Natural Computing **3:** 21–35, 2004.

A study of drift analysis for estimating computation time of evolutionary algorithms

JUN HE and XIN YAO
The Centre of Excellence for Research in Computational Intelligence and Applications (CERCIA), School of Computer Science, the University of Birmingham, Edgbaston, Birmingham B15 2TT, UK (E-mail: x.yao@cs.bham.ac.uk)

Abstract. This paper introduces drift analysis and its applications in estimating average computation time of evolutionary algorithms. Firstly, drift conditions for estimating upper and lower bounds of the mean first hitting times of evolutionary algorithms are presented. Then drift analysis is applied to two specific evolutionary algorithms and problems. Finally, a general classification of easy and hard problems for evolutionary algorithms is given based on the analysis.

Key words: algorithm analysis, combinatorial optimisation, evolutionary computation, meta-heuristics

1. Introduction

An important topic in the theory of evolutionary algorithms (EAs) is their computation time for solving combinatorial optimisation problems, which reveals the number of expected generations needed to reach an optimal solution (Rudolph, 1998; Eiben and Rudolph, 1999). In the last few years, some progresses have been made towards this direction (Droste et al., 1998; Beyer et al., 2002; Droste et al., 2002; He and Yao, 2002). However, most of the tools used in the analysis were somewhat ad hoc. It is important to develop mathematical models and tools for analysing EAs so that insights can be gained into them.

This paper introduces a technique for estimating the computation time of EAs – drift analysis. Drift analysis draws properties of a stochastic process from its mean drift. It has been used to study properties of the general Markov chain (Hajek, 1982; Meyn and Tweedie, 1993) and to estimate the time complexity of simulated annealing algorithms (Sasaki and Hajek, 1988). It has also been applied to the analysis of EAs (He and Yao, 2001, 2003).

The rest of this paper is organised as follows. Section 2 describes mathematical models of EAs. Section 3 discusses drift conditions which are used to estimate upper and lower bounds of the first hitting times. Two examples in Section 4 illustrate the application of drift analysis to specific EAs and problems. One important outcome of our analysis is a new classification of easy and hard problems for EAs, described in Section 5. Finally, Section 6 concludes the paper with a short summary.

2. Mathematical models of EAs

Given an objective function $f : S \to R$, where S is a finite set and R is the real line, a maximisation problem is to find an $x \in S$ such that

$$f(x) = \max\{f(y), y \in S\}.$$

EAs are often used to solve this kind of problems. Denote $\mathbf{x}$ to be a population of individuals, E the set consisting of all populations. Let ξ_t be the t-th generation population, which is a random variable and takes values from E. Given an initial population ξ_0 and let $t = 0$, most EAs can be described by the following three major steps.

Recombination: Individuals in population ξ_t are recombined. An offspring population $\xi_t^{(c)}$ is then obtained.

Mutation: Individuals in population $\xi_t^{(c)}$ are mutated. An offspring population $\xi_t^{(m)}$ is then obtained.

Selection: Each individual in the population $\xi_t^{(m)}$ and original population ξ_t is assigned a survival probability. Then select some individuals as the next generation ξ_{t+1}.

There have been a few mathematical models proposed for EAs. Here we introduce two of them, i.e., Markov chains and supermartingale, because they seem to be most appropriate for estimating EA's computation time.

Markov chains are widely used mathematical models in the theoretical analysis of EAs. The sequence of random variables $\{\xi_t; t = 0, 1, 2, \cdots\}$ can be modelled by a Markov chain (Rudolph, 1998), because the state of the $(t+1)$-th generation often depends only on the t-th generation. The transition probability $\mathbf{P}(\mathbf{x}, \mathbf{y}; t)$ is given by, for any populations $\mathbf{x}, \mathbf{y} \in E$,

$$\mathbf{P}(\mathbf{x}, \mathbf{y}; t) := \mathbf{P}(\xi_{t+1} = \mathbf{y} \mid \xi_t = \mathbf{x}).$$

An EA with elitist selection strategy can be modelled by an absorbing Markov chain.

An alternative way to model EAs is by a supermartingale. This model was first used in discussing the convergence of non-elitist selection strategies (Rudolph, 1994). It is not as commonly used as Markov chains.

Denote E_{opt} to be the set of populations which include an optimal solution. In order to describe how far a population is away from E_{opt}, we need a function $V(\mathbf{x}, E_{opt})$ to measure the distance between a population $\mathbf{x}$ and the optimal set E_{opt}. This *distance function*, without confusion, is denoted as $V(\mathbf{x})$ in short. The distance function can be defined in differents ways, e.g., Hamming distance, or $V(\mathbf{x}) = \min\{| f(\mathbf{x}) - f(\mathbf{y}) |; \mathbf{y} \in E_{opt}\}$, or $V(\mathbf{x}) = 0$ if $\mathbf{x} \in E_{opt}$ and $V(\mathbf{x}) = 1$ if $\mathbf{x} \notin E_{opt}$.

Given a distance function, now we can define the one-step *mean drift* at the t-th generation (assume ξ_t takes the value of $\mathbf{x}$),

$$\mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = \mathbf{x}] \;:=\; V(\mathbf{x}) - \sum_{\mathbf{y} \in E} \mathbf{P}(\mathbf{x}, \mathbf{y}; t) V(\mathbf{y}). \qquad (1)$$

The drift can be decomposed into two parts: positive and negative drifts,

$$\mathbf{E}^{+}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = \mathbf{x}] := V(\mathbf{x}) - \sum_{\{\mathbf{y}: V(\mathbf{y}) < V(\mathbf{x})\}} \mathbf{P}(\mathbf{x}, \mathbf{y}) V(\mathbf{y}),$$

$$\mathbf{E}^{-}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = \mathbf{x}] := V(\mathbf{x}) - \sum_{\{\mathbf{y}: V(\mathbf{y}) > V(\mathbf{x})\}} \mathbf{P}(\mathbf{x}, \mathbf{y}) V(\mathbf{y}).$$

The one-step mean drift is similar to the local performance of EAs (Beyer, 2001), where the positive drift is the rate of the gain of a population towards the optimum and the negative drift is that away from the optimum.

Using the mean drift, we can analyse the convergence, convergence rate and first hitting time of a stochastic process (Hajek, 1982; Meyn and Tweedie, 1993). If the one-step mean drift is always not less than 0, i.e.,

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

then $\{V(\xi_t)\}$ is a supermartingale (Chow and Teicher, 1988).

3. Drift analysis for mean first hitting times

Let $\{\xi_t; t = 0, 1, \cdots\}$ be a Markov chain associated with an EA. Its first hitting time to the optimal set E_{opt}, or the number of generations for the EA to find an optimal solution first time, is defined by

$$\tau := \min\{t \geq 0; \xi_t \in E_{opt}\}.$$

The upper bound of the first hitting time can be estimated through the lower bound of the mean drift (He and Yao, 2001).

LEMMA 1. *Given a distance function $V(\mathbf{x})$, if $\{V(\xi_t); t = 0, 1, 2, \cdots\}$ satisfies: for any time $t \geq 0$ and any population ξ_t with $V(\xi_t) > 0$,*

$$\mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] \geq c_{low}, \tag{2}$$

where $c_{low} > 0$, *then the mean first hitting time satisfies*

$$\mathbf{E}[\tau \mid \xi_0] \leq \frac{V(\xi_0)}{c_{low}}. \tag{3}$$

Similarly, the lower bound of the first hitting time can be estimated by the upper bound of the one-step mean drift (He and Yao, 2001).

LEMMA 2. *Given a distance function* $V(\mathbf{x})$, *if* $\{V(\xi_t); t = 0, 1, 2, \cdots\}$ *satisfies: for any time* $t \geq 0$ *and any population* ξ_t *with* $V(\xi_t) > 0$,

$$\mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] \leq c_{up}, \tag{4}$$

where $c_{up} > 0$, *then the mean first hitting time satisfies*

$$\mathbf{E}[\tau \mid \xi_0] \geq \frac{V(\xi_0)}{c_{up}}. \tag{5}$$

A natural question here is whether there exists such kind of distance function $V(\mathbf{x})$, upper bound c_{up} and lower bound c_{low} in the previous two lemmas. According to the following lemma (Theorem 10, Chapter II §3.1.1 in Syski (1992)), they do exist if an EA could be modelled by a homogeneous absorbing Markov chain.

LEMMA 3. *Assume* $\{\xi_t\}$ *is a homogeneous absorbing Markov chain. If the distance function* $V(\mathbf{x}) = \mathbf{E}[\tau \mid \xi_0 = \mathbf{x}]$, *then* $V(\mathbf{x})$ *satisfies*

$$\begin{cases} V(\mathbf{x}) = 0, & \text{if } \mathbf{x} \in E_{opt}, \\ V(\mathbf{x}) - \sum_{\mathbf{y} \in E} \mathbf{P}(\mathbf{x}, \mathbf{y}) V(\mathbf{y}) = 1, & \text{if } \mathbf{x} \notin E_{opt}. \end{cases} \tag{6}$$

Based on this lemma, we know that Lemmas 1 and 2 hold when $c_{low} = c_{up} = 1$.

4. Drift analysis for selected evolutionary algorithms

4.1. *Analysis of a* $(1 + 1)$*-EA for linear functions*

Consider the following linear function (Droste et al., 1998):

$$f(x) = w_0 + \sum_{i=1}^{n} w_i s_i, \tag{7}$$

where $x = (s_1 \cdots s_n)$ is a binary string, weights $w_1 \geq w_2 \geq \cdots w_n > 0$ and $w_0 \geq 0$. This function is maximal at $(1 \cdots 1)$.

A $(1 + 1)$ EA can be used to solve this problem:

Mutation: At generation t, for individual $\xi_t = (s_1 \cdots s_n)$, flip each of its bits with probability $1/n$. The mutated population (in (1 + 1) EA, a population is an individual) is denoted as $\xi_t^{(m)}$.

Selection: If $f(\xi_t^{(m)}) > f(\xi_t)$, then $\xi_{t+1} = \xi_t^{(m)}$, otherwise, $\xi_{t+1} = \xi_t$.

We use two different distance functions to study this problem. The first one is that: given a binary string $x = (s_1 \cdots s_n)$,

$$V(x) = 4(n-1)^2 \left(\sum_{i=1}^{n} \mid s_i - 1 \mid \right). \tag{8}$$

Using this distance, we can get an $O(n^3)$ upper-bound on the mean first hitting time. Its proof is given in the appendix.

THEOREM 1. *Let τ be the mean number of generations for the (1+1) EA to find an optimal solution, then*

$$\mathbf{E}[\tau \mid \xi_0] = O(n^3). \tag{9}$$

Consider another distance function (without losing generality assume n is an even number):

$$V(x) = n \ln \left(1 + \sum_{i=1}^{n/2} c \mid s_i - 1 \mid + \sum_{i=n/2+1}^{n} \mid s_i - 1 \mid \right), \tag{10}$$

where $c (1 < c \leq 2)$ is a constant. The case of $c = 2$ is used in Droste et al. (2002). The case considered in this paper is more general.

Using this distance function, we can get a tighter upper bound, as given in the following theorem. The proof is in the appendix.

THEOREM 2. *Let τ be the mean number of generations for the* (1 + 1) *EA to find an optimal solution for the first time, then*

$$\mathbf{E}[\tau \mid \xi_0] = O(n \ln n). \tag{11}$$

4.2. *Analysis of an $(n + n)$-EA for the ONE-MAX problem*

It is well known that a (1 + 1)-EA can solve the ONE-MAX problem in time $\Theta(n \ln n)$ (Droste et al., 1998). Here we discuss a population-based $(n + n)$ EA, where n is the length of the string:

Mutation: For each individual $x = (s_1 \cdots s_n)$ in the population ξ_t at generation t, each bit s_i will be flipped with probability $1/n$. The mutated population is denoted as $\xi_t^{(m)}$.

Selection: Sort $2n$ individuals in ξ_t and $\xi_t^{(m)}$ according to their fitness from high to low. Then select n best individuals as the next generation ξ_{t+1}.

Two different distance functions will be used in this study. The first one is that: for an individual x, define the distance function to be

$$V(x) = 3(n-1)\ln\left(1 + \sum_{i=1}^{n} \mid s_i - 1 \mid\right). \tag{12}$$

For a population $\mathbf{x}$, define the distance function to be

$$V(\mathbf{x}) = \min\{V(x); x \in \mathbf{x}\}. \tag{13}$$

Similar to Theorem 2, we have

THEOREM 3. *Let τ be the mean number of generations for the $(n+n)$ EA to find an optimal solution for the first time, then*

$$\mathbf{E}[\tau \mid \xi_0] = O(n \ln n). \tag{14}$$

The second one is a little complex. Denote the Hamming distance between an individual and the optimal solution $(1 \cdots 1)$ as

$$H(x, \mathbf{1}) = \sum_{i=1}^{n} \mid s_i - 1 \mid,$$

for simplicity, write it as $H(x)$ in short. For a population $\mathbf{x}$, define

$$H(\mathbf{x}) = \min\{H(x); x \in \mathbf{x}\}.$$

We decompose the population set E into $n+1$ subsets as follows:

$$E_l = \{\mathbf{x}; H(\mathbf{x}) = l\}, \quad l = 0, 1, \cdots, n.$$

For each subset E_l ($l \in \{1, \cdots, n-1\}$), we can decompose E_l further into n subsets: for $k = 1, \cdots, n$,

$$E_{l,k} = \{\mathbf{x} \in E_l; \mathbf{x} \text{ includes } k \text{ individuals } x \text{ with } H(x) = l\ \}$$

Then we can define the distance function to be

$$V(\mathbf{x}) = d_{l,k}, \quad \text{if } \mathbf{x} \in E_{l,k}, \tag{15}$$

where $d_{l,k}$ is defined by

$$\begin{aligned} d_0 &:= 0, \\ d_{1,n} &:= 4, \\ d_{1,k} &:= 4 + 4\textstyle\sum_{i=k}^{n-1} 1/i, \qquad k = n-1, \cdots, 1, \end{aligned}$$

and for $l = 2, \cdots, n-1$,

$$\begin{aligned} d_{l,k} &:= d_{l-1,1} + 4, & k &= n, \cdots, \lceil n/l \rceil, \\ d_{l,k} &:= d_{l-1,1} + 4 + 4\textstyle\sum_{i=k}^{\lceil (n-1)/l \rceil} 1/i, & k &= \lceil (n-1)/l \rceil + 1, \cdots, 1, \end{aligned}$$

where $\lceil (n-1)/l \rceil$ represents the minimum integer not less than n/l, and finally

$$d_n :\ = d_{n-1,1} + 4.$$

Notice that $d_n = O(n)$.

Using the above distance we can get the following results. The proof is in the appendix.

THEOREM 4. *Let τ be the mean number of generations for the $(n+n)$ EA to find an optimal solution for the first time, then*

$$\mathbf{E}[\tau \mid \xi_0] = O(n). \tag{16}$$

5. Drift analysis for EAs in general: A problem classification

Given an EA, we can divide optimisation problems into two classes based on the mean number of generations needed to solve the problems.

Easy Class: For the given EA, starting from any initial population $\mathbf{x} \in E$, the mean number of generations needed by the EA to solve the problem, i.e., $\mathbf{E}[\tau \mid \xi_0 = \mathbf{x}]$, is polynomial in the problem size.

Hard Class: For the given EA, starting from some initial population $\mathbf{x} \in E$, the mean number of generations needed by the EA to solve the problem, i.e., $\mathbf{E}[\tau \mid \xi_0 = \mathbf{x}]$, is exponential in the problem size.

5.1. *Characteristics of easy problems*

A sufficient and necessary condition for the easy problem is given below.

THEOREM 5. *Given an EA which can be modelled by a homogeneous absorbing Markov chain, a problem belongs to the Easy Class if and only if there exists a distance function $V(\mathbf{x})$ such that*

1. *$V(\mathbf{x}) \leq g_1(n)$, where $g_1(n)$ is polynomial in the problem size n, and*
2. *the one-step mean drift satisfies: for any population ξ_t at generation t with $V(\xi_t) > 0$,*

$$\mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] \geq c_{low},$$

where $c_{low} > 0$ is a constant.

From the above theorem, we see that for a given EA, a problem belongs to the Easy Class if there exists a distance function, and (1) Condition 1 in the theorem shows that under the distance measure, no population is very far away from the optimal set (polynomial in the problem size); (2) Condition 2 shows that the one-step mean drift towards the optimal set is sufficiently large, always greater than a positive constant.

Since every population is only a *short-distance* away from the optimal set, and the one-step mean drift is always greater than a positive constant, the optimal set is easy to reach.

5.2. *Characteristics of hard problems*

Similar to the above analysis, a sufficient and necessary condition for a hard problem is given below.

THEOREM 6. *Given an EA that can be modelled by a homogeneous absorbing Markov chain, a problem belongs to the Hard Class if and only if there exists a distance function $V(\mathbf{x})$ such that*

1. *$V(\mathbf{x})$ satisfies: for some population $\mathbf{x} \in E : V(\mathbf{x}) \geq g_2(n)$, where $g_2(n)$ is exponential in the problem size n, and*
2. *the one-step mean drift satisfies: for any population ξ_t with $V(\xi_t) > 0$,*

$$\mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] \leq c_{up},$$

where c_{up} is a positive constant.

From this theorem we can see that: given an EA, a problem belongs to the Hard Class if there is a distance function, and (1) Condition 1 shows that under this distance, some populations are far away (exponential in the problem size) from the optimal set; (2) Condition 2 shows that the one-step mean drift towards the optimal set is limited and always less than a positive constant.

Since some population are at a *far-distance* from the optimal set and the one-step mean drift is limited by a certain positive constant, the number of generations needed to arrive at the optimal set will be very large (exponential) if starting from these populations.

5. Conclusions and discussions

This paper has shown that drift analysis is a useful tool in estimating the computation time of EAs. The analysis in this paper has shown that a number of results, both specific and general, can be derived using the drift analysis.

Drift analysis reduces the behaviour of EAs in a higher dimensional population space E into a supermartingale on the one-dimensional space. This reduction is implemented by the introduction of a distance function for the population space. The analysis of the one dimension random walk is much easier than analysing the original Markov chain. Two key points in applying drift analysis are: (i) to define a good distance function; and (ii) to estimate the mean drift.

The application of drift analysis in the study of EAs is still at its early days. More research is needed to generalise it in order to analyse more complex EAs on more complex problems in the future.

Appendix

Proof of Theorem 1.

Denote

$$d_l = 4(n-1)^2 l, \quad l = 0, \cdots, n.$$

Assume at generation t, the population ξ_t satisfies: $V(\xi_t) = d_l$ where $l \in \{1, \cdots, n\}$.

Firstly, we consider the positive drift. The following event will lead to a positive drift: k bits ($k = 1, \cdots, l$) among zero-valued bits of ξ_t flip and other bits keep unchanged. The positive drift satisfies:

$$\begin{aligned} \mathbf{E}^+[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] &= \sum_{k=1}^{l} (d_l - d_{l-k}) \mathbf{P}(V(\xi_{t+1}) = d_{l-k} \mid \xi_t) \\ &\geq \sum_{k=1}^{l} 4(n-1)^2 k \binom{l}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k}. \end{aligned}$$

Secondly, we consider the negative drift. Only the following event can produce the negative drift: k bits ($k = 1, \cdots, \min\{l, n-l-1\}$) in zero-valued bits of ξ_t flip, $k+m$ bits ($m = 1, \cdots, n-l-k$) of one-valued bits of ξ_t flip, and all other bits keep unchanged. The negative drift will be no more than

$$\begin{aligned} &\mathbf{E}^-[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] \\ &= \sum_{k=1}^{l} \sum_{m=1}^{n-l-k} (d_l - d_{l+m}) \mathbf{P}(V(\xi_{t+1}) = d_{l+m} \mid \xi_t) \\ &\geq -\sum_{k=1}^{l} \sum_{m=1}^{n-l-k} 4(n-1)^2 m \binom{l}{k} \binom{n-l}{k+m} \left(\frac{1}{n}\right)^{2k+m} \left(1 - \frac{1}{n}\right)^{n-2k-m}. \end{aligned}$$

Hence, the total drift will be

$$\mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t]$$

$$
\begin{aligned}
&\geq \sum_{k=1}^{l} 4(n-1)^2 k \binom{l}{k} \left(\frac{1}{n}\right)^k \left(1-\frac{1}{n}\right)^{n-k} \\
&\quad - \sum_{k=1}^{l} \sum_{m=1}^{n-l-k} 4(n-1)^2 m \binom{l}{k} \binom{n-l}{k+m} \left(\frac{1}{n}\right)^{2k+m} \left(1-\frac{1}{n}\right)^{n-2k-m} \\
&= 4(n-1)^2 \sum_{k=1}^{l} \binom{l}{k} \left(\frac{1}{n}\right)^k \left(1-\frac{1}{n}\right)^{n-k} \\
&\quad \left(k - \sum_{m=1}^{n-l-k} m \binom{n-l}{k+m} \left(\frac{1}{n-1}\right)^{k+m}\right) \\
&\geq 4(n-1)^2 \left(\frac{1}{n}\right) \left(1-\frac{1}{n}\right)^{n-1} \left(1 - \sum_{m=1}^{+\infty} m \binom{n-l}{m+1} \left(\frac{1}{n-1}\right)^{1+m}\right) \\
&> (n-1) \left(1 + \frac{1}{2(n-1)} - \sum_{m=2}^{+\infty} \frac{m}{(m+1)!}\right) \\
&> c_{low} := \frac{1}{2}.
\end{aligned}
$$

According to Lemma 2, we get

$$\mathbf{E}[\tau \mid \xi_0] \leq \frac{d_n}{c_{low}} = 8n(n-1)^2.$$

Proof of Theorem 2.
Denote

$$d_u = n \ln(1+u).$$

We call a bit s_i the left bit if $i \leq n/2$ or the right bit if $i \geq n/2+1$.

Assume at generation t, population ξ_t consists of l_1 ($l_1 \geq 0$) one-valued left bits and l_2 ($l_2 \geq 0$) one-valued right bits. Then $V(\xi_t) = d_{cl_1+l_2}$.

Firstly, we consider the positive drift. A positive drift will happen if the following event happens: $m_1(\geq 0)$ zero-valued left bits in ξ_t flip, $m_2(\geq 0)$ zero-valued right bits flip, and other bits keep unchanged, and $-cm_1 - m_2 < 0$. The probability of this event satisfies

$$
\begin{aligned}
&\mathbf{P}(V(\xi_{t+1}) = d_{cl_1+l_2-cm_1-m_2} \mid \xi_t) \\
&\geq \sum_{cm_1+m_2>0} \binom{l_1}{m_1} \binom{l_2}{m_2} \left(\frac{1}{n}\right)^{m_1+m_2} \left(1-\frac{1}{n}\right)^{n-(m_1+m_2)}.
\end{aligned}
$$

Next we consider the negative drift. A negative drift will happen only if the following event happens: $m_1(\geq 0)$ zero-valued left bits in ξ_t flip and $m_2(\geq 0)$ zero-valued right bits flip; $k_1(\geq 0)$ one-valued left bits in ξ_t flip and k_2 one-valued right bits flip; other bits in ξ_t keep unchanged; and $-cm_1 - m_2 + ck_1 + k_2 > 0$. The probability satisfies

$$
\begin{aligned}
& \mathbf{P}(V(\xi_{t+1}) = d_{cl_1+l_2-cm_1-m_2+ck_1+k_2} \mid \xi_t) \\
\leq & \sum_{-cm_1-m_2+ck_1+k_2>0} \binom{l_1}{m_1}\binom{l_2}{m_2}\binom{n/2-l_1}{k_2}\binom{n/2-l_2}{k_2} \\
& \left(\frac{1}{n}\right)^{m_1+m_2+k_1+k_2} \left(1-\frac{1}{n}\right)^{n-(m_1+m_2+k_1+k_2)}.
\end{aligned}
$$

Denote $u = cl_1 + l_2$, $v = cm_1 + m_2$ and $w = -cm_1 - m_2 + ck_1 + k_2$. $w = ck_1 + k_2 - v$. Then the mean drift satisfies

$$
\begin{aligned}
& \mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] \\
= & \sum_{v>0} (d_u - d_{u-v}) \mathbf{P}(V(\xi_{t+1}) = d_{u-v} \mid \xi_t) \\
& + \sum_{w>0} (d_u - d_{u+w}) \mathbf{P}(V(\xi_{t+1}) = d_{u+w} \mid \xi_t) \\
\geq & \sum_{v>0} \binom{l_1}{m_1}\binom{l_2}{m_2} \left(\frac{1}{n}\right)^{m_1+m_2} \left(1-\frac{1}{n}\right)^{n-(m_1+m_2)} \\
& \left(d_u - d_{u-v} + \sum_{w>0} \frac{d_u - d_{u+w}}{(n-1)^{k_1+k_2}} \binom{n/2-l_1}{k_1}\binom{n/2-l_2}{k_2} \right) \\
\geq & \sum_{v>0} \binom{l_1}{m_1}\binom{l_2}{m_2} \left(\frac{1}{n}\right)^{m_1+m_2} \left(1-\frac{1}{n}\right)^{n-(m_1+m_2)} \\
& \left(d_u - d_{u-v} + \sum_{w>0} (d_u - d_{u+w}) \frac{1}{(2k_1)!!(2k_2)!!} \right). \qquad (17)
\end{aligned}
$$

Since

$$
\begin{aligned}
& d_u - d_{u-v} + \sum_{w>0} (d_u - d_{u+w}) \frac{1}{(2k_1)!!(2k_2)!!} \\
\geq & n \ln(1+u) - n \ln(1+u-v) \\
& + \sum_{w>0} \left(n \ln(1+u) - n \ln(1+u+w) \right) \frac{1}{(2k_1)!!(2k_2)!!} \\
\geq & \frac{n}{1+u} \left(v - \sum_{w>0} \frac{w}{(2k_1)!!(2k_2)!!} \right), \qquad (18)
\end{aligned}
$$

we have the following results.

Case 1: $m_1 \geq 1$ **or** $m_1 = 0$ **and** $m_2 \geq 2$**.**

Then $v \geq c$, $w = ck_1 + k_2 - cm_1 - m_2 \leq ck_1 + k_2 - c$, and (18) is not less than:

$$\begin{aligned}
&\frac{n}{1+u}\left(c - \sum_{ck_1+k_2-c>0} \frac{ck_1+k_2-c}{(2k_1)!!(2k_2)!!}\right) \\
\geq\ &\frac{n}{1+u}\left(c - \sum_{k_1+k_2-1>0} \frac{ck_1+ck_2-c}{(2k_1)!!(2k_2)!!} + \sum_{k_1+k_2-1>0} \frac{(c-1)k_2}{(2k_1)!!(2k_2)!!}\right) \\
\geq\ &\frac{n}{1+u}(c-1)\left(\sum_{k_1+k_2-1>0} \frac{k_2}{(2k_1)!!(2k_2)!!}\right) \\
=\ &\frac{n}{1+u}c_1 > 0, \quad \text{(let } c_1 \text{ be the constant term in above).} \qquad (19)
\end{aligned}$$

Case 2: $m_1 = 0$ **and** $m_2 = 1$.

In this case, k_1 must be 0. Because in this case only 1 zero-valued right bit flips, so if $k_1 \geq 1$, i.e., one or more one-valued left bits flip, then the fitness does not increase, and the new individual won't be accepted. Thus (18) is not less than

$$\begin{aligned}
&\frac{n}{1+u}\left(1 - \sum_{k_2-1>0} \frac{k_2-1}{(2k_2)!!}\right) \\
\geq\ &\frac{n}{1+u}c_2 > 0, \quad \text{(let } c_2 \text{ be the constant term in above).} \qquad (20)
\end{aligned}$$

Since $u = cl_1 + l_2$, either $l_1 \geq u/(1+c)$ or $l_2 \geq u/(1+c)$. Without losing generality, assume $l_2 \geq u/(1+c)$. Then from (17), (19) and (20), we know that (denote $(c_1 \wedge c_2) = \min\{c_1, c_2\}$)

$$\begin{aligned}
&\mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] \\
\geq\ &\binom{l_1}{0}\binom{l_2}{1}\left(\frac{1}{n}\right)\left(1-\frac{1}{n}\right)^{n-1} \frac{n}{1+u}(c_1 \wedge c_2) \\
\geq\ &\frac{u}{1+c}\left(1-\frac{1}{n}\right)^{n-1} \frac{1}{1+u}(c_1 \wedge c_2) \\
\geq\ &c_{low} := \frac{c_1 \wedge c_2}{8(1+c)} \quad (\text{ using } u \geq 1).
\end{aligned}$$

According to Lemma 2, we know that

$$\mathbf{E}[\tau \mid \xi_0] \leq \frac{V(\xi_0)}{c_{low}} \leq \frac{d_n}{c_{low}} = O(n \ln n).$$

Proof of Theorem 4.

We decompose individual set S into $n+1$ subsets as follows:

$$S_l = \{x;\, H(x) = l\}, \quad l = 0, 1, \cdots, n.$$

At generation t, if $\xi_t \in E_n$, i.e., all individuals in the population take the value of $(0 \cdots 0)$, then we have

$$\mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] \geq 4 \sum_{l=0}^{n-1} \mathbf{P}(\xi_{t+1} \in E_l \mid \xi_t).$$

The event of ξ_{t+1} belonging to $\cup_{l=0}^{n-1} E_l$ will happen if at least one individual in population ξ_t enters S_{n-1}. The probability satisfies

$$\sum_{l=0}^{n-1} \mathbf{P}(\xi_{t+1} \in E_l \mid \xi_t) \geq \binom{n}{1} \left(\frac{1}{n}\right) \left(1 - \frac{1}{n}\right)^{n-1}.$$

So for population $\xi_t \in E_n$, the drift satisfies

$$\mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] \geq 1. \tag{21}$$

At generation t, if population $\xi_t \in E_{l,k}$, where $l = 1, \cdots, n-1$ and $k = 1, \cdots, n$, because the selection is $(n+n)$ elitist strategy, we have

$$\begin{aligned}
& \mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] \\
= & \sum_{i=0}^{l-1} (V(\xi_t) - V(\xi_{t+1}))\mathbf{P}(\xi_{t+1} \in E_i \mid \xi_t) \\
& + \sum_{j=k+1}^{n} (V(\xi_t) - V(\xi_{t+1}))\mathbf{P}(\xi_{t+1} \in E_{l,j} \mid \xi_t) \\
\geq & \; 4 \sum_{i=0}^{l-1} \mathbf{P}(\xi_{t+1} \in E_i \mid \xi_t) + 4\frac{1}{k} \sum_{j=k+1}^{n} \mathbf{P}(\xi_{t+1} \in E_{l,j} \mid \xi_t).
\end{aligned}$$

Case 1: $\lceil (n-1)/l \rceil \leq k \leq n$.

The event of ξ_{t+1} beloning to $\cup_{i=0}^{l-1} E_i$ will happen if an individual in population ξ_t is mutated to S_{l-1}. The event of an individual in ξ_t being mutated to S_{l-1} will happen if for one of k best individuals in ξ_t, at least one of its l zero-valued bits is flipped into 1 and $n-l$ one-value bits are remained unchanged. Then

$$\begin{aligned}
\sum_{i=0}^{l-1} \mathbf{P}(\xi_{t+1} \in E_i \mid \xi_t) & \geq \binom{k}{1}\binom{l}{1} \left(1 - \frac{1}{n}\right)^{n-1} \frac{1}{n} \\
& \geq \lceil (n-1)/l \rceil \frac{l}{n-1} \left(1 - \frac{1}{n}\right)^{n} \\
& \geq \left(1 - \frac{1}{n}\right)^{n}.
\end{aligned}$$

So for population $\xi_t \in E_{l,k}$ with $\lceil (n-1)/l \rceil \leq k \leq n$, the drift satisfies:

$$\mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] \geq 4 \sum_{i=0}^{l-1} \mathbf{P}(\xi_{t+1} \in E_i \mid \xi_t) \geq 1. \tag{22}$$

Case 2: $1 \leq k < \lceil (n-1)/l \rceil$.

The event of ξ_{t+1} belonging to $\cup_{j=k+1}^{n} E_{l,j}$ will happen if population ξ_{t+1} includes $k+1$ individuals such that $V(x) = l$ and the event of $\xi_{t+1} \in \cup_{i=0}^{l-1} E_i$ does not happen. This will happen if an individual among k best individuals in ξ_t is kept unchanged during the mutation and none of the individuals is mutated into $\cup_{i=0}^{l-1} E_i$. The probability of this event happening satisfies:

$$
\begin{aligned}
& \sum_{j=k+1}^{n} \mathbf{P}(\xi_{t+1} \in E_{l,j} \mid \xi_t) \\
\geq\ & \binom{k}{1} \left(1 - \frac{1}{n}\right)^n \left(1 - \sum_{i=0}^{l-1} \mathbf{P}(\xi_{t+1} \in E_i \mid \xi_t)\right) \\
\geq\ & \frac{k}{4} \left(1 - \sum_{i=0}^{l-1} \mathbf{P}(\xi_{t+1} \in E_i \mid \xi_t)\right).
\end{aligned}
$$

So if population $\xi_t \in E_{l,k}$ with $1 \leq k \leq \lceil (n-1)/l \rceil$, then the drift satisfies:

$$
\begin{aligned}
& \mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t \in E_{l,k}] \\
\geq\ & 4 \sum_{i=0}^{l-1} \mathbf{P}(\xi_{t+1} \in E_i \mid \xi_t) + \left(1 - \sum_{i=0}^{l-1} \mathbf{P}(\xi_{t+1} \in E_i \mid \xi_t)\right) \geq 1.
\end{aligned}
$$

From Lemma 2, (21), (22) and (23), we know that

$$
\mathbf{E}[\tau \mid \xi_0] \ \leq \ \frac{d_n}{1} = O(n).
$$

Proof of Theorem 5.

(1) Sufficient condition. From Theorem 1, we get

$$
\mathbf{E}[\tau \mid \xi_0] \leq \frac{V(\xi_0)}{c_{low}} \leq \frac{g_1(n)}{c_{low}}.
$$

Since $g_1(n)$ is polynomial in n, then $\mathbf{E}[\tau \mid \xi_0]$ is polynomial in n too.

(2) Necessary condition. Choose the distance function to be $V(\mathbf{x}) = \mathbf{E}[\tau \mid \xi_0]$. Let $g_1(n) = \max\{V(\mathbf{x}); \mathbf{x} \in E\}$, which is polynomial in the problem size.

From Lemma 3, we know that, if at generation t, $\xi_t = \mathbf{x} \notin E_{opt}$, then we have

$$
\mathbf{E}[V(\xi_t) - V(\xi_{t+1}) \mid \xi_t] = V(\mathbf{x}) - \sum_{\mathbf{y} \in E} \mathbf{P}(\mathbf{x}, \mathbf{y}) V(\mathbf{y}) = 1.
$$

Acknowledgements

Work partially supported by an EPSRC grant (GR/S63847/01) and the State Key Lab of Software Engineering, P.R. China.

References

Beyer H-G (2001) The Theory of Evolutionary Strategies. Springer, Berlin

Beyer H-G, Schwefel H-P and Wegener I (2002) How to analyse evolutionary algorithms. Theoretical Computer Science 287(1): 101–130

Chow YS and Teicher H (1988) Probability Theory: Independence, Interchangeability and Martingales, 2nd edn. Springer-Verlag, New York

Droste S, Jansen T and Wegener I (1998) A rigorous complexity analysis of the (1 + 1)-evolutionary algorithm for linear functions with Boolean inputs. Evolutionary Computation 6(2): 185–196

Droste S, Jansen T and Wegener I (2002) On the analysis of the (1 + 1)-evolutionary algorithms. Theoretical Computer Science 276(1–2): 51–81

Eiben AE and Rudolph G (1999) Theory of evolutionary algorithms: A bird's eye view. Theoretical Computer Science 229(1–2): 3–9

Hajek B (1982) Hitting time and occupation time bounds implied by drift analysis with applications. Advances in Applied Probability 14(3): 502–525

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

He J and Yao X (2002) From an individual to a population: An analysis of the first hitting time of population-based evolutionary algorithms. IEEE Transactions on Evolutionary Computation 6(5): 495–511

He J and Yao X (2003) Towards an analytic framework for analysing the computation time of evolutionary algorithms. Artificial Intelligence 145(1–2): 59–97

Meyn SP and Tweedie RL (1993) Markov Chains and Stochastic Stability. Springer-Verlag, New York

Rudolph G (1994) Convergence of non-elitist strategies. In: Proc. of the First IEEE conference on Evolutionary Computation, Vol. 1, pp. 63–66. IEEE Press, Piscataway, NJ

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

Sasaki GH and Hajek B (1988) The time complexity of maximum matching by simulated annealing. J ACM 35(2): 387–403

Syski R (1992) Passage Times for Markov Chains. IOS Press, Amsterdam