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) = \frac{k}{n}q$$
$$\mathbb{P}(X_{t+1} \in S_k \mid X_t = i) = 1 - \left(\frac{k}{n}q + \frac{n-k}{n}p\right)$$
$$\mathbb{P}(X_{t+1} \in S_{k+1} \mid X_t = i) = \frac{n-k}{n}p.$$

When $k = n$, for any $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) = q$$
$$\mathbb{P}(X_{t+1} \in S_k \mid X_t = i) = 1 - q.$$

Introduce an auxiliary homogeneous Markov chain $(Y_t, t = 0, 1, \dots)$ whose state space is $\{0, \dots, n\}$ and transition probabilities are given by

$$\bar{p}_{00} = 1$$
$$\bar{p}_{kh} = \mathbb{P}(X_{t+1} \in S_h \mid X_t \in S_k), k = 1, \dots, n, h = 0, \dots, n.$$

For the Markov chain (Y_t), let its first hitting time to state 0 be $T_Y = \min\{t; Y_t = 0\}$, and the mean first hitting time $m_k = \mathbb{E}[T_Y; T_Y < \infty \mid Y_0 = k]$. Then it is apparent that $m_k = m_{S_k}$ for any $0 \le k \le n$.

Let $D_k = \mathbb{P}(T_Y < \infty \mid Y_0 = k)$. According to (3), the first hitting probabilities are given by (14), as shown at the bottom of the page.

If $p < 1$, we get

$$D_0 = D_1 = D_2 = \cdots D_n = 1.$$

If $p = 1$, then $\bar{p}_{nn} = 1$; this means that state n is an absorbing state and it is impossible for the EA to visit the global optimal state starting from state n. Let $D_n = 0$ for the equation $0D_n = 0$. We get

$$D_0 = 1, D_1 = \frac{1}{n}, D_2 = \cdots D_n = 0.$$

According to (5), the mean first hitting time satisfies (15), shown at the bottom of the page, when $p < 1$.

The above linear equations can be solved as shown in the last equation at the bottom of the page.

If p/q is a constant greater than 1, then $m_1, \dots, m_n$ are exponential functions of n.

When $p = 1$, the mean first hitting time satisfies the following equations:

$$\begin{cases} m_0 = 0, \\ m_1 = \frac{1}{n} + \frac{1}{n}m_0 + \frac{n-1}{n}m_2, \\ \frac{n-k}{n}m_k = 1 + \frac{n-k}{n}m_{k+1}, \quad k = 2, \dots n-1, \\ 0m_n = 0. \end{cases} \tag{16}$$

$$\begin{cases} D_0 = 1 \\ \frac{1}{n}pD_0 - \left(\frac{1}{n}p + \frac{(n-1)}{n}\right)pD_1 + \frac{(n-1)}{n}pD_2 = 0 \\ \frac{k}{n}qD_{k-1} - \left(\frac{k}{n}q + \frac{n-k}{n}p\right)D_k + \frac{n-k}{n}pD_k = 0, \quad k = 2, \dots, n-1 \\ qD_{n-1} - qD_n = 0. \end{cases} \tag{14}$$

$$\begin{cases} m_0 = 0 \\ \left(\frac{1}{n}p + \frac{n-1}{n}p\right)m_1 = 1 + \frac{1}{n}pm_0 + \frac{n-1}{n}pm_2 \\ \left(\frac{k}{n}q + \frac{n-k}{n}p\right)m_k = 1 + \frac{k}{n}qm_{k-1} + \frac{n-k}{n}pm_{k+1}, \quad k = 2, \dots n-1 \\ qm_n = 1 + qm_{n-1}. \end{cases} \tag{15}$$

$$\begin{cases} m_0 = 0 \\ m_1 = \frac{n}{p} + \sum_{j=0}^{n-2} \frac{n}{q(j+2)} \left(\frac{p}{q}\right)^j \left(\prod_{i=0}^{j} \frac{n-1-i}{1+i}\right) \\ m_k = m_{k-1} + \frac{n}{kq} + \sum_{j=0}^{n-k-1} \frac{n}{q(j+k+1)} \left(\frac{p}{q}\right)^{j+1} \left(\prod_{i=0}^{j} \frac{n-k-i}{k+i}\right), \quad k = 2, \dots, n-1 \\ m_n = m_{n-1} + q^{-1}. \end{cases}$$

Since $D_n = 0$, we should let $m_n = 0$. Hence, we obtain

$$\begin{cases} m_0 = 0, \\ m_1 = \frac{1}{n}, \\ m_k = 0, \quad k = 2, \ldots, n, \end{cases}$$

and for any $k \geq 2$, since $D_k = 0$, we have $\mathbb{E}[T_Y \,|\, Y_0 = k] = \infty$, and $\mathbb{E}[T \,|\, X_0 \in S_k] = \infty$.

Since $m_{S_k} = m_k$, we get the conclusion of Part (1).

Part (2) Since the given $(N+N)$ EA always keeps the best and worst individuals, the transition probabilities among S_k can be derived as follows.

For any $1 \leq k \leq n$ and for any population $i \in S_k$, we have

$$\mathbb{P}(X_{t+1} \in S_{k-1} \,|\, X_t = i) \geq \frac{k}{n}$$
$$\mathbb{P}(X_{t+1} \in S_k \,|\, X_t = i) \leq 1 - \frac{k}{n}$$
$$\mathbb{P}(X_{t+1} \in S_{k+1} \,|\, X_t = i) = 0.$$

Introduce an auxiliary homogeneous Markov chain $(Y_t; t = 0, 1, \ldots)$ whose state space is $\{0, \ldots, n\}$ and transition probability $\bar{p}_{kh}$ satisfies

$$\bar{p}_{00} = 1$$

and for $k = 1, \ldots, n$

$$\bar{p}_{kh} = \begin{cases} \frac{k}{n}, & h = k-1 \\ 1 - \frac{k}{n}, & h = k, \\ 0, & \text{otherwise.} \end{cases}$$

For Markov chain (Y_t), let the first hitting time time to state 0 be $T_Y = \min\{t; Y_t = 0\}$, and the mean first hitting time $m_k = \mathbb{E}[T_Y; T_Y < \infty \,|\, Y_0 = k]$. It is apparent that $m_k \geq m_{S_k}$ for $k = 1, \ldots, n$.

For the Markov chain (Y_t), let $D_k = \mathbb{P}(T_Y < \infty \,|\, Y_0 = k)$. According to (3), the first hitting probabilities are given by

$$\begin{cases} D_0 = 1, \\ \frac{k}{n} D_{k-1} - \frac{k}{n} D_k = 0, \quad k = 1, \ldots n. \end{cases} \tag{17}$$

Hence, we get

$$D_0 = D_1 = D_2 = \cdots D_n = 1.$$

According to (5), the mean first hitting times satisfy

$$\begin{cases} m_0 = 0, \\ \frac{k}{n} m_k = 1 + \frac{k}{n} m_{k-1}, \quad k = 1, \ldots, n, \end{cases} \tag{18}$$

from which we can get

$$\begin{cases} m_0 = 0, \\ m_k = \sum_{i=1}^{k} \frac{n}{i} \quad k = 1, \ldots, n, \end{cases}$$

where each m_k is a polynomial function of n.

Since $m_{S_k} \leq m_k$, we come to the conclusion of Part (2). □

The above proposition gives an estimation to the mean first hitting times of the EAs starting from different states. From these expressions, we can get the details of the EA's time complexity. It is easy to see that for the $(N+N)$ EA, $m_{S_1} = O(n)$ and $m_{S_n} = O(n \log n)$. For the $(1+1)$ EA, the order of m_{S_1} is $qp^{-2}(1 + (p/q)^n$. If $p, q > 0$ are constants, the order will be $\Omega((1 + (p/q))^n)$. For other $k > 1, m_{S_k}$ is at least $\Omega((1 + (p/q)^n)$.

2) Population Can Be Beneficial: Another Example: We investigate another example: function (11). The $(1+1)$ EA still uses Mutation I and Selection I for this function. The $(N+N)$ EA uses Mutation I and the following selection.

Selection III: Retain the best individual in the combined population of $X_t^{(m)}$ and X_t. Assign the remaining $2N-1$ individuals survival probabilities $p_2 \geq p_3 \geq \cdots \geq p_{2N} > 0$ based on their fitness from high to low, respectively. Generate the next generation X_{t+1} using such survival probabilities.

It is clear that Selection III includes Selection II.

For convenience, assume that n is an even number in the following proposition. For the case of n being odd, the proof is similar.

Proposition 2: Given the objective function (11), we have the following.

1) For the $(1+1)$ EA with Mutation I and Selection I, let $S_k = \{x \,|\, d(x) = k\}, k = 0, \ldots, n$ and the mean first hitting time $m_{S_k} = \mathbb{E}[T, T < \infty \,|\, X_0 \in S_k]$. Then we have the equation shown at the bottom of the page, where

$$\theta(k) = \begin{cases} p, & \text{if } k \text{ is an odd number in } (0, n) \\ q, & \text{if } k \text{ is an even number in } (0, n) \end{cases}$$

and $m_{S_1}, \ldots, m_{S_n}$ are exponential functions of n.

2) For the $(N+N)$ EA with Mutation I and Selection III, let $S_k = \{Z \,|\, d(Z) = k\}, k = 0, \ldots, n$. Define the mean first hitting time $m_{S_k} = \mathbb{E}[T, T < \infty \,|\, X_0 \in S_k]$, and let $q = \min\{p_1, \ldots, p_{2N}\}$. Then m_{S_k} satisfies the first equation at the bottom of the next page where if q^{-1} is a polynomial function of n or a constant, $Nm_{S_1}, \ldots, Nm_{S_n}$ are polynomial functions of n.

Proof—Part (1): From the description of the above $(1+1)$ EA, the transition probabilities of (X_t) among S_k can be derived. □

When k is an odd number in $(0, n)$, for any $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \,|\, X_t = i) = \frac{k}{n} p$$
$$\mathbb{P}(X_{t+1} \in S_k \,|\, X_t = i) = 1 - p$$
$$\mathbb{P}(X_{t+1} \in S_{k+1} \,|\, X_t = i) = \frac{n-k}{n} p.$$

$$\begin{cases} m_{S_0} = 0, \\ m_{S_k} = m_{S_{k-1}} + \frac{n}{k\theta(k)} + \sum_{j=0}^{n-k-1} \frac{n}{(j+k+1)\theta(j+k+1)} \left(\prod_{i=0}^{j} \frac{n-k-i}{k+i} \right), \quad k = 1, \ldots, n-1 \\ m_{S_n} = m_{S_{n-1}} + q^{-1} \end{cases}$$

When k is an even number in $(0,n)$, for any $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) = \frac{k}{n}q$$
$$\mathbb{P}(X_{t+1} \in S_k \mid X_t = i) = 1 - q$$
$$\mathbb{P}(X_{t+1} \in S_{k+1} \mid X_t = i) = \frac{n-k}{n}q.$$

When $k = n$, for any $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) = q$$
$$\mathbb{P}(X_{t+1} \in S_k \mid X_t = i) = 1 - q.$$

Introduce an auxiliary homogeneous Markov chain $(Y_t, t = 0,1,\ldots)$ whose state space is $\{0,\ldots,n\}$ and transition probabilities are given by

$$\bar{p}_{00} = 1$$
$$\bar{p}_{kh} = \mathbb{P}(X_{t+1} \in S_h \mid X_t \in S_k)$$
$$k, = 1,\ldots,n, \quad h = 0,\ldots,n.$$

Now let $\mathbf{P} = (\bar{p}_{ij})$.

For Markov chain (Y_t), define the first hitting time $T_Y = \min\{t; Y_t = 0\}$ and its mean $m_k = \mathbb{E}[T_Y, T_Y < \infty \mid Y_0 = k]$. It is apparent that $m_k = m_{S_k}$ for any $0 \le k \le n$.

Define the first hitting probability $D_k = \mathbb{P}(T_Y < \infty \mid Y_0 = k)$ for Markov chain (Y_t). From its transition probability matrix $\mathbf{P}$ and (3), we have

$$D_0 = D_1 = \cdots D_n = 1.$$

According to (5), the mean first hitting times satisfy (19), shown at the bottom of the page, from which we can obtain the last equation at the bottom of the page.

It is clear from the above equations that $m_k, k = 1,\ldots,n$ are exponential functions of n for $p \in [0,1]$.

Part (2) From the description of the above $(N+N)$ EA, the transition probabilities of Markov chain $(X_t; t = 0,1,\ldots)$ among S_k can be derived as follows.

When k is an even number in $(0,n)$, for any $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) \ge \frac{k}{n}q,$$
$$\mathbb{P}(X_{t+1} \in S_k \mid X_t = i) \le 1 - \frac{k}{n}q$$

and for $h \ne k-1, k$

$$\mathbb{P}(X_{t+1} \in S_h \mid X_t = i) = 0.$$

When k is an odd number in $(0,n)$, for any $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) \ge \frac{k}{n}$$
$$\mathbb{P}(X_{t+1} \in S_k \mid X_t = i) \ge 0$$
$$\mathbb{P}(X_{t+1} \in S_{k+1} \mid X_t = i) \le \frac{n-k}{n}$$

and for $h \ne k-1, k, k+1$

$$\mathbb{P}(X_{t+1} \in S_h \mid X_t = i) = 0.$$

Introduce an auxiliary homogeneous Markov chain $(Y_t, t = 0,1,\ldots)$ whose state space is $\{0,\ldots,n\}$ and transition probabilities are given by

$$\bar{p}_{00} = 1,$$
$$\bar{p}_{kh} = \begin{cases} \frac{k}{n}q, & h = k-1 \\ 1 - \frac{k}{n}q, & h = k, \\ 0, & \text{otherwise.} \end{cases} \quad \text{for any even } k \text{ in } (0,n]$$
$$\bar{p}_{kh} = \begin{cases} \frac{k}{n}, & h = k-1 \\ \frac{n-k}{n}, & h = k+1 \\ 0, & \text{otherwise.} \end{cases} \quad \text{for any odd } k \text{ in } (0,n).$$

$$\begin{cases} m_{S_0} = 0 \\ m_{S_k} \le m_{S_{k-1}} + \frac{n}{k} + \frac{(n-k)n}{k(k+1)q}, & k \text{ is an odd number in } (2,n) \\ m_{S_k} \le m_{S_{k-1}} + \frac{n}{kq}, & k \text{ is an even number in } (0,n] \end{cases}$$

$$\begin{cases} m_0 = 0 \\ qm_k = 1 + \frac{k}{n}qm_{k-1} + \frac{n-k}{n}qm_{k+1}, & k \text{ is an even number in } (0,n) \\ pm_k = 1 + \frac{k}{n}pm_{k-1} + \frac{n-k}{n}pm_{k+1}, & k \text{ is an odd number in } (0,n) \\ qm_n = 1 + qm_{n-1} \end{cases} \tag{19}$$

$$\begin{cases} m_0 = 0 \\ m_k = m_{k-1} + \frac{n}{k\theta(k)} + \sum_{j=0}^{n-k-1} \frac{n}{(j+k+1)\theta(j+k+1)} \left(\prod_{i=0}^{j} \frac{n-k-i}{k+i} \right), & k = 1,\ldots,n-1 \\ m_n = m_{n-1} + q^{-1}. \end{cases}$$

For Markov chain (Y_t), define T_Y and m_k as the same as those in Part (1). We have $m_k \geq m_{S_k}$ for any $0 \leq k \leq n$.

Let $\mathbf{P} = (\bar{p}_{ij})$ define the first hitting probability $D_k = \mathbb{P}(T_Y < \infty \mid Y_0 = k)$. According to (3)

$$D_0 = D_1 = D_2 = \cdots D_n = 1.$$

According to (5), the mean first hitting time satisfies

$$\begin{cases} m_0 = 0, \\ m_k = 1 + \frac{k}{n} m_{k-1} + \frac{n-k}{n} m_{k+1}, & \text{for any odd } k \text{ in } (0, n) \\ \frac{k}{n} q m_k = 1 + \frac{k}{n} q m_{k-1}, & \text{for any even } k \text{ in } (0, n] \end{cases} \tag{20}$$

from which we get the first equation shown at the bottom of the page. It is easy to see that $m_k, k = 1, \dots, n$, are polynomial functions of n if q^{-1} is a polynomial function of n or a constant. □

From the above proposition, we can also estimate the time complexity of the EAs. For example, for the $(1+1)$ EA, $m_{S_1} = \Omega(2^n/p)$ (assuming $p \geq q$). For the $(N+N)$ EA, $m_{S_k} \leq O(n^2/q)$.

3) Population May Not be Beneficial: In general, if the first hitting time of a $(1+1)$ EA is polynomial in the input size, the first hitting time of the corresponding $(N+N)$ EA will also be polynomial, except for some cases where the selection pressure is very high in the $(N+N)$ EA.

Consider the objective function (10). A $(1+1)$ EA uses Mutation I and the following selection.

Selection IV: If $X_t^{(m)}$ is better than X_t, let $X_{t+1} = X_t^{(m)}$. If $X_t^{(m)}$ is worse than X_t, let $X_{t+1} = X_t^{(m)}$ with probability $1/n$. If they are the same, select X_t and $X_t^{(m)}$ uniformly at random as the next population X_{t+1}.

Selection IV is similar to simulated annealing with a fixed temperature [23]–[25].

The corresponding $(N+N)$ EA still uses Mutation I and Selection III.

Proposition 3: Given objective function (10), we have the following.

1) For the $(1+1)$ EA using Mutation I and Selection IV, let $S_k = \{x \mid d(x) = k\}, k = 0, \dots, n$, and the mean first hitting time $m_{S_k} = \mathbb{E}[T, T < \infty \mid Y_0 \in S_k]$. Then we have the second equation shown at the bottom of the page, where $m_{S_1}, \dots, m_{S_n}$ are polynomials in n.
2) For the $(N+N)$ EA using Mutation I and Selection III, define $S_k = \{Z \mid d(Z) = k\}, k = 0, \dots, n$. Let the mean first hitting time $m_{S_k} = \mathbb{E}[T, T < \infty \mid X_0 \in S_k], q = \max\{p_{N+1}, \dots, p_{2N}\}$. Then m_{S_k} satisfies, for $k = 1, \dots, n-2$, the last equation shown at the bottom of the page, where $Nm_{S_{n-1}}$ and Nm_{S_n} are exponential in n if q^{-1} is exponential in n.

Proof—Part (1): From the description of the above $(1+1)$ EA, the transition probabilities of (X_t) among S_k can be derived.

When $k = 1, \dots, n-2$, for any $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) = \frac{k}{n},$$
$$\mathbb{P}(X_{t+1} \in S_k \mid X_t = i) = 1 - \frac{k}{n} - \frac{(n-k)}{n^2},$$
$$\mathbb{P}(X_{t+1} \in S_{k+1} \mid X_t = i) = \frac{(n-k)}{n^2}.$$

When $k = n-1$, for any $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{n-2} \mid X_t = i) = \frac{n-1}{n}$$
$$\mathbb{P}(X_{t+1} \in S_n \mid X_t = i) = \frac{1}{n}.$$

$$\begin{cases} m_0 = 0, \\ m_k = m_{k-1} + \frac{n}{k} + \frac{n-k}{k} \frac{n}{(k+1)q}, & k \text{ is an odd number in } (0, n) \\ m_k = m_{k-1} + \frac{n}{kq}, & k \text{ is an even number in } (0, n]. \end{cases}$$

$$\begin{cases} m_{S_0} = 0 \\ m_{S_k} = m_{S_{k-1}} + \frac{n}{k} + \sum_{j=0}^{n-k-2} \frac{n}{j+k+1} \left(\prod_{i=0}^{j} \frac{n-k-i}{(k+i)n} \right) + \frac{n}{n-1} \left(\prod_{i=0}^{n-k-2} \frac{n-k-i}{(k+i)n} \right), & k = 1, \dots, n-2 \\ m_{S_{n-1}} = m_{S_{n-2}} + \frac{2n}{n-1} \\ m_{S_n} = m_{S_{n-1}} + n \end{cases}$$

$$\begin{cases} m_{S_0} = 0, \\ m_{S_k} \geq \sum_{i=1}^{k} \left(1 - \left(1 - \frac{i}{n}\right)^N\right)^{-1} \\ m_{S_{n-1}} \geq \sum_{i=1}^{n-1} \left(1 - \left(1 - \frac{i}{n}\right)^N\right)^{-1} + \left(1 - \left(\frac{1}{n}\right)^N\right)^{-1} \left(\frac{1}{n}\right)^N (1 - (1-q)^N)^{-1} \\ m_{S_n} \geq (1 - (1-q)^N)^{-1} \left(1 + (1 - (\frac{1}{n})^N)^{-1} (\frac{1}{n})^N\right) + \sum_{i=1}^{n-1} \left(1 - \left(1 - \frac{i}{n}\right)^N\right)^{-1}. \end{cases}$$

When $k=n$, for any $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{n-1} \mid X_t = i) = \frac{1}{n}$$
$$\mathbb{P}(X_{t+1} \in S_n \mid X_t = i) = \frac{n-1}{n}.$$

Introduce an auxiliary homogeneous Markov chain $(Y_t, t = 0, 1, \ldots)$ whose state space is $\{0, \ldots, n\}$ and transition probabilities are given by

$$\bar{p}_{00} = 1$$
$$\bar{p}_{kh} = \mathbb{P}(X_{t+1} \in S_h \mid X_t \in S_k)$$
$$k = 1, \ldots, n, \quad h = 0, \ldots, n.$$

For Markov chain (Y_t), let the first hitting time to the state 0 be $T_Y = \min\{t; Y_t = 0\}$ and its mean $m_k = \mathbb{E}[T_Y, T_Y < \infty \mid Y_0 = k]$. It is clear that $m_k = m_{S_k}$, for any $0 \le k \le n$.

Let $\mathbf{P} = (\bar{p}_{ij})$, and define the first hitting probability to be $D_k = \mathbb{P}(T_Y < \infty \mid Y_0 = k)$. From (3), we can get

$$D_0 = D_1 = D_2 = \cdots D_n = 1.$$

According to (5), the mean first hitting time satisfies (21), shown at the bottom of the page, from which we can get the second equation shown at the bottom of the page. It can be seen from the above equations that $m_k, k = 1, \ldots, n$ are polynomial in n.

Part (2) From the description of the above $(N+N)$ EA, the transition probabilities of $(X_t; t = 0, 1, \ldots)$ among S_k can be derived.

When $k = 1, \ldots, n-2$, for any $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) \le 1 - \left(1 - \frac{k}{n}\right)^N$$
$$\mathbb{P}(X_{t+1} \in S_k \mid X_t = i) \ge \left(1 - \frac{k}{n}\right)^N$$

When $k = n-1$, for any $i \in S_k$,

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) \le 1 - \left(1 - \frac{n-1}{n}\right)^N$$
$$\mathbb{P}(X_{t+1} \in S_{k+1} \mid X_t = i) \ge \left(\frac{1}{n}\right)^N$$

When $k = n$, for any $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{n-1} \mid X_t = i) \le 1 - (1-q)^N$$
$$\mathbb{P}(X_{t+1} \in S_n \mid X_t = i) \ge (1-q)^N.$$

Introduce an auxiliary homogeneous Markov chain $(Y_t, t = 0, 1, \ldots)$ in the state space $\{0, \ldots, n\}$ whose transition probabilities are given by the last equation at the bottom of the page.

$$\begin{cases} m_0 = 0, \\ \left(\frac{k}{n} + \frac{n-k}{n^2}\right) m_k = 1 + \frac{k}{n} m_{k-1} + \frac{n-k}{n^2} m_{k+1}, \quad k = 1, \ldots n-2 \\ m_{n-1} = 1 + \frac{n-1}{n} m_{n-2} + \frac{1}{n} m_n \\ \frac{1}{n} m_n = 1 + \frac{1}{n} m_{n-1} \end{cases} \tag{21}$$

$$\begin{cases} m_0 = 0 \\ m_k = m_{k-1} + \frac{n}{k} + \sum_{j=0}^{n-k-2} \frac{n}{j+k+1} \left(\prod_{i=0}^{j} \frac{n-k-i}{(k+i)n}\right) + \frac{n}{n-1} \left(\prod_{i=0}^{n-k-2} \frac{n-k-i}{(k+i)n}\right), \quad k = 1, \ldots, n-2 \\ m_{n-1} = m_{n-2} + \frac{2n}{n-1} \\ m_n = m_{n-1} + n. \end{cases}$$

$$\bar{p}_{00} = 1$$
$$\bar{p}_{kh} = \begin{cases} 1 - \left(1 - \frac{k}{n}\right)^N, & h = k-1 \\ \left(1 - \frac{k}{n}\right)^N, & h = k, \\ 0, & \text{otherwise.} \end{cases} \quad \text{for any } k = 1, \ldots, n-2,$$
$$\bar{p}_{kh} = \begin{cases} 1 - \left(\frac{1}{n}\right)^N, & h = k-1 \\ \left(\frac{1}{n}\right)^N, & h = k+1 \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } k = n-1,$$
$$\bar{p}_{nh} = \begin{cases} 1 - (1-q)^N, & h = n-1 \\ (1-q)^N, & h = n, \\ 0, & \text{otherwise.} \end{cases}$$

For Markov chain (Y_t), define the first hitting time to the state 0 to be $T_Y = \min\{t; Y_t = 0\}$ and its mean $m_k = \mathbb{E}[T_Y, T_Y < \infty \,|\, Y_0 = k]$. It is obvious that $m_k \le m_{S_k}$ for any $0 \le k \le n$.

Consider the first hitting probability $D_k = \mathbb{P}(T_Y < \infty \,|\, Y_0 = k)$ and let $\mathbf{P} = (\bar{p}_{ij})$. From (3), we obtain

$$D_0 = D_1 = \cdots D_n = 1.$$

According to (5), the mean first hitting time satisfies (22), shown at the bottom of the page, from which we can derive, for $k = 1, \ldots, n-2$, the last equation at the bottom of the page.

It is clear from the above equations that if q^{-1} is exponential in n, then m_{n-1} and m_n will also be exponential in n.

Since $m_{S_k} \ge m_k$, we complete the proof of Part (2). □

From the above proposition, we can see that for the $(1+1)$ EA, the order of m_{S_1} is $O(n)$ and m_{S_k} is no more than $O(n^2)$. For the $(N+N)$ EA, $m_{S_{n-1}}$ is greater than $\Omega((n(1-(1-q)^N))^{-1})$ and m_{S_n} is greater than $\Omega((1-(1-q)^N)^{-1})$.

It is not surprising that, if q^{-1} is exponential in n, then when the population starts from S_n, i.e., the point $\{(1\cdots 1)\}^N$, the $(N+N)$ EA will take an exponential time to find the optimal point. Furthermore, for all populations starting from S_{n-1}, i.e., the points such as $\{(01\cdots 1)\}^N$ or $\{(1\cdots 10)\}^N$, the $(N+N)$ EA still takes an exponential time.

C. Impact of Population on the First Hitting Probability

One way to analyze the first hitting probability is to examine the failure probability as defined as follows.

Definition 4: Given a Markov chain $(X_t, t = 0, 1, \ldots)$ and its initial distribution $\mu_0(i) = \mathbb{P}(X_0 = i)$, let the first hitting probability $D_i = \mathbb{P}(T < \infty \,|\, X_0 = i)$, then define $D^{(f)} = 1 - \sum_i \mu_0(i) D_i$ as the *mean failure probability*.

For a Markov chain associated with an EA, we can use the mean failure probability to measure the probability of the EA fails to find an optima. Let $D_1^{(f)}$ be the failure probability of a $(1+1)$ EA and $D_N^{(f)}$ be that of an $(N+N)$ EA for the same problem. It is expected that $D_N^{(f)} \le D_1^{(f)}$ because we could simply run N independent $(1+1)$ EAs simultaneously. The interesting questions are how much smaller $D_N^{(f)}$ could be in comparison with $D_1^{(f)}$ and whether an $(N+N)$ EA would have a smaller failure probability than that produced by N independent $(1+1)$ EAs. We can use the *failure rate* to answer the first question. The failure rate is defined as

$$\text{failure rate} = \frac{D_N^{(f)}}{D_1^{(f)}}.$$

The $(1+1)$ EA considered in this section for objective function (7) uses Mutation I and Selection I with $p = 1$.

There are two $(N+N)$ EAs that we will analyze. The first one, called $(N+N)$ EA-I, is simply N independent $(1+1)$ EAs running simultaneously. The second EA, denoted as $(N+N)$ EA-II, uses Mutation I and Selection III.

Proposition 4: Given objective function (7), we have the following.

1) For the $(1+1)$ EA using Mutation I and Selection I with $p = 1$, let $S_k = \{x; d(x) = k\}, k = 0, \ldots, n$, its first hitting probabilities $D_k = \mathbb{P}(T < \infty \,|\, X_0 \in S_k)$ are given by

$$\begin{cases} D_{S_k} = 1, & k = 0, \ldots, \lambda n - 1 \\ D_{S_k} = \lambda, & k = \lambda n, \\ D_{S_k} = 0, & k = \lambda n + 1, \ldots, n. \end{cases}$$

Assume that the initial individual satisfies the uniform distribution in the space $\{0,1\}^n$. Then the failure probability to find the global optimal solution is

$$D_1^{(f)} = 1 - 2^{-n}\left(\lambda C_n^{\lambda n} + \sum_{k=0}^{\lambda n - 1} C_n^k\right)$$

where $C_n^k = (N!/k!(n-k)!)$ is the binomial coefficient.

2) For the $(N+N)$ EA-I, let $S_k = \{Z; d(Z) = k\}, k = 0, \ldots, n$. Its first hitting probabilities $D_k = \mathbb{P}(T < \infty \,|\, X_0 \in S_k)$ are given by

$$\begin{cases} D_{S_k} = 1, & k = 0, \ldots, \lambda n - 1 \\ D_{S_k} = \lambda, & k = \lambda n \\ D_{S_k} = 0, & k = \lambda n + 1, \ldots, n. \end{cases}$$

$$\begin{cases} m_0 = 0, \\ \left(1 - \left(1 - \frac{k}{n}\right)^N\right) m_k = 1 + \left(1 - \left(1 - \frac{k}{n}\right)^N\right) m_{k-1}, \quad k = 1, \ldots, n-2 \\ m_{n-1} = 1 + \left(1 - \left(\frac{1}{n}\right)^N\right) m_{n-2} + \left(\frac{1}{n}\right)^N m_n \\ (1 - (1-q)^N) m_n = 1 + (1 - (1-q)^N) m_{n-1} \end{cases} \tag{22}$$

$$\begin{cases} m_0 = 0, \\ m_k = \sum_{i=1}^{k} \left(1 - \left(1 - \frac{i}{n}\right)^N\right)^{-1} \\ m_{n-1} = \sum_{i=1}^{n-1} \left(1 - \left(1 - \frac{i}{n}\right)^N\right)^{-1} + \left(1 - \left(\frac{1}{n}\right)^N\right)^{-1} \left(\frac{1}{n}\right)^N (1 - (1-q)^N)^{-1} \\ m_n = (1 - (1-q)^N)^{-1} + m_{n-1}. \end{cases}$$

Assume each individual in the initial population satisfies the uniform distribution in the space $\{0,1\}^n$. The failure probability is

$$D_N^{(f)} = \left(1 - 2^{-n}\lambda C_n^{\lambda n} - 2^{-n}\sum_{k=0}^{\lambda n-1} C_n^k\right)^N.$$

3) For the $(N+N)$ EA-II, let $S_k = \{Z; d(Z) = k\}, k = 0,\ldots,n$. Then its first hitting probabilities $D_k = \mathbb{P}(T < \infty \mid X_0 \in S_k)$ are given by

$$D_{S_k} = 1, \quad k = 0,\ldots,n.$$

The failure probability is

$$D_N^{(f)} = 0$$

regardless of the kind of distributions that an initial individual has.

Proof—Part (1): For the $(1+1)$ EA, the transition probabilities among S_k can be derived as follows.

For any k: $1 \le k < \lambda n$, for $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) = \frac{k}{n}$$
$$\mathbb{P}(X_{t+1} \in S_k \mid X_t = i) = 1 - \frac{k}{n}.$$

For $k = \lambda n$, for $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) = \frac{k}{n},$$
$$\mathbb{P}(X_{t+1} \in S_{k+1} \mid X_t = i) = \frac{n-k}{n}.$$

For any k: $\lambda n + 1 \le k < n$, for $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k+1} \mid X_t = i) = \frac{n-k}{n},$$
$$\mathbb{P}(X_{t+1} \in S_k \mid X_t = i) = 1 - \frac{n-k}{n}.$$

For $k = n$, for $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_k \mid X_t = i) = 1.$$

Introduce an auxiliary homogeneous Markov chain $(Y_t, t = 0,1,\ldots)$ in the state–space $\{0,\ldots,n\}$ whose transition probabilities are given by

$$\begin{aligned}&\bar{p}_{00} = 1\\ &\bar{p}_{kh} = \mathbb{P}(X_{t+1} \in S_k \mid X_t \in S_h)\\ &\quad k = 1,\ldots,n, \quad h = 0,\ldots,n.\end{aligned}$$

For Markov chain (Y_t), let $T_Y = \min\{t; Y_t = 0\}$ and $D_k = \mathbb{P}(T < \infty \mid Y_0 = k)$. It is clear that $D_k = D_{S_k}$ for $k = 1,\ldots,n$.

Let $P = (\bar{p}_{ij})$. According to (3), we have

$$\begin{cases} D_0 = 1, \\ \frac{k}{n}D_{k-1} - \frac{k}{n}D_k = 0, \quad k = 1,\ldots \lambda n - 1 \\ \frac{k}{n}D_{k-1} - D_k + \frac{n-k}{n}D_{k+1} = 0, \quad k = \lambda n \\ \frac{n-k}{n}D_{k+1} - \frac{n-k}{n}D_k = 0, \quad k = \lambda n + 1,\ldots n-1, \\ 0D_n = 0. \end{cases} \tag{23}$$

Since state n is an absorbing state, it is impossible to access the global optimal solution starting from this state. We should let $D_n = 0$ for the equation $0D_n = 0$. Then we can get

$$\begin{cases} D_k = 1, \quad k = 0,\ldots,\lambda n - 1 \\ D_k = \lambda, \quad k = \lambda n \\ D_k = 0, \quad k = \lambda n + 1,\ldots,n. \end{cases}$$

Since $D_k = D_{S_k}$ and the distribution of initial individuals is uniform in the space $\{0,1\}^n$, we obtain

$$D_1^{(f)} = 1 - 2^{-n}\left(\lambda C_n^{\lambda n} + \sum_{k=0}^{\lambda n-1} C_n^k\right).$$

The above equations show that starting from a population in $S_k (k > \lambda n)$, the EA fails to find the global optimal solution.

Part (2) Note that each initial individual is chosen uniformly at random from $\{0,1\}^n$. Hence, we have

$$D_N^{(f)} = \left(D_1^{(f)}\right)^N = \left(1 - 2^{-n}\lambda C_n^{\lambda n} - 2^{-n}\sum_{k=0}^{\lambda n-1} C_n^k\right)^N.$$

Part (3) For the $(N+N)$ EA-II, there exist some positive values a_0 (e.g., let $a_0 = \min\{p_2,\ldots,p_{2N}\}$) and b_0 (e.g., let $b_0 = 1-(1-a_0)^N$) such that the transition probabilities among S_k can be bounded as follows.

When $k = 1,\ldots,n$, for $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) \ge a_0$$
$$\mathbb{P}(X_{t+1} \in S_{k+1} \mid X_t = i) \le b_0.$$

When $k = n$, for $i \in S_k$

$$\mathbb{P}(X_{t+1} \in S_{k-1} \mid X_t = i) \ge a_0.$$

Introduce an auxiliary homogeneous Markov chain $(Y_t, t = 0,1,\ldots)$ in the state–space $\{0,\ldots,n\}$ whose transition probabilities are given by

$$\bar{p}_{00} = 1$$
$$\bar{p}_{kh} = \begin{cases} a_0, & h = k-1 \\ 1-(a_0+b_0), & h = k \\ b_0, & h = k+1 \\ 0, & \text{otherwise} \end{cases} \quad \text{for } k = 1,\ldots,n-1$$
$$\bar{p}_{nh} = \begin{cases} a_0, & h = n-1 \\ 1-a_0, & h = n. \end{cases}$$

For Markov chain (Y_t), let $T_Y = \min\{t; Y_t = 0\}$ and $D_k = \mathbb{P}(T_Y < \infty \mid Y_0 = k)$. We have $D_k \le D_{S_k}$ for $k = 1,\ldots,n$.

According to (3), the first hitting probabilities for Markov chain (Y_t) are given by (24), shown at the bottom of the page, from which we can get

$$D_k = 1, \quad k = 0, \ldots, n.$$

Given an initial distribution μ_0, we have

$$D_N^{(f)} = 1 - \sum_k \mu_0(k) D_k = 0$$

which shows that the $(N+N)$ EA-II can find the global optimal solution from any initial population. This is a much improved result as compared to that in Part (2). □

Discussion: Now we discuss the failure rate of the above $(1+1)$ EA and $(N+N)$ EA-I further. We investigate how the population size changes the failure rate. From Parts (1) and (2) in Proposition 4, we have

$$\text{failure rate} = \left(1 - 2^{-n}\lambda C_n^{\lambda n} - 2^n \sum_{k=0}^{\lambda n - 1} C_n^k\right)^{N-1}.$$

When $\lambda = 1/n$

$$\text{failure rate} = (1 - 2^{-n})^{N-1}.$$

When $\lambda = 1/2$

$$\text{failure rate} \simeq \left(\frac{1}{2}\right)^{N-1}.$$

Fig. 2 shows how the failure rate decreases as the population size increases when $n = 20$. It is clear from Fig. 2 that little improvement could be made when N became large for the objective function with $\lambda = 1/n$. However, a much greater improvement could be made as N was increased for the objective function with $\lambda = 1/2$.

D. Impact of Population on the First Hitting Time

In this section, we discuss the impact of population on the average computation time of an EA. We consider the question whether the mean first hitting time of an $(N+N)$ EA would be shorter than that of a $(1+1)$ EA that uses the same mutation operator and, if it is, how much shorter.

Let $\mathbb{E}[T_1]$ be the mean first hitting time for a $(1+1)$ EA and $\mathbb{E}[T_N]$ be that for a $(N+N)$ EA. The speedup of the $(N+N)$ EA over the $(1+1)$ EA can be defined as

$$\text{speedup} = \frac{\mathbb{E}[T_1]}{N\mathbb{E}[T_N]}.$$

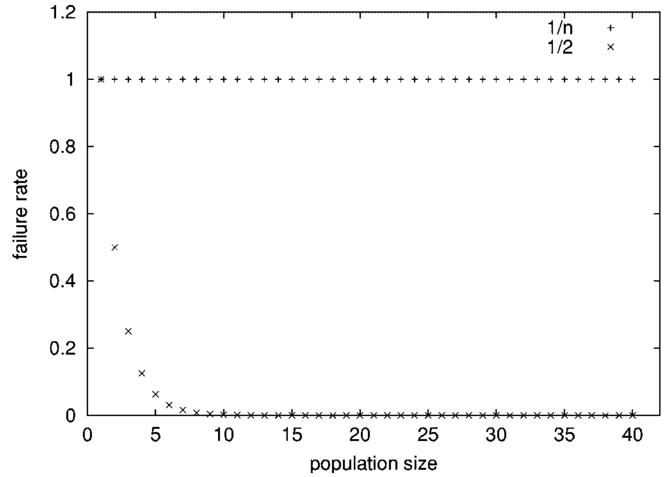

Fig. 2. For the objective function with $\lambda = 1/2$, when N decreases, the failure rate for the $(N+N)$ EA-I to find a global optimal solution in polynomial time decreases very quickly. However, for the objective function (7) with $\lambda = 1/n$, the $(N+N)$ EA-I does not seem to improve much over the $(1+1)$ EA, even for large N.

If we run each individual on one processor, ignoring the communication cost, then the speedup in a parallel computing system is

$$\text{speedup} = \frac{\mathbb{E}[T_1]}{\mathbb{E}[T_N]}.$$

In this paper, we are interested in the later definition of speedup on a hypothetical parallel computer.

Consider the objective function (9). The $(1+1)$ EA for it uses Mutation I and Selection I with $p = 1$ and the $(N+N)$ EA uses Mutation I and the following selection.

Selection V: Select the best $N-1$ individuals in the combined population of X_t and $X_t^{(m)}$ as the next generation X_{t+1}. The best individual is selected twice.

The selection used above is a variant of truncation selection, as often used in evolution strategies.

Proposition 5: Given the objective function (9), we have the following.

1) For the above $(1+1)$ EA, assume the initial population is distributed uniformly at random in $\{0,1\}^n$. Then the mean first hitting time will be

$$\mathbb{E}[T_1] = n 2^{-n} \sum_{k=0}^{n} C_n^k \left(\sum_{j=1}^{k} j^{-1}\right).$$

2) For the above $(N+N)$ EA, also assume each individual in the initial population is distributed uniformly at random

$$\begin{cases} D_0 = 1, \\ a_0 D_{k-1} - (a_0 + b_0) D_k + b_0 D_{k+1} = 0, \quad k = 1, \ldots, n-1 \\ a_0 D_{n-1} - a_0 D_n = 0 \end{cases} \tag{24}$$

in $\{0,1\}^n$. Let $S_k = \{Z; d(Z) = k\}; k = 0, \dots, n$. Then the mean first hitting time will be

$$\mathbb{E}[T_N] \le \sum_{k=0}^{n} \mu_0(S_k) \left(n \sum_{j=1}^{k} j^{-1} \right)$$

where $\mu_0(S_0) = 1 - (1 - 2^{-n})^N$, and for $k = 1, \dots, n$

$$\mu_0(S_k) = \left(1 - \left(1 - C_n^k \Big/ \left(\sum_{i=k}^{n} C_n^i\right)\right)^N\right) \times \left(1 - \sum_{j=0}^{k-1} \mu_0(S_j)\right).$$

3) The speedup of the $(N+N)$ EA over the $(1+1)$ EA is given by

$$\text{speedup} \ge \frac{2^{-n} \sum_{k=0}^{n} C_n^k \left(\sum_{j=1}^{k-1} j^{-1}\right)}{\sum_{k=0}^{n} \mu_0(S_k) \left(\sum_{j=1}^{k} j^{-1}\right)}.$$

Proof—Part (1): Let $S_k = \{x; d(x) = k\}, k = 0, \dots, n$. For the $(1+1)$ EA, we have

$$\mu_0(S_k) = 2^{-n} C_n^k, \quad k = 0, \dots, n$$

since the initial population is distributed uniformly at random.

The mean first hitting time starting from an individual in S_k is

$$m_{S_k} = \sum_{j=1}^{k} n/j.$$

Hence, the mean first hitting time is

$$\mathbb{E}[T] = \sum_{k=0}^{n} \mu_0(S_k) m_{S_k} = n 2^{-n} \sum_{k=0}^{n} C_n^k \left(\sum_{j=1}^{k} j^{-1}\right).$$

Part (2) For the $(N+N)$ EA, since individuals in the initial population are distributed uniformly at random, we have for $k = 0, \dots, n$

$$\begin{aligned}\mu_0(S_k) &= \mathbb{P}(X_0 \in S_k) \\ &= \mathbb{P}((\exists x \in X_0 : x \in S_k) \quad \text{and} \\ &\quad \times (\forall x \in X_0 : x \notin S_0, \dots S_{k-1})) \\ &= \left(1 - \left(1 - C_n^k \Big/ \left(\sum_{i=k}^{n} C_n^i\right)\right)^N\right) \\ &\quad \times \left(1 - \sum_{j=0}^{k-1} \mu_0(S_j)\right).\end{aligned}$$

where $\mu_0(S_0) = 1 - (1 - 2^{-n})^N$.

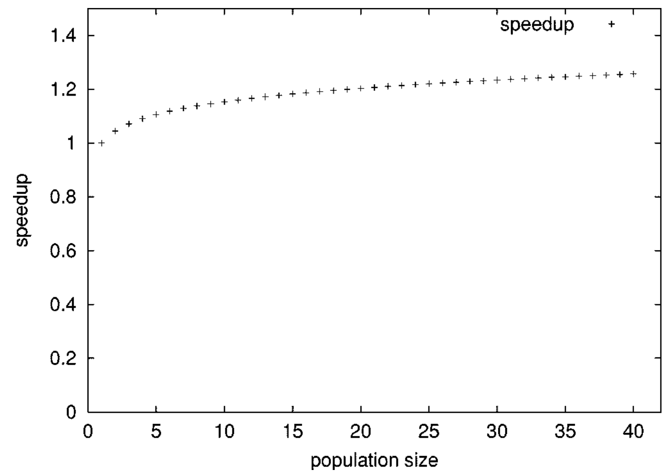

Fig. 3. Speedup $\mathbb{E}[T_1]/\mathbb{E}[T_N]$ of the $(N+N)$ EA using Mutation I and Selection V over the $(1+1)$ EA using Mutation I and Selection I on the objective function (9) when the initial population is distributed uniformly at random.

The mean first hitting time for the $(N+N)$ EA starting from a population in S_k is

$$m_{S_k} \le \sum_{j=1}^{k} n/j.$$

Hence, the mean first hitting time will be

$$\mathbb{E}[T] = \sum_{k=0}^{n} \mu_0(S_k) \left(n \sum_{j=1}^{k} j^{-1} \right).$$

Part (3) It is a direct result of Parts (1) and (2). □

Fig. 3 shows such a speedup function of N when $n = 20$.

However, the estimation on the speedup, more precisely on $\mathbb{E}[T_N]$, given above is not very tight. Further work is needed to derive tighter bounds.

Now we examine another distribution of the initial population for the above EAs and problem. Assume the initial individual x (both for the $(1+1)$ EA and $(N+N)$ EA) takes $(1 \cdots 1)$, i.e., the EA starts from the worst state.

The analysis in the above proposition is not suitable for this case. If we use the above analysis, we would have

$$\mathbb{E}[T_1] = \sum_{k=0}^{n} \mu_0(S_k) m_{S_k} = \mu_0(S_n) m_{S_n} = n \sum_{j=1}^{n} j^{-1}$$

and

$$\mathbb{E}[T_N] \le \sum_{k=0}^{n} \mu_0(S_k) m_{S_k} = \mu_0(S_n) m_{S_n} = n \sum_{j=1}^{n} j^{-1}.$$

Hence

$$\text{speedup} = \frac{\mathbb{E}[T_1]}{\mathbb{E}[T_N]} \ge 1.$$

There would be no speedup at all. A more accurate estimation on $\mathbb{E}[T_N]$ is needed. The following proposition represents our first attempt.

Proposition 6: Given the objective function (9), we have the following.

1) For the aforementioned $(1+1)$ EA, if the initial population is $(1\cdots 1)$, then the mean first hitting time is

$$\mathbb{E}[T_1] = \sum_{i=1}^{n} n/i.$$

2) For the aforementioned $(N+N)$ EA, if all individuals in the initial population are $(1\cdots 1)$, then the mean first hitting time is

$$\mathbb{E}[T_N] \le m_n$$

where m_n is given by, for $l = 1,\ldots N-1$ and $k = 2,\ldots,n-1$, as in the first equation shown at the bottom of the page.

3) The speedup of the $(N+N)$ EA over the $(1+1)$ EA is given by

$$\text{speedup} \ge \frac{n\sum_{j=1}^{k-1} j^{-1}}{m_n}.$$

Proof: The proof of Part (1) is the same as that for Proposition 5.

Part (2) Let $S_0 = \{Z; d(Z) = 0\}, S_{kl} = \{Z; d(Z) = k,$ and the number of individual x with $d(x) = k$ is $l\}$, for $k = 1,\ldots,n-1$, and $l = 1,\ldots,N$, and $S_n = \{Z; d(Z) = n\}$.

According to the given $(N+N)$ EA, the transition probabilities among S_{kl} can be derived as follows.

When $k = 1$, for $l = 1,\ldots,N$ and $i \in S_{1l}$

$$\mathbb{P}(X_{t+1} \in S_0 \,|\, X_t = i) \ge 1 - \left(1 - \frac{1}{n}\right)^l$$
$$\mathbb{P}(X_{t+1} \in S_{1h} \,|\, X_t = i) = 0, \quad h \le l$$
$$\mathbb{P}(X_{t+1} \in S_{1(l+1)} \,|\, X_t = i) \le \left(1 - \frac{1}{n}\right)^l$$
$$\mathbb{P}(X_{t+1} \in S_{1h} \,|\, X_t = i) \ge 0, \quad h > l+1$$
$$\mathbb{P}(X_{t+1} \in S_{mh} \,|\, X_t = i) = 0, \quad m > 1, \quad h = 1,\ldots,N.$$

When $k = 2,\ldots,n-1$, for $l = 1,\ldots,N$ and $i \in S_{kl}$

$$\mathbb{P}(X_{t+1} \in S_{mh} \,|\, X_t = i) = 0, \quad m < k-1, \quad h = 1,\ldots,N$$
$$\mathbb{P}(X_{t+1} \in S_{k-1h} \,|\, X_t = i) = \begin{cases} \ge C_l^h \left(\frac{k}{n}\right)^h \left(1 - \frac{k}{n}\right)^{l-h}, & h \le l \\ = 0, & \text{otherwise} \end{cases}$$
$$\mathbb{P}(X_{t+1} \in S_{kh} \,|\, X_t = i) = \begin{cases} = 0, & h <= l, \\ \le 1 - \sum_{h=1}^{l} C_l^h \left(\frac{k}{n}\right)^h \left(1 - \frac{k}{n}\right)^{l-h}, & h = l+1 \\ \ge 0, & h > l+1 \end{cases}$$
$$\mathbb{P}(X_{t+1} \in S_{ml} \,|\, X_t = i) = 0, \quad m > k, \quad h = 1,\ldots,N.$$

When $k = n$, for any $i \in S_n$

$$\mathbb{P}(X_{t+1} \in S_{(n-1)N} \,|\, X_t = i) = 1.$$

Introduce an auxiliary Markov chain (Y_t) whose state space is $\{0, 11, \ldots, 1N, 21, \ldots, 2N, \ldots, (n-1)1, \ldots, (n-1)N, n\}$ and transition probabilities $\bar{p}_{k,h}$ are given by $\bar{p}_{0,0} = 1$ for $k = 1$ and $l = 1,\ldots,N$

$$\bar{p}_{1l,0} = 1 - \left(1 - \frac{1}{n}\right)^l$$

and

$$\bar{p}_{1l,mh} = \begin{cases} \left(1 - \frac{1}{n}\right)^l, & m = k, \quad h = l+1 \\ 0, & \text{else} \end{cases}$$

for $k = 2,\ldots,n-1$ and $l = 1,\ldots,N$, see the last equation at the bottom of the page.

For Markov chain (Y_t), define the first hitting time to state 0 to be $T_Y = \min\{t; Y_t = 0\}$ and its mean $m_{kl} = \mathbb{E}[T_Y, T_Y < \infty \,|\, Y_0 = kl]$. For Markov chain (X_t), define the mean first hitting time $m_{S_{kl}} = \mathbb{E}[T, T < \infty \,|\, X_0 \in S_{kl}]$, where $k = 1,\ldots,n, l = 1,\ldots,N$. We have $m_{kl} \ge m_{S_{kl}}$.

Let $D_{kl} = \mathbb{P}(T_Y < \infty \,|\, Y_0 = kl)$. According to (3), the first hitting probabilities of Markov chain (Y_t), for $l = 1,\ldots,N-1$,

$$\begin{cases} m_0 = 0 \\ m_{1N} = \left(1 - \left(1 - \frac{1}{n}\right)^N\right)^{-1} \\ m_{1l} = 1 + \left(1 - \frac{1}{n}\right)^l m_{1(l+1)} \\ m_{kN} = \left(\sum_{h=1}^{N} C_N^h \left(\frac{k}{n}\right)^h \left(1 - \frac{k}{n}\right)^{N-h}\right)^{-1} \left(1 + \sum_{h=1}^{N} C_N^h \left(\frac{k}{n}\right)^h \left(1 - \frac{k}{n}\right)^{N-h} m_{(k-1)h}\right) \\ m_{kl} = 1 + \sum_{h=1}^{l} C_l^h \left(\frac{k}{n}\right)^h \left(1 - \frac{k}{n}\right)^{l-h} m_{(k-1)h} + \left(1 - \sum_{h=1}^{l} C_l^h \left(\frac{k}{n}\right)^h \left(1 - \frac{k}{n}\right)^{l-h}\right) m_{k(l+1)} \\ m_n = 1 + m_{(n-1)N}. \end{cases}$$

$$\bar{p}_{kl,mh} = \begin{cases} C_l^h \left(\frac{k}{n}\right)^h \left(1 - \frac{k}{n}\right)^{l-h}, & m = k-1, \quad h \le l \\ 1 - \sum_{h=1}^{l} C_l^h \left(\frac{k}{n}\right)^h \left(1 - \frac{k}{n}\right)^{l-h}, & m = k, \quad h = l+1, \\ 0, & \text{else} \end{cases}$$

and

$$\bar{p}_{n,(n-1)N} = 1.$$

and $k=2,\dots,n-1$, are given by (25), shown at the bottom of the page, from which we can get

$$D_0 = D_n = D_{kl} = 1, \quad k=1,\dots,n-1.$$

According to (5), we know that the mean first hitting time satisfies, for $l=1,\dots,N-1$ and $k=2,\dots,n-1$, (26), shown at the bottom of the page.

The above equations can be simplified, for $l=1,\dots,N$ and $k=2,\dots,n-1$:, as shown in the last equation at the bottom of the page.

Hence

$$\mathbb{E}[T_N] = \sum_{k=0}^{n} \mu_0(S_k) m_{S_k} = m_{S_n} \le m_n.$$

Part (3) This is a direct consequence of Parts (1) and (2). □

Fig. 4 shows such a speedup function of N when $n=20$, based on the above estimation.

Consider another more difficult example given by the objective function (8). The $(1+1)$ EA uses the following mutation and Selection I with $p=1$.

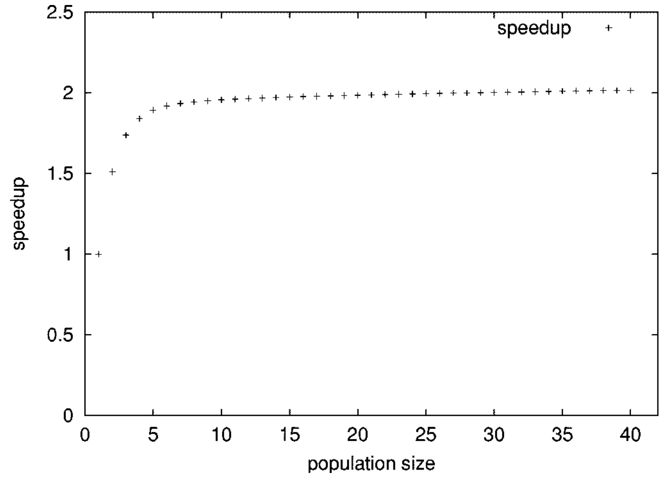

Fig. 4. Speedup $\mathbb{E}[T_1]/\mathbb{E}[T_N]$ of the $(N+N)$ EA using Mutation I and Selection V over the $(1+1)$ EA using Mutation I and Selection I on the objective function (9) when the initial population is taken from S_n.

Mutation II: Given population (only a single individual in the case of $(1+1)$ EA) X_t at generation t, for each individual

$$\begin{cases} D_0 = 1, \\ \left(1-\left(1-\frac{1}{n}\right)^N\right) D_0 - \left(1-\left(1-\frac{1}{n}\right)^N\right) D_{1,N} = 0 \\ \left(1-\left(1-\frac{1}{n}\right)^l\right) D_0 - D_{1l} + \left(1-\frac{1}{n}\right)^l D_{1(l+1)} = 0, \\ \sum_{h=1}^{N} C_N^h \left(\frac{k}{n}\right)^h \left(1-\frac{k}{n}\right)^{N-h} D_{(k-1)h} - \left(\sum_{h=1}^{N} C_N^h \left(\frac{k}{n}\right)^h \left(1-\frac{k}{n}\right)^{N-h}\right) D_{kN} = 0 \\ \sum_{h=1}^{l} C_l^h \left(\frac{k}{n}\right)^h \left(1-\frac{k}{n}\right)^{l-h} D_{(k-1)h} - D_{k,l} + \left(\sum_{h=1}^{l} C_l^h \left(\frac{k}{n}\right)^h \left(1-\frac{k}{n}\right)^{l-h}\right) D_{k(l+1)} = 0, \\ D_{(n-1)N} - D_n = 0 \end{cases} \tag{25}$$

$$\begin{cases} m_0 = 0, \\ \left(1-\left(1-\frac{1}{n}\right)^N\right) m_{1N} = 1, \\ m_{1l} = 1 + \left(1-\frac{1}{n}\right)^l m_{1(l+1)} \\ \left(\sum_{h=1}^{N} C_N^h \left(\frac{k}{n}\right)^h \left(1-\frac{k}{n}\right)^{N-h}\right) m_{kN} = 1 + \sum_{h=1}^{N} C_N^h \left(\frac{k}{n}\right)^h \left(1-\frac{k}{n}\right)^{N-h} m_{(k-1)h} \\ m_{kl} = 1 + \sum_{h=1}^{l} C_l^h \left(\frac{k}{n}\right)^h \left(1-\frac{k}{n}\right)^{l-h} m_{(k-1)h} \\ \qquad + \left(1-\sum_{h=1}^{l} C_l^h \left(\frac{k}{n}\right)^h \left(1-\frac{k}{n}\right)^{l-h}\right) m_{k(l+1)} \\ m_n = 1 + m_{(n-1)N}. \end{cases} \tag{26}$$

$$\begin{cases} m_0 = 0, \\ m_{1N} = \left(1-\left(1-\frac{1}{n}\right)^N\right)^{-1} \\ m_{1l} = 1 + \left(1-\frac{1}{n}\right)^l m_{1(l+1)} \\ m_{kN} = \left(\sum_{h=1}^{N} C_N^h \left(\frac{k}{n}\right)^h \left(1-\frac{k}{n}\right)^{N-h}\right)^{-1} \left(1 + \sum_{h=1}^{N} C_N^h \left(\frac{k}{n}\right)^h \left(1-\frac{k}{n}\right)^{N-h} m_{(k-1)h}\right) \\ m_{kl} = 1 + \sum_{h=1}^{l} C_l^h \left(\frac{k}{n}\right)^h \left(1-\frac{k}{n}\right)^{l-h} m_{(k-1)h} \\ \qquad + \left(1-\sum_{h=1}^{l} C_l^h \left(\frac{k}{n}\right)^h \left(1-\frac{k}{n}\right)^{l-h}\right) m_{k(l+1)} \\ m_n = 1 + m_{(n-1)N}. \end{cases}$$

$(s_1 \cdots s_n)$, flip each of its bits to its complement with probability (c/n), where $c \in (0, n)$. The mutated population is denoted as $X_t^{(m)}$.

The corresponding $(N+N)$ EA uses Mutation II and Selection V. For convenience, we only consider the case of the initial population starting from one point and do not consider the case of the initial population taking a random distribution here.

Proposition 7: Given the objective function (8), we have the following.

1) For the $(1+1)$ EA using Mutation II and Selection I with $p=1$, if the initial population satisfies
$$\mu_0(X_0 = (1 \cdots 1)) = 1$$
then the mean first-hitting time of the $(1+1)$ EA will be
$$\mathbb{E}[T_1] = \left(\frac{c}{n}\right)^{-n}.$$

2) For the $(N+N)$ EA using Mutation II and Selection V, if all individuals in the initial population are $(1 \cdots 1)$, then the mean first-hitting time of the $(N+N)$ EA will be
$$\mathbb{E}[T_N] = \left(1 - \left(1 - \left(\frac{c}{n}\right)^n\right)^N\right)^{-1}.$$

3) The speedup of the $(N+N)$ EA over the $(1+1)$ EA is
$$\text{speedup} = \frac{1 - \left(1 - \left(\frac{c}{n}\right)^n\right)^N}{\left(\frac{c}{n}\right)^n}.$$

Proof—Part (1): Let $S_k = \{x; d(x) = k\}, k = 0, \ldots, n$. We have
$$\mathbb{E}[T_1] = \sum_{k=0}^{n} \mu_0(S_k) m_{S_k} = m_{S_n}$$
and
$$m_{S_n} = \left(\frac{c}{n}\right)^{-n}.$$

Part (2) Let $S_k = \{Z; d(Z) = k\}, k = 0, \ldots, n$. Then
$$\mathbb{E}[T_N] = \sum_{k=0}^{n} \mu_0(S_k) m_{S_k} = m_{S_n}$$
and
$$m_{S_n} = \left(1 - \left(1 - \left(\frac{c}{n}\right)^n\right)^N\right)^{-1}.$$

Part (3) It is a direct consequence of Parts (1) and (2).

Fig. 5 shows such a speedup function of N when $n = 20$ and $c = 1$. It is clear that the speedup increases linearly as N increases.

IV. Conclusion

Many EAs use more than one individual in the population. It has also been argued that one of the key characteristics of EAs is their populations. However, rigorous theoretical results are few regarding the real benefits of populations in EAs. This paper provides a number of results that show when a population may bring benefits to an EA in terms of lower time complexity, higher first hitting probabilities, and shorter first hitting time.

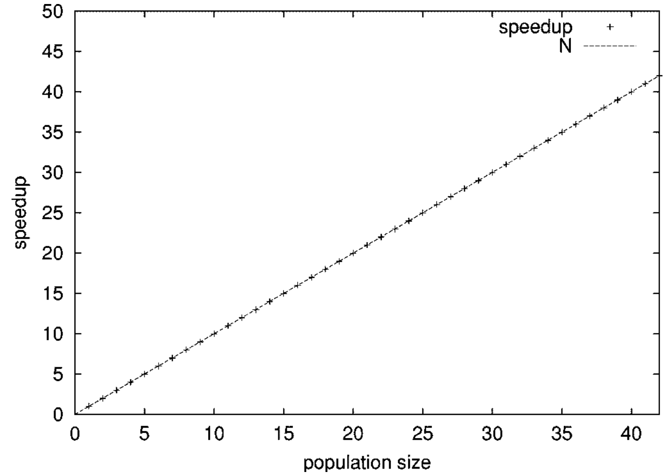

Fig. 5. Speedup $\mathbb{E}[T_1]/\mathbb{E}[T_N]$ of the $(N+N)$ EA using Mutation II and Selection V over the $(1+1)$ EA using Mutation II and Selection I on the objective function (8).

It is shown that a population-based $(N+N)$ EA $(N > 1)$ may take only average polynomial time to solve a problem that would take a $(1+1)$ EA average exponential time to solve, given the same mutation operator in both algorithms. It is also shown that the introduction of a population into an EA can increase the first hitting probability. Given a distribution of initial individuals in an EA, e.g., a uniform distribution, we are able to derive the mean first hitting time of the algorithms. Such analysis enables us to compare the mean first hitting times of the $(1+1)$ and $(N+N)$ EAs under the same initial distribution and show that a population can shorten the mean first hitting time. Our results also represent one of the first attempts toward analysing the average case time complexity of $(N+N)$ EAs $(N > 1)$.

There is much work to be done in the theoretical analysis of population-based EAs. The discussions here are restricted to some simple objective functions and EAs. This paper considers $(N+N)$ EAs with only mutation and selection in order to have the population-based EAs as close to the $(1+1)$ EA as possible so that the impact of a population can be isolated and studied. Different selection schemes in $(N+N)$ EAs are studied and shown to have different impact on EA's performance (in terms of complexity). Our future work includes two major directions. One is to study the impact of recombination and (μ, λ) strategies on EA's computation time. The other is to carry out similar theoretical analysis for other combinatorial optimization problems, especially those often discussed in the classical combinatorial optimization field, e.g., maximum matching and other problems [26], [27].

References

[1] X. Yao, "An overview of evolutionary computation," *Chinese J. Advanced Software Research*, vol. 3, no. 1, pp. 12–29, 1996.

[2] X. Yao, Ed., *Evolutionary Computation: Theory and Applications*, Singapore: World Scientific, 1999.

[3] E. Falkenauer, *Genetic Algorithms and Grouping Problems*, Chichester, U.K.: Wiley, 1998.

[4] K.-H. Liang, X. Yao, C. Newton, and D. Hoffman, A new evolutionary approach to cutting stock problems with and without contiguity, in *Comput. Oper. Res.*, vol. 29, no. 12, pp. 1641–1659, 2002.

[5] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. New York: Dover, 1998.

[6] G. Rudolph, "Finite Markov chain results in evolutionary computation: A tour d'Horizon," *Fundamenta Informaticae*, vol. 35, no. 1–4, pp. 67–89, 1998.

[7] A. E. Eiben and G. Rudolph, "Theory of evolutionary algorithms: A bird eye view," *Theoret. Comput. Sci.*, vol. 229, pp. 3–9, Nov. 1999.

[8] J. He and X. Yao, "Drift analysis and average time complexity of evolutionary algorithms," *Artific. Intell.*, vol. 127, pp. 57–85, Mar. 2001.

[9] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, 1997.

[10] S. Droste, T. Jansen, and I. Wegener, On the analysis of the $(1+1)$ evolutionary algorithms, in Theoretical Computer Science, 2001. to appear.

[11] T. Bäck, "The interaction of mutation rate, selection and self-adaptation within a genetic algorithm," in *Parallel Problem Solving from Nature (PPSN) II*, R. Männer and B. Manderick, Eds. Amsterdam, The Netherlands: North Holland, 1992, pp. 85–94.

[12] H. Mühlenbein, "How genetic algorithms really works I: Mutation and hill-climbing," in *Parallel Problem Solving from Nature (PPSN) II*, R. Männer and B. Manderick, Eds. Amsterdam, : North Holland, 1992, pp. 15–25.

[13] G. Rudolph, *Convergence Properties of Evolutionary Algorithms*. Hamburg, Germany: Springer-Verlag, 1997.

[14] J. Garnier, L. Kallel, and M. Schoenauer, "Rigorous hitting times for binary mutations," *Evol. Comput.*, vol. 7, no. 2, pp. 167–203, 1999.

[15] S. Droste, T. Jansen, and I. Wegener, "A rigorous complexity analysis of the $(1+1)$ evolutionary algorithm for linear functions with Boolean inputs," *Evol. Comput.*, vol. 6, no. 2, pp. 185–196, 1998.

[16] G. Rudolph, "How mutation and selection solve long path problems in polynomial expected time," *Evol. Comput.*, vol. 4, no. 2, pp. 207–211, 1996.

[17] J. Garnier and L. Kallel, "Statistical distribution of the convergence time of evolutionary algorithms for longpath problems," *IEEE Trans. Evol. Comput.*, vol. 4, pp. 16–30, Apr. 2000.

[18] R. Syski, *Passage Times for Markov Chains*. Amsterdam, The Netherlands: IOS Press, 1992.

[19] J. He and L. Kang, "On the convergence rate of genetic algorithms," *Theoret. Comput. Sci.*, vol. 229, pp. 23–39, Nov. 1999.

[20] E. Cantú-Paz, "Markov chain of parallel genetic algorithms," *IEEE Trans. Evol. Comput.*, vol. 4, pp. 216–226, Sept. 2000.

[21] B. Naudts and L. Kallel, "A comparison of predictive measure of problem difficulty in evolutionary algorithms," *IEEE Trans. Evol. Comput.*, vol. 4, pp. 1–15, Apr. 2000.

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

[23] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[24] X. Yao, "Simulated annealing with extended neighborhood," *Int. J. Comput. Math.*, vol. 40, pp. 169–189, 1991.

[25] ——, "A new simulated annealing algorithm," *Int. J. Comput. Math.*, vol. 56, pp. 161–168, 1995.

[26] G. H. Sasaki and B. Hajek, "The time complexity of maximum matching by simulated annealing," *J. ACM*, vol. 35, pp. 387–403, 1988.

[27] A. H. Wright, R. K. Thompson, and J. Zhang, "The computational complexity of N-K fitness functions," *IEEE Trans. Evol. Comput.*, pp. 373–379, Nov. 2000.

Jun He received the M.Sc. degree in mathematics and the Ph.D. degree in computer science from Wuhan University, Wuhan, China, in 1992 and 1995, respectively.

He is currently a Research Fellow with the School of Computer Science, University of Birmingham, Birmingham, U.K. His research interests include evolutionary computation, network security and parallel algorithms.

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

He is currently a Professor of Computer Science at the University of Birmingham, Birmingham, U.K., and a Guest Professor at USTC. He was a Lecturer, Senior Lecturer, and Associate Professor at University College, University of New South Wales, the Australian Defence Force Academy (ADFA), Canberra, Australia, between 1992 and 1999. He held post-doctoral fellowships from the Australian National University (ANU), Canberra, and the Commonwealth Scientific and Industrial Research Organization (CSIRO), Melbourne, between 1990 and 1992. His major research interests include evolutionary computation, neural network ensembles, coevolution, evolvable hardware, time complexity of evolutionary algorithms, and data mining. He has published extensively in these areas.

Dr. Yao is an Associate Editor or Member of the Editorial Boards of six international journals, including IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, and a Guest Editor or co-editor of Special Isue journals. He has chaired/co-chaired over 20 international conferences in evolutionary computation and computational intelligence, including CEC'02, ICCIMA'01, PPSN VI'00, IEEE ECNN'00, and CEC'99. He is the Chair of the IEEE NNS Technical Committee on Evolutionary Computation, and the Past President of the Evolutionary Programming Society. He received the IEEE Donald G. Fink Prize Paper Award in 2001.