Analysis of Computational Time of Simple Estimation of Distribution Algorithms

Tianshi Chen, *Student Member, IEEE,* Ke Tang, *Member, IEEE,*
Guoliang Chen, and Xin Yao, *Fellow, IEEE*

NOTE: This paper was part of the Special Issue on Evolutionary Algorithms Based on Probabilistic Models that should have appeared in the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, Vol. 13, No. 6, December 2009.

Abstract—Estimation of distribution algorithms (EDAs) are widely used in stochastic optimization. Impressive experimental results have been reported in the literature. However, little work has been done on analyzing the computation time of EDAs in relation to the problem size. It is still unclear how well EDAs (with a finite population size larger than two) will scale up when the dimension of the optimization problem (problem size) goes up. This paper studies the computational time complexity of a simple EDA, i.e., the univariate marginal distribution algorithm (UMDA), in order to gain more insight into EDAs complexity. First, we discuss how to measure the computational time complexity of EDAs. A classification of problem hardness based on our discussions is then given. Second, we prove a theorem related to problem hardness and the probability conditions of EDAs. Third, we propose a novel approach to analyzing the computational time complexity of UMDA using discrete dynamic systems and Chernoff bounds. Following this approach, we are able to derive a number of results on the first hitting time of UMDA on a well-known unimodal pseudo-boolean function, i.e., the LeadingOnes problem, and another problem derived from LeadingOnes, named BVLeadingOnes. Although both problems are unimodal, our analysis shows that LeadingOnes is easy for the UMDA, while BVLeadingOnes is hard for the UMDA. Finally, in order to address the key issue of what problem characteristics make a problem hard for UMDA, we discuss in depth the idea of "margins" (or relaxation). We prove theoretically that the UMDA with margins can solve the BVLeadingOnes problem efficiently.

Index Terms—Computational time complexity, estimation of distribution algorithms, first hitting time, heuristic optimization, univariate marginal distribution algorithms.

T. Chen, K. Tang, and G. Chen are with the Nature Inspired Computation and Applications Laboratory, School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, China (e-mail: cetacy@mail.ustc.edu.cn; ketang@ustc.edu.cn; glchen@ustc.edu.cn).

X. Yao is with the Nature Inspired Computation and Applications Laboratory, School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, China, and also with the Center of Excellence for Research in Computational Intelligence and Applications, School of Computer Science, University of Birmingham, Edgbaston, Birmingham B15 2TT, U.K. (e-mail: x.yao@cs.bham.ac.uk).

I. INTRODUCTION

ESTIMATION of distribution algorithms (EDAs) [25], [28] are population-based stochastic algorithms that incorporate learning into optimization. Unlike evolutionary algorithms (EAs) that rely on variation operators to produce offspring, EDAs create offspring through sampling a probabilistic model that has been learned so far in the optimization process. Obviously, the performance of an EDA depends on how well we have learned the probabilistic model that tries to estimate the distribution of the optimal solutions. The general procedure of EDAs can be summarized in Table I. In recent years, many variants of EDAs have been proposed. On one hand, they have been shown experimentally to outperform other existing algorithms on many benchmark test functions. On the other hand, there were also experimental observations that showed EDAs did not scale well to large problems. In spite of a large number of experimental studies, theoretical analysis of EDAs has been few, especially on the computational time complexity of EDAs.

The importance of the time complexity of EDAs was recognized by several researchers. Mühlenbein and Schlierkamp-Voosen [31] studied the convergence time of constant selection intensity algorithms on the ONEMAX function. Later, Mühlenbein [27] studied the response to selection equation of the univariate marginal distribution algorithm (UMDA) on the ONEMAX function through experiments as well as theoretical analysis. Pelikan *et al.* [32] studied the convergence time of Bayesian optimization algorithm on the ONEMAX function. Rastegar and Meybodi [35] carried out a theoretical study of the global convergence time of a limit model of EDAs using drift analysis, but they did not investigate any relations between the problem size and computation time of EDAs. In addition to convergence time, the time complexity of EDAs can be measured by the first hitting time (FHT), which is defined as the first time for a stochastic optimization algorithm to reach the global optimum. Although recent work pointed out the significance of studying the FHT of EDAs [29], [33], few results have been reported. Droste's results [8] on the compact genetic algorithm (cGA) are a rare example. He analyzed rigorously the FHT of cGA with population size 2 [14] on linear functions. The other example is González's doctoral dissertation [13], where she analyzed the FHT of EDAs on the pseudo-boolean injective function using the analytical Markov chain framework proposed by He and Yao [17]. González [13] proved an important result that the worst-case mean FHT

TABLE I
GENERAL PROCEDURE OF EDA

$\xi_1 \leftarrow N$ individuals are generated by the initial probability distribution;
% Beginning of the 0th generation.
$t \leftarrow 1$;   % End of the 0th generation.
REPEAT
$\xi_t^{(s)} \leftarrow M$ individuals are selected from the N individuals in ξ_t;
% Beginning of the tth generation $(t \geq 1)$.
$p(x|\xi_t^{(s)}) \leftarrow$ The joint probability distribution is estimated from $\xi_t^{(s)}$;
$\xi_{t+1} \leftarrow N$ individuals are sampled from $p(x|\xi_t^{(s)})$;
$t \leftarrow t + 1$;   % End of the tth generation.
UNTIL THE STOPPING CRITERION IS MET.

ξ_t and $\xi_t^{(s)}$ are the populations before and after selection at the tth generation.

is exponential in the problem size for four commonly used EDAs. However, no specific problem was analyzed theoretically. Instead, González *et al.* [10] studied experimentally the mean FHT of three different types of EDAs, including the UMDA, on the LINEAR function, LEADINGONES function [4], [7], [16], [37], and UNIMAX (long-path) function [22].

This paper concerns theoretical analysis of the FHT of EDAs on the optimization problems with a unique global optimum. First, we provide a classification of problem hardness based on the FHT of EDAs, so that we can relate the problem characteristics to EDAs. This is very important for investigating the principles of when to use which EDAs for a given problem. Given such a classification (with respect to an EDA), we then investigate the relationship between EDAs probability conditions and problem hardness. Specifically, the time complexity of a simple EDA, the UMDA with truncation selection, is analyzed on two unimodal problems. The first problem is the LEADINGONES problem [37], which has frequently been studied in the field of time complexity analysis of EAs [7], [16]–[18]. The other problem is a variant of LEADINGONES, namely BVLEADINGONES.

Our analysis can be briefly summarized from two aspects. First, we propose a general approach to time complexity analysis of EDAs with *finite* populations. In the domain of EDAs, lots of theoretical results are based on infinite population assumption (e.g., [3], [11], [45]), while few consider the more realistic scenario that employs finite populations. Though we restrict our analysis to UMDA, our approach may also be useful for other EDAs. Second, both LEADINGONES and BVLEADINGONES are unimodal problems, and hence are usually expected to be easy for EDAs [11]. Our analysis confirms that LEADINGONES is easy for the UMDA studied. However, we interestingly find that BVLEADINGONES is hard for the UMDA. To deal with this issue, we relax the UMDA by the so-called *margins*, and prove that BVLEADINGONES becomes easy for this relaxed version of UMDA.

The rest of the paper is organized as follows. Section II discusses why FHT is more appropriate for time complexity analysis of EDAs and presents the classification of problem hardness and the corresponding probability conditions for EDAs. Section III presents the new approach to analyzing EDAs with *finite* populations and describes the UMDA studied in this paper. Then, UMDA is analyzed on LEADINGONES and BVLEADINGONES problems in Sections IV and V, respectively. Section VI studies the relaxation form of the UMDA on

the BVLEADINGONES problem. Finally, Section VII concludes the paper.

II. TIME COMPLEXITY MEASURES FOR EDAS

A. How to Measure the Time Complexity of EDAs

The concept of "convergence" is often used to measure the limit behaviors of EAs, including EDAs, which was derived from the concept of convergence of random sequences [37]. For EDAs, the following formal definition of "convergence" was given by Zhang and Mühlenbein [45]:

> If $\lim_{t \to \infty} \bar{F}(t) = g^*$ holds for a given EDA, where $\bar{F}(t)$ is the average fitness of individuals in the tth generation and g^* is the fitness of the global optimum, then we say that the EDA converges to the global optimum.

There has been some work concerning such convergence of EDAs [12], [30]. It is worth noting that the above definition of convergence requires *all* individuals of a population to reach the global optimum. If we assume that an EDA on a problem converges to the global optimum, we can then measure the EDAs time complexity using the minimal number of generations that is needed for it to converge. This concept is called the convergence time (CT), denoted by T in this paper. For EDAs, the CT is formally defined by

$$T \triangleq \min\{t; p(x^*|\xi_t^{(s)}) = 1\} \tag{1}$$

where x^* is the global optimum of a given problem, and $\xi_t^{(s)}$ is the population after selection at the tth generation. $p(x^*|\xi_t^{(s)})$ is the estimated probability (of generating x^*) by the EDA at the tth generation.

In addition to CT, the FHT is also a commonly used concept for measuring the time complexity of EAs [16], [17]. The FHT [16], [17], [43], denoted by τ, is defined for the general procedure of EDA shown in Table I

$$\tau \triangleq \min\{t; x^* \in \xi_{t+1}\} \tag{2}$$

where ξ_{t+1} is the population generated at the end of tth generation. In the domain of EA, the FHT records the smallest number of generations needed to find the optimum, which is by a factor N smaller than another commonly used measure named number of fitness evaluations, where N is the number of fitness evaluations in every generation [9]. As González pointed out in [13], the FHT can also be used to measure the time complexity of EDAs.

Since EDAs are stochastic algorithms, both CT T and FHT τ are random variables. Noting that the FHT measures the time for the global optimum to be found for the first time, thus the CT is no smaller than FHT

$$T \geq \tau \qquad (3)$$

which implies a natural way to bound CT from below by FHT or bound FHT from above by the CT.

In practical optimization, we are most interested in the time spent in finding the global optimum, not in waiting for the whole population to converge to the global optimum. Hence, the FHT is a better measure for analyzing the time complexity of the EDAs. It is worth noting that for a given EDA on a problem, it may have a small FHT but large CT. In other words, the population may take a long time (even infinite) to converge to the global optimum. In such cases, the analysis of FHT is still valid while the analysis of CT is rather uninteresting. It is possible that an EDA could find the global optimum efficiently (in polynomial time), but the population does not converge to the global optimum. We will discuss such an example in Section VI.

B. Probability Conditions for EDA–Hardness

In order to understand better the relationship between problem characteristics and algorithmic features of an EDA, we introduce a problem classification for a given EDA. However, we should introduce some notations first.

Denote $Poly(n)$ as the polynomial function class of the problem size n and $SuperPoly(n)$ as the super-polynomial function class of the problem size n. For a function $f(n)$ (where $f(n) > 1$ always holds, and when $n \to \infty$, $f(n) \to \infty$), denote the following:

1) $f(n) \prec Poly(n)$ and $g(n) = \frac{1}{f(n)} \succ \frac{1}{Poly(n)}$ if and only if $\exists a, b \in \mathbb{R}^+$, $n_0 \in \mathbb{N}$: $\forall n > n_0$, $f(n) \leq an^b$;
2) $f(n) \succ SuperPoly(n)$ and $g(n) = \frac{1}{f(n)} \prec \frac{1}{SuperPoly(n)}$ if and only if $\forall a, b \in \mathbb{R}^+$: $\exists n_0 \in \mathbb{N}$: $\forall n > n_0$, $f(n) > an^b$.

Based on the above definitions, we know that "$\prec$" and "$\succ$" imply "$<$" and "$>$" respectively, when n is *sufficiently large*. $Poly(n)$ [$SuperPoly(n)$] implies that there *exists* a monotonically increasing function that is polynomial (super-polynomial) in the problem size n. Note that $g(n) = \frac{1}{f(n)} \in (0, 1)$, and its asymptotic form $g(n) \succ \frac{1}{Poly(n)}$ or $g(n) \prec \frac{1}{SuperPoly(n)}$, can be used to measure the asymptotic order of a probability (e.g., the probability of generating a certain individual), since a probability always takes its value in the interval $[0, 1]$.[1] Then we provide the following problem classification for a given EDA.

---

[1] For $g(n) \in [0, 1]$, there are more detailed asymptotic orders in the interval $[0, 1]$:

1) $g(n) \prec \frac{1}{SuperPoly(n)}$;
2) $\frac{1}{Poly(n)} \prec g(n) \prec 1 - \frac{1}{Poly(n)}$ [if and only if $\exists a_1, b_1, a_2, b_2 \in \mathbb{R}^+$, $n_0, n_1 \in \mathbb{N}$: $\forall n > \max\{n_0, n_1\}$, $1/(a_1 n^{b_1}) \leq g(n) \leq 1 - 1/(a_2 n^{b_2})$];
3) $g(n) \succ 1 - \frac{1}{SuperPoly(n)}$ [if and only if $\forall a, b \in \mathbb{R}^+$: $\exists n_0 \in \mathbb{N}$: $\forall n > n_0$, $g(n) \geq 1 - 1/(an^b)$].

If necessary, these detailed asymptotic orders can be obtained by considering the regions $c \pm \frac{1}{Poly(n)}$ and $c \pm \frac{1}{SuperPoly(n)}$, where $0 < c < 1$.

1) *EDA-easy Class*. For a given EDA, a problem is *EDA-easy* if, and only if, with the probability of $1 - 1/SuperPoly(n)$, the FHT needed to reach the global optimum is polynomial in the problem size n.
2) *EDA-hard Class*. For a given EDA, a problem is *EDA-hard* if, and only if, with the probability of $1/Poly(n)$, the FHT needed to reach the global optimum is super-polynomial in the problem size n.

The above classification can be considered as a direct generalization of the following EA-hardness classification for EAs proposed by He and Yao [18].

1) *EA-easy Class*. For a given EA, a problem is *EA-easy* if, and only if, the mean FHT needed to reach the global optimum is polynomial in the problem size n.
2) *EA-hard Class*. For a given EA, a problem is *EA-hard* if, and only if, the mean FHT needed to reach the global optimum is super-polynomial in the problem size n.

We see that He and Yao's classification for EAs is based on mean FHT, while our classification for EDAs concerns more detailed characteristics of the probability distribution of FHT. Given a problem, if the FHT of an EDA is polynomial with a probability super-polynomially close to 1 (the probability will be called "*an overwhelming probability*" in the following parts of the paper), then we can say that in most of independent runs, the EDA can find the optimum of the problem efficiently. On the other hand, if the FHT of an EDA is super-polynomial with a probability that is polynomially large. i.e., $1/Poly(n)$, then it is very likely that the EDA cannot find the optimum of the problem efficiently. A similar idea can be found in [42], which defined efficiency measures for randomized search heuristics.

From the definition of expectation in probability theory, we know that for an algorithm, the problems belonging to the EDA-hard class in our classification will still be hard under the classification based on mean FHT. But our classification defines EDA-easy differently from the classification based on mean FHT. In practice, it is possible that an EDA finds the optimum efficiently in most of the independent runs, while spends extremely long time in the other runs. This kind of problems will considered to be "hard" cases if using mean FHT for classification. However, in our classification, such problems are considered to be easy cases, which is more likely to fit the practitioners' point of view.

We now establish conditions under which a problem is EDA-hard (or EDA-easy) for a given EDA. Let $\mathbb{P}(\tau = t)$ $(t \in \mathbb{N})$ be the probability distribution of the FHT, which is determined by the probabilistic model at the tth generation. An EDA can be regarded as a random process $K = \{K_t : t \in \mathbb{N}\}$, where K_t is the probabilistic model (including the parameters) maintained at the tth generation. Obviously, K_t implies the probability of generating the global optimum *in one sampling* at the tth generation, denoted by P_t^*

$$\forall t \in \mathbb{N}: K_t \vdash P_t^*. \qquad (4)$$

Meanwhile, to obtain the probability distribution of the FHT τ, we let P_t' be the probability of generating the global optimum in one sampling at the tth generation, conditional on the event $\tau \geq t$ (i.e., the global optimum has not been

generated before the tth generation). Consequently, we obtain the following lemma:

Lemma 1: The probability distribution of the FHT τ satisfies

$$\forall t \geq 0 : \mathbb{P}(\tau = t) = \left(1 - (1 - P'_t)^N\right) \prod_{j=0}^{t-1} (1 - P'_j)^N. \quad (5)$$

Proof: Let x^* be the global optimum. As Table I and (2), we also let ξ_{t+1} be the generated population at the end of tth generation ($t \in \mathbb{N}$). According to the FHT defined in (2), for any $t \in \mathbb{N}^+$ we have

$$\mathbb{P}(\tau = t) = \mathbb{P}\left(x^* \in \xi_{t+1}, x^* \notin \xi_t, \ldots, x^* \notin \xi_2, x^* \notin \xi_1\right)$$

$$= \mathbb{P}\left(x^* \in \xi_{t+1}, x^* \notin \xi_t, \ldots, x^* \notin \xi_2 \mid x^* \notin \xi_1\right)$$
$$\cdot \mathbb{P}\left(x^* \notin \xi_1\right)$$

$$= \mathbb{P}\left(x^* \in \xi_{t+1}, x^* \notin \xi_t, \ldots, x^* \notin \xi_3 \mid x^* \notin \xi_2, x^* \notin \xi_1\right)$$
$$\cdot \mathbb{P}\left(x^* \notin \xi_2 \mid x^* \notin \xi_1\right) \mathbb{P}\left(x^* \notin \xi_1\right)$$

$$= \mathbb{P}\left(x^* \in \xi_{t+1} \mid x^* \notin \xi_t, \ldots, x^* \notin \xi_1\right) \mathbb{P}\left(x^* \notin \xi_1\right)$$
$$\cdot \prod_{j=1}^{t-1} \mathbb{P}\left(x^* \notin \xi_{j+1} \mid x^* \notin \xi_j, \ldots, x^* \notin \xi_1\right)$$

$$= \mathbb{P}\left(x^* \in \xi_{t+1} \mid \tau \geq t\right) \prod_{j=0}^{t-1} \mathbb{P}\left(x^* \notin \xi_{j+1} \mid \tau \geq j\right)$$

$$= \left(1 - (1 - P'_t)^N\right) \prod_{j=0}^{t-1} (1 - P'_j)^N$$

where N is the population size, the item $1 - (1 - P'_t)^N$ is the probability that the optimum is found at the tth generation, conditional on the event $\tau \geq t$, and the item $\prod_{j=0}^{t-1}(1 - P'_j)^N$ is the probability that the optimum has not been found before the tth generation. Combining the above result with the fact $\mathbb{P}(\tau = 0) = 1 - (1 - P'_0)^N$, we have proven the lemma. ∎

Moreover, let us consider the following lemma:

Lemma 2: If $\mathbb{P}(\tau \prec Poly(n)) \succ 1 - \frac{1}{SuperPoly(n)}$, then $\exists t' \leq \lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 1$ such that

$$\mathbb{P}(\tau = t') \succ \frac{1}{Poly(n)}.$$

Proof: Assume that $\forall t \leq \lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 1$, $\mathbb{P}(\tau = t) \prec \frac{1}{SuperPoly(n)}$, then we know that

$$\max \left\{ \mathbb{P}(\tau = t); t \leq \lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 1 \right\}$$
$$\prec \frac{1}{SuperPoly(n)}.$$

Hence, we can obtain

$$\mathbb{P}(\tau \leq \lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 1)$$
$$= \sum_{t=0}^{\lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 1} \mathbb{P}(\tau = t)$$
$$\leq \left(\lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 2 \right)$$

$$\cdot \max \left\{ \mathbb{P}(\tau = t); t \leq \lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 1 \right\}$$
$$\prec \frac{Poly(n)}{SuperPoly(n)}.$$

Now we can estimate the expectation of the FHT τ

$$\mathbb{E}[\tau \mid \tau \prec Poly(n)] = \sum_{t=0}^{+\infty} t\mathbb{P}(\tau = t \mid \tau \prec Poly(n))$$

$$= \sum_{t=0}^{Poly(n)} \frac{t\mathbb{P}(\tau = t, \tau \prec Poly(n))}{\mathbb{P}(\tau \prec Poly(n))}$$

$$= \sum_{t=0}^{Poly(n)} \frac{t\mathbb{P}(\tau = t)}{\mathbb{P}(\tau \prec Poly(n))} \geq \sum_{t=0}^{Poly(n)} t\mathbb{P}(\tau = t)$$

$$= \sum_{t=0}^{\lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 1} t\mathbb{P}(\tau = t)$$

$$+ \sum_{t=\lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 2}^{Poly(n)} t\mathbb{P}(\tau = t)$$

$$> (\lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 2)$$
$$\cdot \mathbb{P}\left(Poly(n) \succ \tau > \lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 1 \right)$$

$$= (\lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 2)\left(\mathbb{P}(\tau \prec Poly(n)) \right.$$
$$\left. - \mathbb{P}(\tau \leq \lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 1) \right)$$

$$= (\lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 2)$$
$$\cdot \left(1 - \frac{1}{SuperPoly(n)} - \frac{Poly(n)}{SuperPoly(n)} \right)$$

$$\succ (\lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 2) - \frac{Poly(n)}{SuperPoly(n)}$$
$$- \frac{Poly(n)Poly(n)}{SuperPoly(n)}.$$

As $n \to \infty$, $\frac{Poly(n)}{SuperPoly(n)} \to 0$ and $\frac{Poly(n)Poly(n)}{SuperPoly(n)} \to 0$. Hence, there exists a sufficiently large problem size n such that

$$\mathbb{E}[\tau \mid \tau \prec Poly(n)] > \lceil \mathbb{E}[\tau \mid \tau \prec Poly(n)] \rceil + 1 \quad (6)$$

which is an obvious contradiction. So we have proven the lemma. ∎

Formally, an optimization problem can be denoted by $I = (\Omega, f)$, where Ω is the search space and f the fitness function. Following He *et al.* [19], we use $\mathcal{P} = (\Omega, f, \mathcal{A})$ to indicate an algorithm $\mathcal{A}$ on a fitness function f in the search space Ω. Let the FHT of $\mathcal{A}$ on I be $\tau(\mathcal{P})$. The following theorem describes the relation between EDA-hardness and probability P_i^*.

Theorem 1: For a given $\mathcal{P}$, if the population size N of the EDA $\mathcal{A}$ is polynomial in the problem size n, then:

1) if I is EDA-easy for $\mathcal{A}$, then $\exists t'' \leq \lceil \mathbb{E}[\tau(\mathcal{P}) \mid \tau(\mathcal{P}) \prec Poly(n)] \rceil + 1$ such that

$$P_{t''}^* \succ \frac{1}{Poly(n)};$$

2) if $\forall t = t(n) \prec Poly(n)$, $P_t^* \prec \frac{1}{SuperPoly(n)}$, then I is EDA-hard for $\mathcal{A}$.

Proof: Note that the second part of this theorem is a corollary of the first part. We only need to prove the first part.

According to Lemma 1, we have

$$\mathbb{P}(\tau(\mathcal{P}) = i) < 1 - (1 - P_i')^N.$$

On the other hand, according to Lemma 2, we know that $\exists t' \leq \lceil \mathbb{E}[\tau(\mathcal{P}) \mid \tau(\mathcal{P}) \prec Poly(n)] \rceil + 1$ such that

$$\mathbb{P}(\tau(\mathcal{P}) = t') \succ \frac{1}{Poly(n)}.$$

Thus, we can define t'' as follows:

$$t'' = \min \left\{ t'; t' \leq \lceil \mathbb{E}[\tau(\mathcal{P}) \mid \tau(\mathcal{P}) \prec Poly(n)] \rceil + 1, \right.$$
$$\left. \mathbb{P}(\tau(\mathcal{P}) = t') \succ \frac{1}{Poly(n)} \right\}. \qquad (7)$$

Since $\mathbb{P}(\tau(\mathcal{P}) = t'') \succ \frac{1}{Poly(n)}$, we have

$$1 - (1 - P_{t''}')^N \succ \frac{1}{Poly(n)}. \qquad (8)$$

Let us assume that $P_{t''}^* \prec \frac{1}{SuperPoly(n)}$. Here we let $\mathcal{E}$ represent the event "the global optimum is generated in one sampling at the t''-th generation," then according to the definitions of $P_{t''}^*$ and $P_{t''}'$ mentioned in Section II-B, we obtain the following inequality:

$$\begin{aligned} P_{t''}^* = \mathbb{P}(\mathcal{E}) &\geq \mathbb{P}(\mathcal{E}, \tau(\mathcal{P}) \geq t'') \\ &= \mathbb{P}(\mathcal{E} \mid \tau(\mathcal{P}) \geq t'') \mathbb{P}(\tau(\mathcal{P}) \geq t'') \\ &= P_{t''}' \mathbb{P}(\tau(\mathcal{P}) \geq t''). \end{aligned} \qquad (9)$$

Meanwhile, (7) implies that

$$\mathbb{P}(\tau(\mathcal{P}) \geq t'') \geq \mathbb{P}(\tau(\mathcal{P}) = t'') \succ \frac{1}{Poly(n)}. \qquad (10)$$

Combining (9) and (10) together, we know that $P_{t''}^* \prec \frac{1}{SuperPoly(n)}$ yields $P_{t''}' \prec \frac{1}{SuperPoly(n)}$.

Now $\forall f(n) \prec Poly(n)$, we estimate

$$\lim_{n \to \infty} \frac{1 - \left(1 - P_{t''}'\right)^N}{1/f(n)} \qquad (11)$$

where $N = N(n) \prec Poly(n)$ is the population size of the EDA. Equation (11) can be calculated as follows:

$$\lim_{n \to \infty} \frac{1 - \left(1 - P_{t''}'\right)^{N(n)}}{1/f(n)}$$
$$= \lim_{n \to \infty} \frac{1 - \left(\left(1 - P_{t''}'\right)^{\frac{1}{P_{t''}'}}\right)^{P_{t''}' N(n)}}{1/f(n)}$$
$$= \lim_{n \to \infty} \left(f(n) - f(n) e^{-P_{t''}' N(n)} \right)$$
$$= \lim_{n \to \infty} \left(f(n) - f(n) \left(1 - P_{t''}' N(n) \right. \right.$$
$$\left. \left. + \frac{(P_{t''}' N(n))^2}{2} + o\left((P_{t''}' N(n))^2\right) \right) \right)$$
$$= \lim_{n \to \infty} f(n) P_{t''}' N(n) - \lim_{n \to \infty} \frac{f(n)(P_{t''}' N(n))^2}{2}$$
$$- \lim_{n \to \infty} o\left(f(n) \left(P_{t''}' N(n)\right)^2 \right)$$

$$\prec \lim_{n \to \infty} \frac{Poly^2(n)}{SuperPoly(n)} - \lim_{n \to \infty} \frac{Poly^3(n)}{SuperPoly^2(n)}$$
$$- \lim_{n \to \infty} o\left(\frac{Poly^3(n)}{SuperPoly^2(n)} \right) = 0.$$

Hence, we know that $1 - (1 - P_{t''}')^N$ is smaller than $\frac{1}{f(n)} \succ \frac{1}{Poly(n)}$ when $n \to \infty$. In other words

$$1 - (1 - P_{t''}')^N \prec \frac{1}{SuperPoly(n)}$$

where we obtain a contradiction to (8).

So we have

$$P_{t''}^* \succ \frac{1}{Poly(n)}.$$

The theorem is proven. ∎

The theorem above provides us with two simple probability conditions related to the problem classification in terms of EDA-hardness. Later, we will use this theorem to obtain more specific results related to EDA-hardness for the UMDA.

III. TIME COMPLEXITY ANALYSIS OF EDAS WITH FINITE POPULATION SIZES

A. A General Approach to Analyzing EDAs With Finite Population Sizes

In the domain of EA, several different approaches have been proposed for analyzing theoretically the FHT, such as drift analysis [16], [18], analytical Markov chain [17], Chernoff bounds [7], [23], [24], and convergence rate [15], [43]. Some of them have been applied to EDAs as well. González used the analytical Markov chain to study the worst case exponential FHT of some EDAs [13]. Droste employs drift analysis and Chernoff bounds to analyze the time complexity of cGA (with a population size of two) on linear pseudo-boolean functions [8]. However, those existing techniques might not be sufficient for time complexity analysis of EDAs, because EDAs do not use any variation operators (e.g., mutation and crossover) but rely on sampling successive probabilistic models. Hence, some new ideas are needed to deal with probabilistic models.

One of the main difficulties of analyzing probabilistic models is due to the errors brought by the random sampling processes. Such random errors may occur when a probabilistic model is updated via random sampling. An intuitive idea of handling the random errors is to assume infinite population sizes for EDAs. This assumption has been adopted in the most existing literature, such as the well-known example of ONEMAX given by Mühlenbein and Schlierkamp-Voosen [31], and Zhang's convergence analysis of EDAs [45]. Two exceptions are the aforementioned Droste's results on cGA [8] and González's general worst case analysis of EDAs [13].

In this section, we will provide a general approach to analyzing theoretically EDAs with finite population sizes. The approach is closely related to Chernoff bounds and the discrete dynamic system model of population-based incremental learning (PBIL) [1]. PBIL is a more general version of UMDA and its discrete dynamic system model was first presented by González *et al.* [11]–[13]. Assume there is a function $\mathcal{G} : \mathbb{R}^n \to \mathbb{R}^n$, then $A(t + 1) = \mathcal{G}(A(t))$ $(t = 0, 1, \ldots)$ is called

a discrete dynamic system [39]. In [11]–[13], two discrete dynamic system were discussed. The first one considered PBIL as a function $\mathcal{G}_1 : [0, 1]^n \rightarrow [0, 1]^n$. $\mathcal{G}_1$ includes the random effects. Hence, even if the initial probability distribution and algorithm parameters of PBIL are fixed, the system is still stochastic. This is an exact model of PBIL, but hard to analyze directly. So the authors considered the second dynamic system with the function $\mathcal{G}_2 : [0, 1]^n \rightarrow [0, 1]^n$, which removes the random effects by assuming an infinite population size and thereby becomes deterministic. Although the deviation (caused by the random sampling errors) between the two dynamic systems has been estimated, so as to study the fixed point of the first dynamic system by investigating that of the second system, their method does not relate the deviation to the computation time of PBIL. Hence, it is not applicable to time complexity analysis.

Although González *et al.* [11]–[13] did not analyze the time complexity of EDAs, their mathematical models (using the discrete dynamic systems) can be used to develop a feasible approach to analyzing the time complexity of EDAs. Such an approach can be summarized by two major steps.

1) Build an easy-to-analyze discrete dynamic system for the EDA. The idea is to de-randomize the EDA and build a deterministic[2] dynamic system.
2) Analyze the deviations caused by de-randomization. Note that EDAs are stochastic algorithms. Concretely, tail probability techniques, such as Chernoff bounds, can be used to bound the deviations.

In this paper, we will use UMDA as an example of EDAs to illustrate the analysis of EDAs time complexity using the above approach. The analysis will show that our approach provides a feasible way of estimating the random errors brought by finite populations in UMDA, and thus shed some light on analyzing other EDAs with finite populations. However, it should be noted that much work remains to be done to achieve such a goal.

B. Univariate Marginal Distribution Algorithm

The UMDA was originally proposed as a discrete EDA [28], [44]. As one of the earliest and simplest EDAs, UMDA has attracted a lot of research attention. The UMDA studied in this paper adopts binary encoding and one of the most commonly used selection strategies—the truncation selection, which is described below.

> *Sort the N individuals in the population by their fitness from high to low. Then select the best M of them for estimating the probability distribution.*

The general procedure of UMDA studied in our paper is shown in Table II, where $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \{0, 1\}^n$ represents an individual, $p_{t,i}(1)$ ($p_{t,i}(0)$) is the estimated marginal probability of the ith bit of an individual to be 1 (0) at the tth generation. We can also define the indicators $\delta(x_i|1)$ as follows:

$$\delta(x_i|1) \triangleq \begin{cases} 1, & x_i = 1 \\ 0, & x_i = 0. \end{cases}$$

[2]In our discussions, "deterministic" is always in the sense that we have fixed the initial values of all the parameters of the non-self-adaptive EDA.

The marginal probabilities $p_{t,i}(1)$ and $p_{t,i}(0)$ are given by

$$p_{t,i}(1) \triangleq \frac{\sum_{\mathbf{x} \in \xi_t^{(s)}} \delta(x_i|1)}{M}, \quad p_{t,i}(0) \triangleq 1 - p_{t,i}(1).$$

Let

$$\mathbf{P}_t(\mathbf{x}) \triangleq \left(p_{t,1}(x_1), p_{t,2}(x_2), \ldots, p_{t,n}(x_n) \right)$$

where $\mathbf{P}_t(\mathbf{x})$ is a probability vector, which is made up of n random variables (that is because, UMDA is a stochastic algorithm). Then the probability of generating individual $\mathbf{x}$ in the tth generation is

$$p_t(\mathbf{x}) = \prod_{i=1}^{n} p_{t,i}(x_i).$$

C. Analyzing Time Complexity of UMDA

The UMDA given in the former section can be analyzed following the general idea presented in Section III-A. First, we define a function $\gamma : [0, 1]^n \rightarrow [0, 1]^n$ such that $\gamma = \mathcal{S} \circ \mathcal{D}$, where $\mathcal{S} : [0, 1]^n \rightarrow [0, 1]^n$ is the function that represents the effect of selection, and $\mathcal{D} : [0, 1]^n \rightarrow [0, 1]^n$ is the function that is used in eliminating the stochastic effects of the random sampling. Then we obtain a deterministic discrete dynamic system $\{\hat{\mathbf{P}}_t(\mathbf{x}^*); t = 0, 1, \ldots\}$ related to the marginal probabilities of generating the global optimum

$$\hat{\mathbf{P}}_0(\mathbf{x}^*) = \mathbf{P}_0(\mathbf{x}^*) \tag{12}$$

$$\hat{\mathbf{P}}_{t+1}(\mathbf{x}^*) = \gamma\left(\hat{\mathbf{P}}_t(\mathbf{x}^*)\right) = \mathcal{S}\left(\mathcal{D}\left(\hat{\mathbf{P}}_t(\mathbf{x}^*)\right)\right) \tag{13}$$

$$\hat{\mathbf{P}}_t(\mathbf{x}^*) = \gamma^t\left(\hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \tag{14}$$

where $\hat{\mathbf{P}}_t(\mathbf{x}) = \left(\hat{p}_{t,1}(x_1), \ldots, \hat{p}_{t,n}(x_n) \right)$ is the marginal probability vector of the deterministic system for generating an individual $\mathbf{x}$, and $\mathbf{x}^*$ is the global optimum. Since UMDA is usually initialized with a uniform distribution, we consider $\hat{\mathbf{P}}_0(\mathbf{x}) = \mathbf{P}_0(\mathbf{x}) = \left(\frac{1}{2}, \ldots, \frac{1}{2}\right)$ in this paper. Correspondingly, the probability of generating an individual $\mathbf{x}$ is

$$\hat{p}_t(\mathbf{x}) = \prod_{i=1}^{n} \hat{p}_{t,i}(x_i).$$

Note that $p_t(\mathbf{x})$ in the former section corresponds to the original UMDA, while $\hat{p}_t(\mathbf{x})$ is obtained from the deterministic dynamic system after de-randomization. Following the first step of our general approach, we need to estimate the time complexity of the de-randomized UMDA.

To relate the time complexity result obtained by the deterministic system to the original UMDA, we should estimate the deviation of the de-randomized UMDA from the original UMDA. Since time complexity of the former totally depends on $\{\hat{\mathbf{P}}_t(\mathbf{x}^*); t = 0, 1, \ldots\}$, such deviation arises from the difference between $\{\hat{\mathbf{P}}_t(\mathbf{x}^*); t = 0, 1, \ldots\}$ and $\{\mathbf{P}_t(\mathbf{x}^*); t = 0, 1, \ldots\}$. Ideally, we want to exactly calculate the difference between the two sequences of marginal probability vectors. However, this is a non-trivial work (if not impossible). Alternatively, we resort to estimating the probabilities that the deviations are smaller than some specific values. Two crucial lemmas for this task are given below.

TABLE II

UNIVARIATE MARGINAL DISTRIBUTION ALGORITHM (UMDA) WITH TRUNCATION SELECTION

$p_{0,i}(x_i) \leftarrow$ Initial values $(\forall i = 1, \ldots, n)$
$\xi_1 \leftarrow N$ individuals are sampled according to the distribution
$$p_0(\mathbf{x}) = \prod_{i=1}^{n} p_{0,i}(x_i)$$
$t \leftarrow 1;$
REPEAT
    $\xi_t^{(s)} \leftarrow$ The best M individuals are selected from the N individuals in ξ_t $(N > M)$
    $p_{t,i}(1) \leftarrow \dfrac{\sum_{\mathbf{x} \in \xi_t^{(s)}} \delta(x_i|1)}{M}, \; p_{t,i}(0) \leftarrow 1 - p_{t,i}(1) \quad (\forall i = 1, \ldots, n)$
    $\xi_{t+1} \leftarrow N$ individuals are sampled according to the distribution
    $$p_t(\mathbf{x}) = \prod_{i=1}^{n} p_{t,i}(x_i)$$
    $t \leftarrow t + 1;$
UNTIL THE STOPPING CRITERION IS MET

Lemma 3 ([26]): Chernoff Bounds. Let $X_1, X_2, \ldots, X_k \in \{0, 1\}$ be k independent random variables (take the value of either 0 or 1) with a same distribution

$$\forall i \neq j : \mathbb{P}(X_i = 1) = \mathbb{P}(X_j = 1)$$

where $i, j \in \{1, \ldots, k\}$. Let X be the sum of those random variables, i.e., $X = \sum_{i=1}^{k} X_i$, then we have:

1) $\forall 0 < \delta < 1$

$$\mathbb{P}\left(X < (1 - \delta)\mathbb{E}[X]\right) < e^{-\mathbb{E}[X]\delta^2/2};$$

2) $\forall \delta \leq 2e - 1$

$$\mathbb{P}\left(X > (1 + \delta)\mathbb{E}[X]\right) < e^{-\mathbb{E}[X]\delta^2/4}.$$

Lemma 4 ([21], [38]): Consider sampling without replacement from a finite population $(X_1, \ldots, X_N) \in \{0, 1\}^N$. Let $(Y_1, \ldots, Y_M) \in \{0, 1\}^M$ be a sample of size M get randomly without replacement from the whole population, $Y^{(M)}$ and $X^{(N)}$ be the sums of the random variables in the sample and population, respectively, i.e., $Y^{(M)} = \sum_{i=1}^{M} Y_i$ and $X^{(N)} = \sum_{i=1}^{N} X_i$, then we have

$$\mathbb{P}\left(Y^{(M)} - \frac{MX^{(N)}}{N} \geq M\delta\right) \leq e^{-\frac{2M\delta^2}{1-(M-1)/N}}$$
$$< e^{-2M\delta^2}$$
$$\mathbb{P}\left(\left|Y^{(M)} - \frac{MX^{(N)}}{N}\right| > M\delta\right) \leq 2e^{-\frac{2M\delta^2}{1-(M-1)/N}}$$
$$< 2e^{-2M\delta^2}$$

where $\delta \in [0, 1]$ is some constant.[3]

Another issue that will be involved in our further analysis is to estimate the probability of the following events:

$$\forall t \in \mathbb{N}_0 : p_t(\mathbf{x}^*) \oplus \hat{p}_t(\mathbf{x}^*) \quad (15)$$

where $\oplus \in \{\leq, \geq\}$. As we will show soon, they can be handled on the basis of estimation of the probabilities of deviations. Finally, before presenting the case studies in detail, it should be noted that we always consider finite population sizes throughout this paper. Although we will sometimes utilize a statement like "when the problem size becomes sufficiently large," that does not mean that we assume infinite population

---

[3]The first inequality can be found in Corollary 1.1 in [38], or a similar form can be found in [21], and the second inequality is in (3.3) in [38].

sizes, it is merely used to obtain the asymptotic order of a function of the problem size n. The main difference is that the infinite population assumption implies infinite population sizes for all problem sizes (so that the random sampling errors are removed), while in our case the population size will be infinite only if the problem size has become infinite.

IV. WORST CASE ANALYSIS OF UMDA ON THE LEADINGONES PROBLEM

The first maximization problem we investigate is called the LEADINGONES problem, formally defined as follows:

$$\text{LEADINGONES}(\mathbf{x}) \triangleq \sum_{i=1}^{n} \prod_{j=1}^{i} x_j, \quad x_j \in \{0, 1\}. \quad (16)$$

The global optimum of LEADINGONES is $\mathbf{x}^* = (1, \ldots, 1)$. The fitness of an individual is determined by the number of the leading 1-bits in the individual, and it is not influenced by any bits right to the leftmost 0-bit of the individual. The value of the bits right to the leftmost 0-bit will not influence the output of fitness-based selection operators in EAs. Due to this characteristic, a population will begin to converge to 1 at a bit if the bits left to it have almost converged to 1's, and thus a sequential convergence phenomenon, namely *Domino convergence* [3], [36], [41], will happen.

In the literature of EDAs, the LEADINGONES problem has been investigated empirically [10], but no rigorous theoretical result exists. This section will provide the first theoretical result that put a sound foundation to the time complexity analysis of the UMDA on this problem.

First, we introduce the following concept.

Definition 1 (b-Promising Individual): In the population that contains N individuals, the b-promising individuals are those individuals with fitness no smaller than a threshold b.

Since the UMDA adopts the truncation selection, we have the following lemma.

Lemma 5: For the UMDA with truncation selection, the poportion of the b-promising individuals after selection at the tth generation satisfies

$$Q_{t,b}^{(s)} = \begin{cases} \frac{Q_{t,b}N}{M}, & Q_{t,b} \leq \frac{M}{N} \\ 1, & Q_{t,b} > \frac{M}{N} \end{cases} \quad (17)$$

where $Q_{t,b} \leq 1$ is the proportion of the b-promising individuals before the truncation selection.

Define the i-convergence time T_i to be the number of generations for a discrete EDA to converge to the globally optimal value on the ith bit of the solution. It is defined formally as

$$T_i \triangleq \min\{t; p_{t,i}(x_i^*) = 1\}.$$

Let $T_0 = 0$.

Moreover, in the following parts of the paper, we use the notation "ω" to demonstrate the relationship between the asymptotic orders of two functions [5], [24]. Given two positive functions of the problem size n, say $f = f(n)$ and $g = g(n)$, $f = \omega(g)$ holds if and only if $\lim_{n \to \infty} g(n)/f(n) = 0$. Now we reach the following theorem.

Theorem 2: Given the population sizes $N = \omega(n^{2+\alpha} \log n)$, $M = \omega(n^{2+\alpha} \log n)$ (where α can be any positive constant) and $M = \beta N$ ($\beta \in (0, 1)$ is some constant), for the UMDA with truncation selection on the LEADINGONES problem, initialized with a uniform distribution, at least with the probability of

$$\left(1 - n^{-\omega(n^{2+\alpha})\delta^2}\right)^{\bar{\tau}} \left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \omega(1)}\right)^{2(n-1)\bar{\tau}}$$

its FHT satisfies

$$\tau < \bar{\tau} = \frac{n\left(\ln \frac{eM}{N} - \ln(1-\delta)\right)}{\ln(1-\delta) + \ln\left(\frac{N}{M}\right)} + 2n$$

where $\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - \frac{M}{N})$ is a positive constant, and $\bar{\tau}$ represents an upper bound[4] of the random variable τ. In other words, the LEADINGONES problem is EDA-easy for the UMDA.

Proof: The basic idea of the proof is based on the approach outlined in the former section. We first de-randomize the UMDA. Since the LEADINGONES problem is associated with the domino convergence property, we can further divide the optimization process into n stages. The ith stage starts when all bits at the left side of the ith bit have converged to 1's, and ends when the ith bit has converged. Suppose generation $t + 1$ belongs to the ith stage, then the marginal probabilities at the generation are

$$\hat{\mathbf{P}}_{t+1}(\mathbf{x}^*) = \gamma_i(\hat{\mathbf{P}}_t(\mathbf{x}^*)) = \left(\hat{p}_{t,1}(x_1^*), \ldots, \hat{p}_{t,i-1}(x_{i-1}^*),\right.$$
$$\left.[G\hat{p}_{t,i}(x_i^*)], R\hat{p}_{t,i+1}(x_{i+1}^*), \ldots, R\hat{p}_{t,n}(x_n^*)\right)$$

where $\mathbf{x}^* = (x_1^*, \ldots, x_n^*) = (1, \ldots, 1)$ is the global optimum of the LEADINGONES problem, $G = (1 - \delta)\frac{N}{M}$ ($\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - \frac{M}{N})$ is a constant), and $R = (1 - \eta)(1 - \eta')$ ($\eta < 1$ and $\eta' < 1$ are positive functions of the problem size n). We consider three different cases in the above equation.

1) $j \in \{1, \ldots, i-1\}$. In the deterministic system above, the marginal probabilities $\hat{p}_{t,j}(x_j^*)$ have converged to 1, thus at the next generation they will not change.
2) $j = i$. In the deterministic system above, the marginal probability $\hat{p}_{t,i}(x_i^*)$ is converging, and we use the factor $G = (1 - \delta)\frac{N}{M}$ to demonstrate the impact of selection

pressure on this converging marginal probability,[5] where $\frac{N}{M}$ represents the influence of the selection operator (see Lemma 5).

3) $j \in \{i + 1, \ldots, n\}$. The jth bits of individuals are not exposed to selection pressure, and we use the factor $R = (1-\eta)(1-\eta')$ to demonstrate the impact of genetic drift[6] on these marginal probabilities.

In Case 3, we consider the jth marginal probability $p_{.,j}(x_j^*)$ ($j \in \{i + 1, \ldots, n\}$) which is not affected by the selection pressure. This is rather pessimistic, because the UMDA tends to preserve the value of $x_j^* = 1$ that leads to higher fitness, and thus tends to increase $p_{.,j}(x_j^*)$. Utilizing the idea mentioned in (15), we will study the time complexity of the UMDA by studying the above deterministic system, and estimate the deviation between the deterministic system and the real UMDA in terms of the probability that the stochastic marginal probabilities of the UMDA are bounded by the corresponding deterministic marginal probabilities of the deterministic system. Before our analysis, we first provide the formal definition of the deterministic system.

With $\hat{\mathbf{P}}_0(\mathbf{x}^*) = \left(\frac{1}{2}, \ldots, \frac{1}{2}\right)$, we have

$$\hat{\mathbf{P}}_t(\mathbf{x}^*) = \gamma_i^{t-T_{i-1}}\left(\hat{\mathbf{P}}_{T_{i-1}}(\mathbf{x}^*)\right)$$

where $T_{i-1} < t \leq T_i$ ($i = 1, \ldots, n$). Since $\{\gamma_i\}_{i=1}^n$ de-randomizes the whole optimization process, $\{T_i\}_{i=1}^n$ in the above equation are no longer random variables. For the sake of clarity, we rewrite the above equation as

$$\hat{\mathbf{P}}_t(\mathbf{x}^*) = \gamma_i^{t-\hat{T}_{i-1}}\left(\hat{\mathbf{P}}_{\hat{T}_{i-1}}(\mathbf{x}^*)\right)$$

where $\hat{T}_{i-1} < t \leq \hat{T}_i$ ($i = 1, \ldots, n$). As we will show immediately, $\hat{T}_i$ ($1 \leq i \leq n$) is an upper bound of the random variable T_i with some probability. Since $T_n \geq \tau$, our task finally becomes calculating the $\hat{T}_n$ and the probability that $\hat{T}_n$ holds as an upper bound of T_n.

Now we present the proof in detail. First, we estimate $\hat{T}_1$ and T_1 for the UMDA, which is the first stage of our analysis. Consider the 1-promising individuals. Note that the first bits of the 1-promising individuals are 1's. The sampling procedure of the UMDA can be considered as a large number of events resulting in either 0 or 1. Hence, when $p_{t-1,1}(1) \leq \frac{M}{N(1-\delta)}$, for the sampling procedure of the UMDA, by noting Lemma 5, we can apply Chernoff bounds to obtain the following:

$$\mathbb{P}\left(Mp_{t,1}(1) \geq (1-\delta)p_{t-1,1}(1)N \mid p_{t-1,1}(1) \leq \frac{M}{N(1-\delta)}\right)$$
$$> 1 - e^{-\frac{p_{t-1,1}(1)N}{2}\delta^2}$$

where $N = \omega(n^2 \log n)$, thus the probability above is super-polynomially close to 1, i.e., an overwhelming probability. An

---

[4]Given the values of the population sizes and the constant δ, the value of $\bar{\tau}$ is then determined by the problem size n. Thus, $\bar{\tau}$ is not a random variable.

[5]The notation "[ ]" can be interpreted as follows: given $a > 1$, $[a] = 1$; given $a \in (0, 1)$, $[a] = a$. For the sake of brevity, we will omit this notation but implicitly restrict the value of a probability not to exceed 1 in the following parts of the paper.

[6]When there is no selection pressure, the proportion of alleles in a population with finite genes will fluctuate due to the errors brought by random sampling. For more details, one can refer to [6], [41].

TABLE III
CALCULATION OF PROBABILITY THAT $p_{t,1}(1)$ IS LOWER BOUNDED BY $\hat{p}_{t,1}(1)$

$$\mathbb{P}\left(p_{t,1}(1) \geq \hat{p}_{t,1}(1) \mid p_{0,1}(1) = \hat{p}_{0,1}(1) \right)$$

$$= \sum_{\forall t' < t: a_{t'} \in \left\{0, \frac{1}{M}, \frac{2}{M}, \cdots, 1\right\}} \mathbb{P}\left(p_{t,1}(1) \geq G^t p_{0,1}(1), p_{t-1,1}(1) = a_{t-1}, \cdots, p_{1,1}(1) = a_1 \mid p_{0,1}(1) = \hat{p}_{0,1}(1) \right)$$

$$> \mathbb{P}\left(p_{t,1}(1) \geq G p_{t-1,1}(1), \cdots, p_{1,1}(1) \geq G p_{0,1}(1) \mid p_{0,1}(1) = \hat{p}_{0,1}(1) \right)$$

$$= \mathbb{P}\left(p_{t-1,1}(1) \geq G p_{t-2,1}(1), \cdots, p_{1,1}(1) \geq G p_{0,1}(1) \mid p_{0,1}(1) = \hat{p}_{0,1}(1) \right)$$

$$\mathbb{P}\left(p_{t,1}(1) \geq G p_{t-1,1}(1) \mid p_{t-1,1}(1) \geq G p_{t-2,1}(1), \cdots, p_{1,1}(1) \geq G p_{0,1}(1), p_{0,1}(1) = \hat{p}_{0,1}(1) \right)$$

$$= \mathbb{P}\left(p_{1,1}(1) \geq G p_{0,1}(1) \mid p_{0,1}(1) = \hat{p}_{0,1}(1) \right)$$

$$\prod_{k=2}^{t} \mathbb{P}\left(p_{k,1}(1) \geq G p_{k-1,1}(1) \mid p_{k-1,1}(1) \geq G p_{k-2,1}(1), \cdots, p_{1,1}(1) \geq G p_{0,1}(1), p_{0,1}(1) = \hat{p}_{0,1}(1) \right)$$

$$= \mathbb{P}\left(p_{1,1}(1) \geq G p_{0,1}(1) \mid p_{0,1}(1) = \hat{p}_{0,1}(1) \right)$$

$$\prod_{k=2}^{t} \mathbb{P}\left(p_{k,1}(1) \geq G p_{k-1,1}(1) \mid p_{k-1,1}(1) \geq \hat{p}_{k-1,1}(1) = G^{k-1} \hat{p}_{0,1}(1) \right)$$

$$> \prod_{k=1}^{t} \left(1 - e^{-\hat{p}_{k-1,1}(1) N \delta^2/2} \right) = \prod_{k=1}^{t} \left(1 - e^{-G^{k-1} \hat{p}_{0,1}(1) N \delta^2/2} \right) > \left(1 - e^{-\hat{p}_{0,1}(1) N \delta^2/2} \right)^t$$

TABLE IV
CALCULATION OF PROBABILITY THAT T_1 IS UPPER BOUNDED BY $\hat{T}_1$

$$\mathbb{P}\left(T_1 \leq \hat{T}_1 \mid p_{0,1}(1) = \hat{p}_{0,1}(1) \right) \tag{18}$$

$$> \mathbb{P}\left(p_{\hat{T}_1-1,1}(1) \geq \frac{M}{N(1-\delta)} \mid p_{0,1}(1) = \hat{p}_{0,1}(1) \right) \left(1 - e^{-\frac{\hat{p}_{0,1}(1)N}{2} \delta^2} \right) \tag{19}$$

$$> \mathbb{P}\left(p_{\hat{T}_1-1,1}(1) \geq \hat{p}_{\hat{T}_1-1,1}(1) = G^{\hat{T}_1-1} p_{0,1}(1) > \frac{M}{N(1-\delta)} \mid p_{0,1}(1) = \hat{p}_{0,1}(1) \right) \left(1 - e^{-\frac{\hat{p}_{0,1}(1)N}{2} \delta^2} \right)$$

$$> \mathbb{P}\left(p_{\hat{T}_1-1,1}(1) \geq \hat{p}_{\hat{T}_1-1,1}(1) \mid p_{0,1}(1) = \hat{p}_{0,1}(1), \hat{p}_{\hat{T}_1-1,1}(1) > \frac{M}{N(1-\delta)} \right)$$

$$\cdot \mathbb{P}\left(\hat{p}_{\hat{T}_1-1,1}(1) > \frac{M}{N(1-\delta)} \mid p_{0,1}(1) = \hat{p}_{0,1}(1) \right) \left(1 - e^{-\frac{\hat{p}_{0,1}(1)N}{2} \delta^2} \right)$$

TABLE V
BOUNDING $N_{t,j}^{(s)}(x_j^*)$ FROM BELOW WITH AN OVERWHELMING PROBABILITY

$$\mathbb{P}\left(N_{t,j}^{(s)}(x_j^*) > (1 - \eta') \frac{(1-\eta) p_{t-1,j}(x_j^*) N}{N} M \mid N_{t,j}(x_j^*) \geq \left(1 - \left(\frac{1}{n} \right)^{1+\frac{\alpha}{2}} \right) p_{t-1,j}(x_j^*) N, p_{t-1,j}(x_j) \right)$$

$$= \mathbb{P}\left(\frac{(1-\eta) p_{t-1,j}(x_j^*) N}{N} M - N_{t,j}^{(s)}(x_j^*) < \eta' \frac{(1-\eta) p_{t-1,j}(x_j^*) N}{N} M \mid N_{t,j}(x_j^*) \geq \left(1 - \left(\frac{1}{n} \right)^{1+\frac{\alpha}{2}} \right) p_{t-1,j}(x_j^*) N, p_{t-1,j}(x_j^*) \right)$$

$$> 1 - 2 e^{-2(1-\eta)^2 p_{t-1,j}^2(x_j^*) \eta'^2 M}$$

TABLE VI
CALCULATION OF THE JOINT PROBABILITY THAT T_1 IS BOUNDED ABOVE BY $\hat{T}_2$

$$\mathbb{P}\left(T_2 \leq \hat{T}_2, T_1 \leq \hat{T}_1, p_{\hat{T}_1,2}(1) \geq \hat{p}_{\hat{T}_1,2}(1) > \frac{1}{e} \mid p_{0,1}(1) = \hat{p}_{0,1}(1), p_{0,2}(1) = \hat{p}_{0,2}(1)\right) \tag{20}$$

$$> \mathbb{P}\left(p_{\hat{T}_2-1,2}(1) \geq \frac{M}{N(1-\delta)} \mid p_{0,1}(1) = \hat{p}_{0,1}(1), p_{0,2}(1) = \hat{p}_{0,2}(1), T_1 \leq \hat{T}_1, p_{\hat{T}_1,2}(1) \geq \hat{p}_{\hat{T}_1,2}(1) > \frac{1}{e}\right)$$

$$\cdot \left(1 - e^{-\frac{\omega(n^{2+\alpha}\log n)}{2e}\delta^2}\right)^{\hat{T}_1}\left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2\omega(1)}\right)^{2\hat{T}_1}\left(1 - e^{-\frac{\hat{p}_{\hat{T}_1,2}(1)N}{2}\delta^2}\right) \tag{21}$$

$$> \mathbb{P}\left(p_{\hat{T}_2-1,2}(1) \geq \hat{p}_{\hat{T}_2-1,2}(1) = G^{\hat{T}_2-\hat{T}_1-1}p_{\hat{T}_1,2}(1) > \frac{M}{N(1-\delta)} \mid p_{\hat{T}_1,1}(1) = 1, p_{\hat{T}_1,2}(1) \geq \hat{p}_{\hat{T}_1,2}(1) > \frac{1}{e}\right)$$

$$\cdot \left(1 - e^{-\frac{\omega(n^{2+\alpha}\log n)}{2e}\delta^2}\right)^{\hat{T}_1}\left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2\omega(1)}\right)^{2\hat{T}_1}\left(1 - e^{-\frac{\hat{p}_{\hat{T}_1,2}(1))N}{2}\delta^2}\right)$$

$$> \mathbb{P}\left(p_{\hat{T}_2-1,2}(1) \geq \hat{p}_{\hat{T}_2-1,2}(1) \mid p_{\hat{T}_1,1}(1) = 1, p_{\hat{T}_1,2}(1) \geq \hat{p}_{\hat{T}_1,2}(1) > \frac{1}{e}, \hat{p}_{\hat{T}_2-1,2}(1) > \frac{M}{N(1-\delta)}\right) \tag{22}$$

$$\mathbb{P}\left(\hat{p}_{\hat{T}_2-1,2}(1) > \frac{M}{N(1-\delta)} \mid p_{\hat{T}_1,1}(1) = 1, p_{\hat{T}_1,2}(1) \geq \hat{p}_{\hat{T}_1,2}(1) > \frac{1}{e}\right)$$

$$\cdot \left(1 - e^{-\frac{\omega(n^{2+\alpha}\log n)}{2e}\delta^2}\right)^{\hat{T}_1}\left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2\omega(1)}\right)^{2\hat{T}_1}\left(1 - e^{-\frac{\omega(n^{2+\alpha}\log n)}{2e}\delta^2}\right)$$

TABLE VII
BOUNDING $N_{t,q}^{(s)}(x_q^*)$ FROM ABOVE WITH AN OVERWHELMING PROBABILITY

$$\mathbb{P}\left(N_{t,q}^{(s)}(x_q^*) < (1+\eta')\frac{(1+\eta)p_{t-1,q}(x_q^*)N}{N}M \mid N_{t,q}(x_q^*) \leq \left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)p_{t-1,q}(x_q^*)N, p_{t-1,q}(x_q^*)\right)$$

$$= \mathbb{P}\left(N_{t,q}^{(s)}(x_q^*) - \frac{(1+\eta)p_{t-1,q}(x_q^*)N}{N}M < \eta'\frac{(1+\eta)p_{t-1,q}(x_q^*)N}{N}M \mid N_{t,q}(x_q^*) \leq \left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)p_{t-1,q}(x_q^*)N, p_{t-1,q}(x_q^*)\right)$$

$$> 1 - e^{-2(1+\eta)^2 p_{t-1,q}^2(x_q^*)\eta'^2 M} \tag{23}$$

equivalent form of the equation above is

$$\mathbb{P}\left(p_{t,1}(1) \geq (1-\delta)\frac{p_{t-1,1}(1)N}{M} \mid p_{t-1,1}(1) \leq \frac{M}{N(1-\delta)}\right)$$
$$> 1 - e^{-\frac{p_{t-1,1}(1)N}{2}\delta^2}$$

which demonstrates with an overwhelming probability the marginal probability $p_{t,1}(1)$ is lower bounded by $Gp_{t-1,1}(1) = (1-\delta)\frac{p_{t-1,1}(1)N}{M}$. Furthermore, given $\hat{p}_{t,1}(1) = G^t\hat{p}_{0,1}(1)$ and $G > 1$, we can obtain the inequality in Table III.

We now study the distribution of T_1. Considering the probability that T_1 is bounded by a value, say $\hat{T}_1$: given $T_1 < \hat{T}_1$, then according to Lemma 5, at the $(\hat{T}_1 - 1)$th generation, the marginal probability $p_{\hat{T}_1-1,1}(1)$ should be at least $\frac{M}{N(1-\delta)}$. The above proposition is presented in Table IV, where in (19) the factor $\left(1 - e^{-\frac{\hat{p}_{0,1}(1)N}{2}\delta^2}\right)$ is added since we apply Chernoff bounds once at the end of the $(\hat{T}_1 - 1)$th generation and obtain the probability that $\hat{p}_{\hat{T}_1,1}(1) = 1$, under the condition $\hat{p}_{\hat{T}_1-1,1}(1) \geq \frac{M}{N(1-\delta)}$. Now let us consider the following item. Noting that $\hat{p}_{\hat{T}_1-1,1}(1)$ is deterministic, we

know

$$\mathbb{P}\left(\hat{p}_{\hat{T}_1-1,1}(1) > \frac{M}{N(1-\delta)} \mid p_{0,1}(1) = \hat{p}_{0,1}(1)\right) \tag{24}$$

must be either 0 or 1, and we need to find the value of $\hat{T}_1$ that makes the probability above 1. Given that $\hat{p}_{0,1}(1) = \frac{1}{2}$, the condition that $\forall t < \hat{T}_1 - 1 : \frac{M}{N(1-\delta)} > \hat{p}_{t,1}(1) > (1-\delta)\frac{\hat{p}_{t-1,1}(1)N}{M}$ and Lemma 5 together imply the following inequalities.

$$G^{\hat{T}_1-2}\hat{p}_{0,1}(1) = (1-\delta)^{\hat{T}_1-2}\left(\frac{N}{M}\right)^{\hat{T}_1-2}\hat{p}_{0,1}(1)$$
$$< \frac{M}{N(1-\delta)}$$
$$G^{\hat{T}_1-1}\hat{p}_{0,1}(1) = (1-\delta)^{\hat{T}_1-1}\left(\frac{N}{M}\right)^{\hat{T}_1-1}\hat{p}_{0,1}(1)$$
$$\geq \frac{M}{N(1-\delta)}.$$

Solving the inequalities above, we get

$$\hat{T}_1 \leq \frac{\ln\frac{2M}{N} - \ln(1-\delta)}{\ln(1-\delta) + \ln\left(\frac{N}{M}\right)} + 2$$

where $\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - \frac{M}{N})$ is a constant, and it is easy to show that $\hat{T}_1 = \Theta(1)$. On the other hand, recall the inequalities in Table III, we can continue to estimate the corresponding probability mentioned in (18)

$$\mathbb{P}\left(T_1 \leq \hat{T}_1 \mid p_{0,1}(1) = \hat{p}_{0,1}(1)\right)$$

$$> \mathbb{P}\left(p_{\hat{T}_1-1,1}(1) \geq \hat{p}_{\hat{T}_1-1,1}(1) \mid p_{0,1}(1) = \hat{p}_{0,1}(1)\right)$$

$$\cdot \left(1 - e^{-\frac{\hat{p}_{0,1}(1)N}{2}\delta^2}\right)$$

$$> \left(1 - e^{-\frac{\hat{p}_{0,1}(1)N}{2}\delta^2}\right)^{\hat{T}_1}. \tag{25}$$

The analysis above tells us, the probability to which the marginal probability converges before the $\hat{T}_1$th generation $(T_1 < \hat{T}_1)$ is at least $\left(1 - e^{-\frac{N}{4}\delta^2}\right)^{\hat{T}_1}$. Since $N = \omega(n^{2+\alpha} \log n)$, $M = \beta N$ ($\beta \in (0, 1)$ is a constant) and $\hat{T}_1$ is polynomial in the problem size n, we know that the probability is overwhelming.

At every stage, the bits on the right-hand side of the currently converging bit are not exposed to selection pressure. However, we should still consider the errors brought by the repeated sampling procedures in UMDA, which is related to the genetic drift [6], [41].

Take the first stage as an example. The jth bit ($j = 2, \ldots, n$) is affected by genetic drift. First, we utilize Chernoff bounds to study the deviations brought by the random sampling procedures of the UMDA

$$\mathbb{P}\left(N_{t,j}(x_j^*) \geq (1 - \eta)p_{t-1,j}(x_j^*)N \mid p_{t-1,j}(x_j^*)\right)$$

$$> 1 - e^{-\frac{p_{t-1,j}(1)N}{2}\eta^2}$$

where η is a parameter that controls the size of deviation, and $N_{t,j}(x_j)$ is the number of individuals that takes the value x_j in their jth bit in the population before selection, ξ_t. Here we set $\eta = \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}$, and obtain

$$\mathbb{P}\left(N_{t,j}(x_j^*) \geq \left(1 - \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)p_{t-1,j}(x_j^*)N \mid p_{t-1,j}(x_j^*)\right)$$

$$> 1 - e^{-\frac{p_{t-1,j}(x_j^*)\omega(\log n)}{2}} = 1 - n^{-\frac{p_{t-1,j}(x_j^*)\omega(1)}{2}}.$$

Second, we further consider the selection procedure, since it may also bring some deviations. In our worst case analysis, the jth bits of individuals are considered to not be exposed to the selection pressure, then for these bits the selection procedure can be regarded as get a simple random sample of M individuals from a finite population with N individuals [34]. More precisely, since one individual cannot be selected more than once by the truncation selection, this procedure is known as random sampling *without* replacement from a finite population [34] in the field of statistics. From Lemma 4, we can bound from below the probability such that the number of individuals taking the value x_j^* on their jth bits after selection [denoted by $N_{t,j}^{(s)}(x_j^*)$] is lower bounded, which is shown by the inequalities in Table V, where η' is a parameter that controls the size of deviation, and $N_{t,j}^{(s)}(x_j^*) = p_{t,j}(x_j^*)M$. By

setting $\eta' = \eta = \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}$, since $M = \omega(n^{2+\alpha} \log n)$ we obtain

$$\mathbb{P}\left(p_{t,j}(x_j^*) \geq \left(1 - \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 p_{t-1,j}(x_j^*) \mid p_{t-1,j}(x_j^*)\right)$$

$$> \left(1 - n^{-p_{t-1,j}(x_j^*)\omega(1)}\right)$$

$$\cdot \left(1 - n^{-\left(1-\left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 p_{t-1,j}^2(x_j^*)\omega(1)}\right)$$

$$> \left(1 - n^{-\left(1-\left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 p_{t-1,j}^2(x_j^*)\omega(1)}\right)^2.$$

Since the factor $R = \left(1 - \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 < 1$, for $\forall j = 2 \ldots, n$ and $t = 1, \ldots, \hat{T}_1$, similar to the analysis shown in Table III, we further obtain

$$\mathbb{P}\left(p_{t,j}(x_j^*) \geq \left(1 - \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^{2t} p_{0,j}(x_j^*)\right.$$

$$\left. \mid p_{0,j}(x_j^*) = \hat{p}_{0,j}(x_j^*)\right)$$

$$> \left(1 - n^{-\left(1-\left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 \hat{p}_{t-1,j}^2(x_j^*)\omega(1)}\right)^{2t}. \tag{26}$$

Given any $t = O(n)$, according to the definition of the deterministic system, we know

$$\hat{p}_{t,j}(x_j^*) \geq \left(1 - \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^{O(n)} \hat{p}_{0,j}(x_j^*) > \frac{1}{e}$$

holds. The above inequality implies that within the number of generations $t = O(n)$, the probability in (26) is an overwhelming one.

To generalize the above analysis to other stages, let us consider the ith ($i \in \{2, \ldots, n\}$) stage is about to start. Due to the genetic drift, the marginal probability $p_{t,j}(x_j^*)$ ($j \in \{i, \ldots, n\}$) has dropped to a lower level than the initial value $\frac{1}{2}$ by multiplying the factor R^t. We concern the value of $p_{t,i}(x_i^*)$. For any $t = O(n)$, similar to (26), the probability that $p_{t,i}(x_i^*)$ maintains a level of

$$p_{t,i}(x_i^*) \geq \left(1 - \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^{O(n)} \hat{p}_{0,i}(x_i^*) > \frac{1}{e} \tag{27}$$

is super-polynomially close to 1 (an overwhelming probability).

According to (27), we know that $p_{t,i}(x_i^*)$ is above $\frac{1}{e}$ with an overwhelming probability. Consequently, the joint probability that the first bit has converged to 1 and the genetic drift cannot reduce $p_{\hat{T}_1,2}(1)$ to be smaller than $\frac{1}{e}$ by the end of the first stage is

$$\left(1 - e^{-\frac{\omega(n^{2+\alpha} \log n)}{2e}\delta^2}\right)^{\hat{T}_1} \left(1 - n^{-\left(1-\left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 \omega(1)}\right)^{2\hat{T}_1} \tag{28}$$

which is again an overwhelming probability. Now we have finished the analysis of the first stage.

As the dynamic system we described at the beginning of the proof, in the second stage, for $\hat{T}_1 < t \leq \hat{T}_2$, we have

$$\hat{p}_{t,2}(1) = G\hat{p}_{t-1,2}(1).$$

Given $\hat{T}_1$ and the corresponding marginal probabilities, we consider the joint probability that T_2 is bounded above by $\hat{T}_2$ by inequalities presented in Table VI.

Let us consider the following item of the probability estimated in Table VI:

$$\mathbb{P}\left(\hat{p}_{\hat{T}_2-1,2}(1) > \frac{M}{N(1-\delta)} \mid p_{\hat{T}_1,1}(1) = 1, \right.$$

$$\left. p_{\hat{T}_1,2}(1) \geq \hat{p}_{\hat{T}_1,2}(1) > \frac{1}{e} \right)$$

since $\{\hat{p}_{t,2}(1)\}_{t=0}^{\infty}$ is a deterministic sequence, the above item must be either 0 or 1. Noting that $\hat{p}_{\hat{T}_1,2}(1) > \frac{1}{e}$, given the condition that $\forall t : \hat{T}_1 < t < \hat{T}_2 - 1 : \frac{M}{N(1-\delta)} > \hat{p}_{t,2}(1) = (1-\delta)\frac{\hat{p}_{t-1,2}(1)N}{M}$, we can solve the following inequalities to obtain $\hat{T}_2$

$$G^{\hat{T}_2-\hat{T}_1-2}\hat{p}_{\hat{T}_1,2}(1)$$

$$= \left((1-\delta)\left(\frac{N}{M}\right) \right)^{\hat{T}_2-\hat{T}_1-2} \hat{p}_{\hat{T}_1,2}(1) < \frac{M}{N(1-\delta)}$$

$$G^{\hat{T}_2-\hat{T}_1-1}\hat{p}_{\hat{T}_1,2}(1)$$

$$= \left((1-\delta)\left(\frac{N}{M}\right) \right)^{\hat{T}_2-\hat{T}_1-1} \hat{p}_{\hat{T}_1,2}(1) \geq \frac{M}{N(1-\delta)}.$$

Moreover, another item in (22)

$$\mathbb{P}\left(p_{\hat{T}_2-1,2}(1) \geq \hat{p}_{\hat{T}_2-1,2}(1) \mid p_{\hat{T}_1,1}(1) = 1, \right.$$

$$\left. p_{\hat{T}_1,2}(1) \geq \hat{p}_{\hat{T}_1,2}(1) > \frac{1}{e}, \hat{p}_{\hat{T}_2-1,2}(1) > \frac{M}{N(1-\delta)} \right)$$

should be estimated. This can be done similarly as we have done in Table III. Then we obtain that

$$T_2 < \hat{T}_2 \leq \frac{2\ln\frac{eM}{N} - 2\ln(1-\delta)}{\ln(1-\delta) + \ln\left(\frac{N}{M}\right)} + 4$$

holds with the probability [the product of the items mentioned in (22)]

$$\left(1 - e^{-\frac{\omega(n^{2+\alpha}\log n)}{2e}\delta^2} \right)^{\hat{T}_2} \left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2\omega(1)} \right)^{2\hat{T}_1}.$$

The above analysis can be readily extended to other stages. To be specific, at the ith stage, the i-promising individuals are taken into account. We have

$$\hat{p}_{t,i}(1) = G\hat{p}_{t-1,i}(1).$$

For induction, assume that at the $(i-1)$th stage

$$T_{i-1} < \hat{T}_{i-1} \leq \frac{(i-1)\ln\frac{eM}{N} - (i-1)\ln(1-\delta)}{\ln(1-\delta) + \ln\left(\frac{N}{M}\right)}$$
$$+ 2(i-1) \quad (29)$$

holds with the probability

$$\left(1 - e^{-\frac{\omega(n^{2+\alpha}\log n)}{4}\delta^2} \right)^{\hat{T}_{i-1}}$$

$$\cdot \prod_{k=1}^{i-2} \left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2\omega(1)} \right)^{2\hat{T}_k}.$$

To estimate $\hat{T}_i$, we solve the following inequalities:

$$G^{\hat{T}_i-\hat{T}_{i-1}-2}\hat{p}_{\hat{T}_{i-1},i}(1)$$

$$= (1-\delta)^{\hat{T}_i-\hat{T}_{i-1}-2}\left(\frac{N}{M}\right)^{\hat{T}_i-\hat{T}_{i-1}-2} \hat{p}_{\hat{T}_{i-1},i}(1)$$

$$< \frac{M}{N(1-\delta)}$$

$$G^{\hat{T}_i-\hat{T}_{i-1}-1}\hat{p}_{\hat{T}_{i-1},i}(1)$$

$$= (1-\delta)^{\hat{T}_i-\hat{T}_{i-1}-1}\left(\frac{N}{M}\right)^{\hat{T}_i-\hat{T}_{i-1}-1} \hat{p}_{\hat{T}_{i-1},i}(1)$$

$$\geq \frac{M}{N(1-\delta)}$$

where $\hat{p}_{\hat{T}_{i-1},i}(1) > \frac{1}{e}$ [similar to (27)], since $\hat{T}_{i-1} = O(n)$ [our assumption for induction in (29) shows that it is $O(n)$]. Similar to the discussion at the second stage, we can get that

$$T_i < \hat{T}_i \leq \frac{i\ln\frac{eM}{N} - i\ln(1-\delta)}{\ln(1-\delta) + \ln\left(\frac{N}{M}\right)} + 2i$$

holds with the probability

$$\left(1 - e^{-\frac{\omega(n^{2+\alpha}\log n)}{2e}\delta^2} \right)^{\hat{T}_i}$$

$$\cdot \prod_{k=1}^{i-1} \left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2\omega(1)} \right)^{2\hat{T}_k}.$$

Finally, the FHT τ is upper bounded by

$$\tau < \hat{T}_n = \frac{n\left(\ln\frac{eM}{N} - \ln(1-\delta)\right)}{\ln(1-\delta) + \ln\left(\frac{N}{M}\right)} + 2n$$

with a probability of

$$\left(1 - e^{-\frac{\omega(n^{2+\alpha}\log n)}{4}\delta^2} \right)^{\hat{T}_n}$$

$$\cdot \prod_{k=1}^{n-1} \left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2\omega(1)} \right)^{2\hat{T}_k}$$

$$> \left(1 - n^{-\omega(n^{2+\alpha})\delta^2} \right)^{\hat{T}_n}$$

$$\cdot \left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2\omega(1)} \right)^{2(n-1)\hat{T}_n}$$

which is an overwhelming probability. ∎

In the proof above, we have proven that a bound holds for the FHT with an overwhelming probability. Furthermore, the proof also shows the convergence of UMDA on LEADINGONES: the UMDA will converge to the optimum with an overwhelming probability. The convergence property is ensured by using population sizes of $\omega(n^{2+\alpha}\log n)$, and considering all the random sampling errors in the pessimistic way.

V. BEST CASE ANALYSIS OF UMDA ON THE BVLEADINGONES PROBLEM

The previous section has shown that the LEADINGONES problem is EDA-easy for the UMDA. In this section, we will study another *maximization* problem that is unimodal but EDA-hard for the UMDA. The problem, which is called BVLEADINGONES (BVLO for short), can be regarded as the

LEADINGONES problem with one bit's variation. It is defined as follows:

$$\text{BVLO}(\mathbf{x}) = \begin{cases} \text{LO}(\mathbf{x}) + n, & \text{LO}(\mathbf{x}) \le n - 1, x_n = 0 \\ \text{LO}(\mathbf{x}), & \text{LO}(\mathbf{x}) < n - 1, x_n = 1 \\ 3n, & \text{LO}(\mathbf{x}) = n \end{cases} \quad (30)$$

where $\forall i = 1, \ldots, n : x_i \in \{0, 1\}$ and LO stands for LEADINGONES. The BVLEADINGONES is a unimodal function whose global optimum is $\mathbf{x}^* = (x_1^*, \ldots, x_n^*) = (1, \ldots, 1)$. In this section, we will prove that BVLEADINGONES is EDA-hard for the UMDA.

Let us look at (30) again. The nth bits of the individuals are exposed to the selection pressure from the very beginning. During the optimization process, an individual whose last bit is 0 always has higher fitness than any individuals with its last bit being 1, unless the first $n - 1$ bits of the latter are all 1's. In other words, the nth marginal probability $p_{.,n}(\bar{x}_n^*)$ starts converging to 1 from the beginning of optimization, where $\bar{x}_n^* = 1 - x_n^* = 0$. Once $p_{.,n}(\bar{x}_n^*)$ reaches 1, the UMDA will miss the global optimum forever. Therefore, we need to check whether an individual whose first $n - 1$ bits are all 1's can be generated before $p_{.,n}(\bar{x}_n^*)$ reaches 1.

We start from analyzing the converging speed of the first $n - 1$ bits of individuals, given polynomial population sizes $M = \omega(n^{2+\alpha} \log n)$, $N = \omega(n^{2+\alpha} \log n)$ (where α can be any positive constant), and $M = \beta N$ ($\beta \in (0, 1)$ is some constant) for the UMDA. These bits can be classified into two categories. The first category is exposed to the selection pressure, and the second one is affected by the genetic drift. Unlike the previous section, here we analyze from an optimistic viewpoint: all bits of the first category will converge in one generation, and the genetic drift will promote the marginal probabilities of generating the optimal value on the remaining bits. We first consider the genetic drift of a typical marginal probability, say $p_{.,q}(x_q^*)$ (the qth bits belong to the second category). Using Chernoff bounds to study the deviations brought by the random sampling procedures, we have

$$\mathbb{P}\left(N_{t,q}(x_q^*) \le (1 + \eta) p_{t-1,q}(x_q^*) N \mid p_{t-1,q}(x_q^*) \right)$$
$$> 1 - e^{-\frac{p_{t-1,q}(x_q^*) N}{4} \eta^2}$$

where η is a parameter that controls the size of deviation, and $N_{t,q}(x_q^*)$ is the number of individuals that takes the value x_q^* in their qth bit in the population before selection. Set $\eta = \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}$, we obtain

$$\mathbb{P}\left(N_{t,q}(x_q^*) \le \left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right) p_{t-1,q}(x_q^*) N \right.$$
$$\left. \mid p_{t-1,q}(x_q^*) \right)$$
$$> 1 - e^{-\frac{p_{t-1,q}(x_q^*) \omega(\log n)}{4}} = 1 - n^{-\frac{p_{t-1,q}(x_q^*) \omega(1)}{4}}.$$

The selection procedure may also bring some deviations. Since the qth bits of individuals are not exposed to the selection pressure, then for these bits the selection procedure can be regarded as Simple Random Sampling without replacement.

Lemma 4 can be used to estimate the probability that the number of individuals taking the value x_q^* on their qth bits after selection [denoted by $N_{t,j}^{(s)}(x_q^*)$] is bounded from above, which is lower bounded by $1 - e^{-2(1+\eta)^2 p_{t-1,q}^2(x_q^*) \eta'^2 M}$ estimated by (23) in Table VII, where η' is a parameter that controls the size of deviation, and $N_{t,q}^{(s)}(x_q^*) = p_{t,q}(x_q^*) M$. Let $\eta' = \eta = \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}$, since $M = \omega(n^{2+\alpha} \log n)$ we get

$$\mathbb{P}\left(p_{t,q}(x_q^*) \le \left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 p_{t-1,q}(x_q^*) \mid p_{t-1,q}(x_q^*) \right)$$
$$> \left(1 - n^{-p_{t-1,q}(x_q^*)\omega(1)}\right)$$
$$\cdot \left(1 - n^{-\left(1+(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 p_{t-1,q}^2(x_q^*)\omega(1)}\right)$$
$$> \left(1 - n^{-\left(1+(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 p_{t-1,q}^2(x_q^*)\omega(1)}\right)^2.$$

Since $R = \left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 > 1$ (thus we know that $\hat{p}_{t-1,q}(x_q^*) > \hat{p}_{0,q}(x_q^*)$ in the above inequality), similar to the analysis shown in Table III, we further have

$$\mathbb{P}\left(p_{t,q}(x_q^*) \le \left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^{2t} p_{0,q}(x_q^*) \right.$$
$$\left. \mid p_{0,q}(x_q^*) = \hat{p}_{0,q}(x_q^*) \right)$$
$$> \left(1 - n^{-\left(1+(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \hat{p}_{0,q}^2(x_q^*)\omega(1)}\right)^{2t}.$$

Given any polynomial t, the above probability is an overwhelming one. Specifically, $\forall t = O(n)$, $p_{t,q}(x_q^*)$ is upper bounded as

$$p_{t,q}(x_q^*) \le \left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^{O(n)} \hat{p}_{0,q}(x_q^*)$$
$$= \frac{1}{2} + \Theta\left(\frac{1}{n^{\alpha/2}}\right) + o\left(\frac{1}{n^{\alpha/2}}\right) < c < 1 \quad (31)$$

with an overwhelming probability (where c is some positive constant, and the qth bits are not exposed to the selection pressure).

Another key issue of our analysis is the time T_n' for the nth marginal probability $p_{.,n}(\bar{x}_n^*)$ to converge to 1. We can prove the following lemma.

Lemma 6: The number of generations required by the marginal probability $p_{.,n}(\bar{x}_n^*)$ to converge to 1, i.e. T_n', is upper bounded by

$$U = \frac{\ln \frac{2M}{N} - \ln(1 - \delta)}{\ln(1 - \delta) + \ln\left(\frac{N}{M}\right)} + 2$$

with an overwhelming probability, if no global optimum is generated before the Uth generation, where $\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - \frac{M}{N})$ is a positive constant.

The proof is provided in the Appendix. Given polynomial population sizes $M = \omega(n^{2+\alpha} \log n)$, $N = \omega(n^{2+\alpha} \log n)$ (where α can be any positive constant), and $M = \beta N$ ($\beta \in (0, 1)$ is some constant), Lemma 6 implies that $U = \Theta(1)$. Now we reach the following theorem.

Theorem 3: Given polynomial population sizes $M = \omega(n^{2+\alpha} \log n)$, $N = \omega(n^{2+\alpha} \log n)$ (where α can be any positive constant), and $M = \beta N$ ($\beta \in (0, 1)$ is some constant), the FHT

of the UMDA with truncation selection on the BVLEADIN-GONES problem is infinity with an overwhelming probability. In other words, the UMDA with truncation selection cannot find the optimum of the BVLEADINGONES problem with an overwhelming probability.

Proof: We have proven that the number of generations required for $p_{.,n}(\bar{x}_n^*)$ to reach 1 (denoted by T_n') is upper bounded by a constant function U with an overwhelming probability, under the condition that no global optimum is generated before the Uth generation. We now further prove that the probability that no global optimum is generated before the Uth generation is also overwhelming.

As mentioned before, we classify the first $n - 1$ bits of individuals into two categories. The first category, which contains the bits being exposed to the selection, further contains two types of bits. The first type contains the bits which have already converged to the optimal values, and the second type contains the bits that are exposed to the selection pressure but have not converged to the optimal values yet. In our best case analysis, for the bits of the second type, we consider that only one generation is needed for the corresponding marginal probabilities (to the optimal values) to converge. In other words, before the Uth generation, the marginal probabilities (of the first $n - 1$ bits of individuals) are either 1 or no more than the constant c. Noting that $U = \Theta(1)$, according to (31), $c \in (\frac{1}{2}, 1)$, and it demonstrates the result of genetic drift within $O(n)$ generations.

From an optimistic viewpoint, we further consider that in every generation, besides the marginal probability $p_{.,n}(\bar{x}_n^*)$, at most $\log^2 n$ other marginal probabilities[7] are also converging with an overwhelming probability. $\log^2 n$ is used here because the joint probability of generating $\log^2 n$ consecutive 1's (so as to produce the selection pressure on the corresponding bits) by $\log^2 n$ non-converged marginal probabilities is no more than $c^{\log^2 n}$, which is super-polynomially small.

The above result implies that the probability of generating the global optimum in one generation is also super-polynomially small. Noting that $U = \Theta(1)$, then the probability of generating the optimum before the Uth generation is also super-polynomially small. Combining this probability with the conditional probability mentioned in Lemma 6, we know that the joint probability that no global optimum is generated before the Uth generation, and $p_{.,n}(\bar{x}_n^*)$ converges to 1 no later than the Uth generation, is super-polynomially close to 1, i.e., an overwhelming probability. Combining with the fact that once the nth marginal probability $p_{.,n}(x_n^*)$ has already converged to 0, the probability of finding the optimum will drop to 0, we have proven the theorem.

According to Theorem 1, given polynomial population sizes $M = \omega(n^{2+\alpha} \log n)$ and $N = \omega(n^{2+\alpha} \log n)$ ($M = \beta N$, $\beta \in (0, 1)$ is a constant.), BVLEADINGONES is EDA-hard for the UMDA. ∎

For the sake of consistence, we also provide the formal description of the deterministic dynamic system utilized in this section. Considering the ith stage ($i \leq \min\{T_n', \frac{n-1}{\log^2 n}\}$)

[7]For the sake of brevity, we assume that $\log^2 n$ is an integer and thus omit the notation "⌈ ⌉."

which starts when all the marginal probabilities $p_{.,k}(x_k^*)$ ($k \leq (i - 1)\log^2 n$) have just converged to 1 and ends when all the marginal probabilities $p_{.,j}(x_j^*)$ ($j \leq i\log^2 n$) have just converged to 1, we can obtain $\hat{\mathbf{P}}_{t+1}(\mathbf{x}^*)$ by defining γ_i as follows.

$$\hat{\mathbf{P}}_{t+1}(\mathbf{x}^*) = \gamma_i(\hat{\mathbf{P}}_t(\mathbf{x}^*)) =$$

$$\Big(\hat{p}_{t,1}(x_1^*), \ldots, \hat{p}_{t,(i-1)\log^2 n}(x_{(i-1)\log^2 n}^*), 1, \ldots, 1,$$

$$R\hat{p}_{t,i\log^2 n+1}(x_{i\log^2 n+1}^*), \ldots, R\hat{p}_{t,n-1}(x_{n-1}^*),$$

$$1 - G(1 - \hat{p}_{t,n}(x_n^*))\Big)$$

where $R = (1 + \eta)(1 + \eta')$ ($\eta < 1$ and $\eta' < 1$ are positive functions of the problem size n), and $G = (1 - \delta)\frac{N}{M}$ ($\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - \frac{M}{N})$ is a constant). In the above equation, we consider four different cases.

1) $j \in \{1, \ldots, (i - 1)\log^2 n\}$. In the deterministic system above, the marginal probabilities $\hat{p}_{t,j}(x_j^*)$ have converged to 1, thus at the next generation they will not change.
2) $j \in \{(i - 1)\log^2 n + 1, \ldots, i\log^2 n\}$. In the deterministic system above, the marginal probabilities $\hat{p}_{t,j}(x_j^*)$ are converging to the optimum, and they will converge in one generation in the best case analysis.
3) $j \in \{i\log^2 n + 1, \ldots, n - 1\}$. The jth bits of individuals are not exposed to selection pressure, and we use the factor $R = (1 + \eta)(1 + \eta')$ to demonstrate the impact of genetic drift in the deterministic system above.
4) $j = n$. The marginal probability $\hat{p}_{t,n}(\bar{x}_n^*) = 1 - \hat{p}_{t,n}(x_n^*)$ is converging, and we use the factor $G = (1 - \delta)\frac{N}{M}$ to demonstrate the impact of selection pressure on this converging marginal probability in the deterministic system above, which is a best case style for $\hat{p}_{t,n}(x_n^*)$.

With $\hat{\mathbf{P}}_0(\mathbf{x}^*) = \left(\frac{1}{2}, \ldots, \frac{1}{2}\right)$, noting that one stage actually refers to one generation (thus $i = t$), we have

$$\hat{\mathbf{P}}_t(\mathbf{x}^*) = \gamma_t \circ \gamma_{t-1} \ldots \circ \gamma_1\Big(\hat{\mathbf{P}}_0(\mathbf{x}^*)\Big)$$

where $t \leq \min\left\{T_n', \frac{n-1}{\log^2 n}\right\}$. Since $\{\gamma_i\}_{i=1}^t$ de-randomizes the whole optimization process, T_n' in the above equation is no longer random variable. For the sake of clarity, we rewrite the above equation as

$$\hat{\mathbf{P}}_t(\mathbf{x}^*) = \gamma_t \circ \gamma_{t-1} \ldots \circ \gamma_1\Big(\hat{\mathbf{P}}_0(\mathbf{x}^*)\Big)$$

where $t \leq \min\left\{\hat{T}_n', \frac{n-1}{\log^2 n}\right\} \leq \min\left\{U, \frac{n-1}{\log^2 n}\right\}$.

VI. A MODIFIED UMDA: RELAXATION by MARGINS

So far we have seen both EDA-easy and EDA-hard problems for the UMDA. This section will analyze more in-depth the relationship between EDA-hardness and the algorithms. The BVLEADINGONES problem, which has proven to be EDA-hard for the UMDA with finite populations, will be employed as the target problem in this section. We will show that a simple "relaxed" version of UMDA with truncation

selection can solve the BVLEADINGONES problem efficiently. The "relaxation" is implemented by adding some "margins" to the marginal probabilities of the UMDA. That is, the highest level the marginal probabilities can reach is $1 - \frac{1}{M}$ and the lowest level the marginal probabilities can drop to is $\frac{1}{M}$. Any marginal probabilities higher than $1 - \frac{1}{M}$ are set to be $1 - \frac{1}{M}$, and any marginal probabilities lower than $\frac{1}{M}$ are set to be $\frac{1}{M}$. We denote such a UMDA with margin as UMDA$_M$. The margins here aim to avoid the premature convergence, which is similar to the upper and lower bounds of the pheromone information in Max-Min Ant System [40] and Laplace correction [2]. It is noteworthy that we are not trying to propose a new algorithm here. Instead, by an example, we are trying to demonstrate theoretically that some approaches proposed to avoid premature convergence of EDAs, can actually help to promote the performance of the algorithms.

We have seen in the previous section that the original UMDA cannot solve BVLEADINGONES efficiently. Interestingly, by adding the margins, the UMDA$_M$ can solve BVLEADINGONES efficiently. The following theorem summarizes the main result.

Theorem 4: Given polynomial population sizes $N = \omega(n^{2+\alpha} \log n)$, $M = \omega(n^{2+\alpha} \log n)$ (where α can be any positive constant) and $M = \beta N$ ($\beta \in (0, 1)$ is some constant), then for any constant δ that satisfies $\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - e^{\frac{1}{\epsilon(n)}} \frac{M}{N})$ (where $\epsilon(n) = \frac{M}{n}$), the first hitting time τ of the UMDA$_M$ with truncation selection (initialized with a uniform distribution) satisfies

$$\tau < \bar{\tau} = \frac{\left(\ln \frac{e(M-1)}{N} - \ln(1-\delta) \right) n \epsilon(n) + n}{\epsilon(n) \ln(1-\delta) + \epsilon(n) \ln\left(\frac{N}{M}\right) - 1} + \frac{M}{N} \ln^2 n + 2n$$

with the overwhelming probability

$$\left(1 - n^{-e^{-1/\epsilon(n)} \omega(n^{2+\alpha}) \delta^2/2e} \right)^{2\bar{\tau}}$$

$$\cdot \left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \omega(1)} \right)^{2(n-1)\bar{\tau}}$$

$$\cdot \left(1 - \left(\frac{1}{e}\right)^{\omega(\ln n)} \right).$$

Proof: In order to proof the above theorem, we define $n+1$ random variables t_0 and t_i ($i = 1, \ldots, n$) as follows:

$$t_0 \triangleq \min\left\{ t; p_{t,n}(\bar{x}_n^*) = 1 - \frac{1}{M} \right\}$$

$$t_i \triangleq \min\left\{ t; p_{t,i}(x_i^*) = 1 - \frac{1}{M} \right\}.$$

The proof follows our basic idea introduced in Section III-A, and thus is similar to the proof of Theorem 2. However, the maximal value that a marginal probability can reach drops to $1 - \frac{1}{M}$, and the minimal value that a marginal probability can reach increases to $\frac{1}{M}$. We will then de-randomize the UMDA$_M$.

In the analysis, we ignore the possibility that the optimum is found before the t_0th generation (which will make the FHT smaller), and we divide the optimization process into $(n+1)$th stages. The 1st stage begins when the optimization begins, and ends when the marginal probability $\hat{p}_{.,n}(\bar{x}_n^*)$ reaches $1 - \frac{1}{M}$

for the first time. The 2nd stage follows the 1st stage, and ends when the marginal probability $\hat{p}_{.,1}(x_1^*)$ reaches $1 - \frac{1}{M}$ for the first time. The qth stage ($q \in \{2, \ldots, n\}$) begins when the marginal probability $\hat{p}_{.,q-2}(x_{q-2}^*)$ reaches $1 - \frac{1}{M}$ for the first time, and ends when the marginal probability $\hat{p}_{.,q-1}(x_{q-1}^*)$ reaches $1 - \frac{1}{M}$ for the first time.

Let us consider the deterministic system. Suppose generation $t+1$ belongs to the ith stage ($i \in \{1, \ldots, n+1\}$), then the marginal probabilities at this generation are updated from the marginal probabilities at generation t by γ_i. When $i = 1$, we have

$$\hat{\mathbf{P}}_{t+1}(\mathbf{x}^*) = \gamma_1(\hat{\mathbf{P}}_t(\mathbf{x}^*)) =$$
$$\left(R \hat{p}_{t,1}(x_1^*), \ldots, R \hat{p}_{t,n-1}(x_{n-1}^*), \right.$$
$$\left. 1 - G_1(1 - \hat{p}_{t,n}(x_n^*)) \right)$$

where $R = (1 - \eta)(1 - \eta')$ ($\eta < 1$ and $\eta' < 1$ are positive functions of the problem size n), and $G_1 = (1 - \delta)\frac{N}{M}$ ($\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - e^{\frac{1}{\epsilon(n)}} \frac{M}{N})$ is a constant). In the above equation, we consider two different cases.

1) $j \in \{1, \ldots, n-1\}$. In the deterministic system above, the jth bits of individuals are not exposed to selection pressure, and we use the factor $R = (1 - \eta)(1 - \eta')$ to demonstrate the impact of genetic drift on these marginal probabilities.

2) $j = n$. In the deterministic system above, the marginal probability $\hat{p}_{t,n}(\bar{x}_n^*) = 1 - \hat{p}_{t,n}(x_n^*)$ is increasing, and we use the factor $G_1 = (1 - \delta)\frac{N}{M}$ to demonstrate the impact of selection pressure on the increasing marginal probability $\hat{p}_{.,n}(\bar{x}_n^*)$ ($\hat{p}_{t+1,n}(\bar{x}_n^*) = G_1 \hat{p}_{t,n}(\bar{x}_n^*)$, thus $\hat{p}_{t+1,n}(x_n^*) = 1 - G_1 \hat{p}_{t,n}(\bar{x}_n^*) = 1 - G_1(1 - \hat{p}_{t,n}(x_n^*))$ holds).

When $i \in \{2, \ldots, n\}$, we have

$$\hat{\mathbf{P}}_{t+1}(\mathbf{x}^*) = \gamma_i(\hat{\mathbf{P}}_t(\mathbf{x}^*))$$
$$= \left(\hat{p}_{t,1}(x_1^*), \ldots, \hat{p}_{t,i-2}(x_{i-2}^*), \right.$$
$$G_2 \hat{p}_{t,i-1}(x_{i-1}^*), R \hat{p}_{t,i}(x_i^*), \ldots,$$
$$\left. R \hat{p}_{t,n-1}(x_{n-1}^*), \hat{p}_{t,n}(x_n^*) \right)$$

where $G_2 = (1-\delta)(1-\frac{1}{M})^n \frac{N}{M}$ ($\delta \in (\max\{0, 1-\frac{2M}{N}\}, 1-e^{\frac{1}{\epsilon(n)}} \frac{M}{N})$ is a constant), and $R = (1 - \eta)(1 - \eta')$ ($\eta < 1$ and $\eta' < 1$ are positive functions of the problem size n). In the above equation, we consider four different cases for the deterministic system above.

1) $j \leq i - 2, j \in \mathbb{N}^+$. The marginal probabilities $\hat{p}_{t,j}(x_j^*)$ have reached $1 - \frac{1}{M}$, and at the next generation they will not change (we will soon prove this).

2) $j = i - 1$. The marginal probability $\hat{p}_{t,j}(x_j^*)$ is increasing, and we use the factor $G_2 = (1 - \delta)(1 - \frac{1}{M})^n \frac{N}{M}$ to demonstrate the impact of selection pressure on this increasing marginal probability.

3) $j \in \{i, \ldots, n-1\}$. The jth bits of individuals are not exposed to selection pressure, and we use the factor $R = (1-\eta)(1-\eta')$ to demonstrate the impact of genetic drift on these marginal probabilities.

4) $j = n$. The marginal probabilities $\hat{p}_{t,n}(\bar{x}_n^*)$ and $\hat{p}_{t,n}(x_n^*)$ have reached $1 - \frac{1}{M}$ and $\frac{1}{M}$ respectively, and at the next

TABLE VIII

CALCULATION OF PROBABILITY THAT t_0 IS UPPER BOUNDED BY $\hat{t}_0$

$$\mathbb{P}\left(t_0 \leq \hat{t}_0 \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right) \tag{32}$$

$$> \quad \mathbb{P}\left(p_{\hat{t}_0-1,1}(1) \geq \frac{M-1}{N(1-\delta)} \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right) \left(1 - e^{-\frac{\hat{p}_{0,1}(1)N}{2}\delta^2} \right) \tag{33}$$

$$> \mathbb{P}\left(p_{\hat{t}_0-1,1}(1) \geq \hat{p}_{\hat{t}_0-1,1}(1) = G^{\hat{t}_0-1}p_{0,n}(\bar{x}_n^*) > \frac{M-1}{N(1-\delta)} \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right) \left(1 - e^{-\frac{\hat{p}_{0,1}(1)N}{2}\delta^2} \right)$$

$$> \mathbb{P}\left(p_{\hat{t}_0-1,1}(1) \geq \hat{p}_{\hat{t}_0-1,n}(\bar{x}_n^*) \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*), \hat{p}_{\hat{t}_0-1,n}(\bar{x}_n^*) > \frac{M-1}{N(1-\delta)} \right)$$

$$\cdot \mathbb{P}\left(\hat{p}_{\hat{t}_0-1,n}(\bar{x}_n^*) > \frac{M-1}{N(1-\delta)} \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right) \left(1 - e^{-\hat{p}_{0,n}(\bar{x}_n^*)N\delta^2/2} \right)$$

TABLE IX

CALCULATION OF (34) AND (35)

$$G_2^{\hat{t}_i-\hat{t}_{i-1}-2}\hat{p}_{\hat{t}_{i-1},i}(x_i^*) = (1-\delta)^{\hat{t}_i-\hat{t}_{i-1}-2}\left(1-\frac{1}{M}\right)^{(\hat{t}_i-\hat{t}_{i-1}-2)n}\left(\frac{N}{M}\right)^{\hat{t}_i-\hat{t}_{i-1}-2}\hat{p}_{\hat{t}_{i-1},i}(x_i^*) < \frac{M-1}{N(1-\delta)(1-\frac{1}{M})^n} \tag{34}$$

$$G_2^{\hat{t}_i-\hat{t}_{i-1}-1}\hat{p}_{\hat{t}_{i-1},i}(x_i^*) = (1-\delta)^{\hat{t}_i-\hat{t}_{i-1}-1}\left(1-\frac{1}{M}\right)^{(\hat{t}_i-\hat{t}_{i-1}-1)n}\left(\frac{N}{M}\right)^{\hat{t}_i-\hat{t}_{i-1}-1}\hat{p}_{\hat{t}_{i-1},i}(x_i^*) \geq \frac{M-1}{N(1-\delta)(1-\frac{1}{M})^n} \tag{35}$$

generation they will not change (we will soon prove this).

Consider the $(n + 1)$th stage, we have

$$\hat{\mathbf{P}}_{t+1}(\mathbf{x}^*) = \gamma_{n+1}(\hat{\mathbf{P}}_t(\mathbf{x}^*))$$
$$= \left(\hat{p}_{t,1}(x_1^*), \ldots, \hat{p}_{t,n-1}(x_{n-1}^*), \hat{p}_{t,n}(x_n^*) \right)$$

where we consider two different cases for this deterministic system.

1) $j \in \{1, \ldots, n - 1\}$. The marginal probabilities $\hat{p}_{t,j}(x_j^*)$ have reached $1 - \frac{1}{M}$, and at the next generation they will not change (we will soon prove this).
2) $j = n$. The marginal probability $\hat{p}_{t,n}(x_n^*)$ is always no smaller than $\frac{1}{M}$.

With $\hat{\mathbf{P}}_0(\mathbf{x}^*) = \left(\frac{1}{2}, \ldots, \frac{1}{2} \right)$, we have

$$\hat{\mathbf{P}}_t(\mathbf{x}^*) = \gamma_i^{t-t_{i-2}}\left(\hat{\mathbf{P}}_{t_{i-2}}(\mathbf{x}^*) \right)$$

where $t_{i-2} < t \leq t_{i-1}$ $(i = 1, \ldots, n + 1)$, and we let $t_{-1} = 0$ represent the beginning of the optimization process. Since $\{\gamma_i\}_{i=1}^{n+1}$ de-randomizes the whole optimization process, $\{t_i\}_{i=0}^n$ in the above equation are no longer random variables. For the sake of clarity, we rewrite the above equation as

$$\hat{\mathbf{P}}_t(\mathbf{x}^*) = \gamma_i^{t-\hat{t}_{i-2}}\left(\hat{\mathbf{P}}_{\hat{t}_{i-2}}(\mathbf{x}^*) \right)$$

where $\hat{t}_{i-2} < t \leq \hat{t}_{i-1}$ $(i = 1, \ldots, n + 1)$. As we will show immediately, $\hat{t}_i$ $(0 \leq i \leq n)$ is an upper bound of the random variable t_i with some probability. Once all $\hat{t}_i$ can be estimated, and all the marginal probabilities $p_{t,j}(x_j^*)$ $(j = 1, \ldots, n)$ have

reached $1 - \frac{1}{M}$, the optimum might already be found, or it will take only a few steps to generate the optimum. Thus, if we can prove that once the marginal probabilities $p_{t,j}(x_j^*)$ $(j = 1, \ldots, n - 1)$ have reached $1 - \frac{1}{M}$, it will never reduce again, our task finally becomes calculating the $\hat{t}_n$, the probability that $\hat{t}_n$ holds as an upper bound of t_n.

We now provide the formal proof stage by stage. At the 1st stage, we analyze the case with the nth bit. At the tth generation (which belongs to the 1st stage), according to Lemma 5 and Chernoff bounds, we have

$$\mathbb{P}\left(p_{t,n}(\bar{x}_n^*) \geq (1-\delta)\frac{p_{t-1,n}(\bar{x}_n^*)N}{M} \right.$$
$$\left. \mid p_{t-1,n}(\bar{x}_n^*) \leq \frac{M-1}{N(1-\delta)} \right)$$
$$> 1 - e^{-p_{t-1,n}(\bar{x}_n^*)N\delta^2/2}$$

where $\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - e^{\frac{1}{\epsilon(n)}}\frac{M}{N})$ is a positive constant, and $p_{t,n}(\bar{x}_n^*) \leq 1 - \frac{1}{M}$ (since the UMDA adopts margins) yields the condition that $p_{t-1,n}(\bar{x}_n^*) \leq \frac{M-1}{N(1-\delta)}$. Similar to Table III in the proof of Theorem 2 we can obtain

$$\mathbb{P}\left(p_{t,n}(\bar{x}_n^*) \geq G_1^t p_{0,n}(\bar{x}_n^*) \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right)$$
$$> \left(1 - e^{-p_{0,n}(\bar{x}_n^*)N\delta^2/2} \right)^t. \tag{36}$$

Consider the probability that t_0 is upper bounded by some value, say $\hat{t}_0$, we obtain the inequalities estimated in Table VIII, where in (33) the factor $\left(1 - e^{-\hat{p}_{0,n}(\bar{x}_n^*)N\delta^2/2} \right)$ is

added since we apply Chernoff bounds at the end of the $(\hat{t}_0 - 1)$th generation. Now we consider the following item:

$$\mathbb{P}\left(\hat{p}_{\hat{t}_0-1,n}(\bar{x}_n^*) > \frac{M-1}{N(1-\delta)} \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*)\right)$$

$$= \mathbb{P}\left(\hat{p}_{\hat{t}_0-1,n}(\bar{x}_n^*) > \frac{M-1}{N(1-\delta)}\right). \tag{37}$$

Since $\{\hat{p}_{t,n}(\bar{x}_n^*)\}_{t=0}^{\infty}$ is a deterministic sequence, the probability above must be either 0 or 1. We need to find the value of $\hat{t}_0$ that makes the above probability 1. Given that $\hat{p}_{0,n}(\bar{x}_n^*) = \frac{1}{2}$, the definition of $\hat{t}_0$ (it is an upper bound of t_0 defined at the beginning of the proof) and the condition that $\forall t < \hat{t}_0 - 1 : \frac{M-1}{N(1-\delta)} > \hat{p}_{t,n}(\bar{x}_n^*) > (1-\delta)^{\frac{\hat{p}_{t-1,n}(\bar{x}_n^*)N}{M}}$ together imply

$$G_1^{\hat{t}_0-2}\hat{p}_{0,n}(\bar{x}_n^*)$$
$$= \left((1-\delta)\left(\frac{N}{M}\right)\right)^{\hat{t}_0-2}\hat{p}_{0,n}(\bar{x}_n^*) < \frac{M-1}{N(1-\delta)}$$

$$G_1^{\hat{t}_0-1}\hat{p}_{0,n}(\bar{x}_n^*)$$
$$= \left((1-\delta)\left(\frac{N}{M}\right)\right)^{\hat{t}_0-1}\hat{p}_{0,n}(\bar{x}_n^*) \geq \frac{M-1}{N(1-\delta)}.$$

Hence, we obtain the value of $\hat{t}_0$

$$\hat{t}_0 \leq \frac{\ln\frac{2M-2}{N} - \ln(1-\delta)}{\ln(1-\delta) + \ln\left(\frac{N}{M}\right)} + 2.$$

Now we can continue to estimate the probability mentioned in (32), which can provide us the probability that t_0 is upper bounded by $\hat{t}_0$. Similar to (25) in the proof of Theorem 2, according to (36), we can obtain that the probability is at least $\left(1 - e^{-p_{0,n}(\bar{x}_n^*)N\delta^2/2}\right)^{\hat{t}_0}$.

On the other hand, we can deal with the genetic drift in the same way as we did for Theorem 2: since $\hat{t}_0 = \Theta(1)$, when $t = \hat{t}_0$, for the marginal probabilities of other bits, a level of $\frac{1}{e}$ can be maintained at least with the overwhelming probability of

$$\left(1 - e^{-\frac{\omega(n^{2+\alpha}\log n)}{2e}\delta^2}\right)^{\hat{t}_0}\left(1 - n^{-\left(1-\left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2\omega(1)}\right)^{2\hat{t}_0}$$

where the second factor $\left(1 - n^{-\left(1-\left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2\omega(1)}\right)^{2\hat{t}_0}$ comes from the analysis of genetic drift (please refer to (26) for details). The proof details will be very similar to those in the proof of Theorem 2. For the sake of brevity, we omit the details. Now we have finished the analysis of the 1st stage.

After the marginal probability $p_{.,n}(\bar{x}_n^*)$ has reached $1 - \frac{1}{M}$, i.e., $t \geq \hat{t}_0$, $p_{.,n}(\bar{x}_n^*)$ will not drop to a level that is smaller than $1 - \frac{1}{M}$ again unless the algorithm has found the optimum. In fact, for other marginal probabilities, similar fact also holds. In order to prove it, let us consider the $(i + 1)$th stage ($1 \leq i < n$), and we use the factor G_2 to demonstrate the impact of selection, by which the interactions among bits are taken into account. For the ith bit, at the kth generation, we can investigate the following situation:

$$p_{k,i}(x_i^*) < 1 - \frac{1}{M},$$
$$\forall j \leq i - 1 : p_{k,j}(x_j^*) = 1 - \frac{1}{M}.$$

We will then prove that once $\forall 1 \leq j \leq i - 1, p_{.,j}(x_j^*)$ reach $1 - \frac{1}{M}$, with an overwhelming probability, none of them will decrease again with an overwhelming probability. Let $r_{k+1}\left((1^{i-1} * * \cdots * 1)\right)$ be the proportion of individuals $(1^{i-1} * * \cdots * 1)$ before selection at the $(k + 1)$th generation, where $*$ must be either 0 or 1. According to Chernoff bounds, and with $N > M = \epsilon(n)n$, we have

$$\mathbb{P}\left(r_{k+1}\left((1^{i-1} * * \cdots * 1)\right) > (1-\delta)\left(1 - \frac{1}{M}\right)^i\right.$$

$$\left. \mid p_{k,n}(\bar{x}_n^*) = 1 - \frac{1}{M}, \forall j \leq i - 1 : p_{k,j}(x_j^*) = 1 - \frac{1}{M}\right)$$

$$> 1 - e^{-(1-\frac{1}{M})^i N\delta^2/2} > 1 - e^{-(1-\frac{1}{M})^n N\delta^2/2}$$
$$> 1 - e^{-(1-\frac{1}{\epsilon(n)n})^n \epsilon(n)n\delta^2/2}$$
$$\to 1 - e^{-e^{-1/\epsilon(n)}\epsilon(n)n\delta^2/2}$$

which is an overwhelming probability when $n \to \infty$. Since $\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - e^{\frac{1}{\epsilon(n)}}\frac{M}{N})$, we know that

$$r_{k+1}\left((1^{i-1} * * \cdots * 1)\right)$$
$$> (1-\delta)\left(1 - \frac{1}{M}\right)^i$$
$$> (1-\delta)\left(1 - \frac{1}{M}\right)^n > \frac{M}{N}$$

holds with an overwhelming probability $1 - e^{-e^{-1/\epsilon(n)}\epsilon(n)n\delta^2/2}$. At the same time, it is obvious that the individuals $(1^{i-1} * * \cdots * 1)$ have the highest fitness in the population. After truncation selection, according to Lemma 5, we obtain that (note that we use margins for the marginal probabilities)

$$\mathbb{P}\left(\forall j \leq i - 1 : p_{k+1,j}(x_j^*) = 1 - \frac{1}{M} \mid p_{k,n}(\bar{x}_n^*) = 1 - \frac{1}{M},\right.$$

$$\left. \forall j \leq i - 1 : p_{k,j}(x_j^*) = 1 - \frac{1}{M}\right)$$

$$> 1 - e^{-e^{-1/\epsilon(n)}\epsilon(n)n\delta^2/2} \tag{38}$$

which means with an overwhelming probability, the marginal probabilities $p_{.,j}(x_j^*)$ ($\forall j \leq i - 1$) will no longer change once they reach $1 - \frac{1}{M}$.

Now we consider the $(i + 1)$th stage ($i \leq n - 1$), at which the ith bits of individuals are of our interest. Similar to the case of the 1st stage, in which the marginal probability $\hat{p}_{.,n}(\bar{x}_n^*)$ is investigated, we can estimate the time that $\hat{p}_{.,i}(x_i^*)$ reaches $1 - \frac{1}{M}$, i.e., $\hat{t}_i$ ($1 \leq i < n$). As presented in Table IX, it is not hard to obtain (34) and (35).

In order to obtain $\hat{t}_i$. we need to know $\hat{p}_{\hat{t}_{i-1},i}(x_i^*)$ so as to solve (34) and (35). It is worth noting that $\hat{p}_{\hat{t}_{i-1},i}(x_i^*)$ is related to the genetic drift. Similar to what we did in Section IV, when the bits are not exposed to selection pressure, given that $\hat{t}_{i-1} = O(n)$, the marginal probability $\hat{p}_{.,i}(x_i^*)$ will remain to

be as $\frac{1}{e}$. [8] Hence, we have $p_{\hat{t}_{i-1},i}(x_i^*) > \frac{1}{e}$ holds with the overwhelming probability of

$$\prod_{k=0}^{i-1}\left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \omega(1)}\right)^{2\hat{t}_k} \tag{39}$$

where the item

$$\left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \omega(1)}\right)^{2\hat{t}_k}$$

represents the probability that the $(k+1)$th marginal probability is at least $\frac{1}{e}$ after genetic drift. Detailed analysis can be found in the proof of Theorem 2.

Now we can solve the equations given in (34) and (35), and get

$$\hat{t}_i = \hat{t}_0 + \sum_{k=1}^{i}(\hat{t}_k - \hat{t}_{k-1})$$

$$< \frac{(i+1)\left(\ln\frac{e(M-1)}{N} - \ln(1-\delta) + \frac{1}{\epsilon(n)}\right)}{\ln(1-\delta) + \ln\left(\frac{N}{M}\right) - \frac{1}{\epsilon(n)}}$$

$$+2(i+1) \tag{40}$$

where $i \leq n - 1$ holds.

Next, we need to estimate the joint probability that the random variable t_i is upper bounded by $\hat{t}_i$. Since similar work has been done in (32) and (33), and (20) in the proof of Theorem 2, we only informally describe it here for the sake of brevity. This joint probability contains four parts.

1) The probability that $\forall k \in \{1, \ldots, i-1\} : t_k < \hat{t}_k$. (It can be obtained by induction. For more details, please refer to (20).)

2) The probability that after genetic drift of the ith bit, the marginal probability $p_{\hat{t}_{i-1},i}(x_i^*)$ is larger than $\frac{1}{e}$. (We have already mentioned it in (39).)

3) The probability that after the marginal probabilities $p_{\cdot,j}(x_j^*)$ ($j \neq n$) have reached $1 - \frac{1}{M}$, they will never drop to a lower level again. (We can utilize the result given in (38).)

4) The probability that $p_{t,i}(x_i^*)$ is lower bounded by $\hat{p}_{t,i}(x_i^*)$ ($\hat{t}_{i-1} < t \leq \hat{t}_i$), given the condition that $p_{\hat{t}_{i-1},i}(x_i^*) \geq \hat{p}_{\hat{t}_{i-1},i}(x_i^*)$.

Now we briefly estimate the probability mentioned in Item 4 (and a more detailed example can be found in Table III in the proof of Theorem 2). As the first step, we consider the relation between $p_{t,i}(x_i^*)$ and $p_{t-1,i}(x_i^*)$ ($\hat{t}_{i-1} < t \leq \hat{t}_i$) by applying Chernoff bounds twice. As a result, we obtain the inequalities presented in Table X, where we utilize "min" to take into account the situation in which $(1-\delta)\frac{N}{M}p_{t-1,i}(x_i^*)p_{t-1,n}(\bar{x}_n^*)\prod_{j=1}^{i-1}p_{t-1,j}(x_j^*) > 1 - \frac{1}{M}$ holds. In this case, noting that the UMDA has adopted margins, the

marginal probability $p_{t,i}(x_i^*)$ is set to be $1 - \frac{1}{M}$. By setting the condition of the above probability as $p_{t-1,i}(x_i^*) \geq \hat{p}_{t-1,i}(x_i^*) = G_2^{t-\hat{t}_{i-1}-1}\hat{p}_{\hat{t}_{i-1},i}(x_i^*)$, the above inequality further implies that

$$\mathbb{P}\left(p_{t,i}(x_i^*) \geq \min\left\{G_2 p_{t-1,i}(x_i^*), 1 - \frac{1}{M}\right\}\right.$$

$$\left.\mid p_{t-1,i}(x_i^*) \geq G_2^{t-\hat{t}_{i-1}-1}\hat{p}_{\hat{t}_{i-1},i}(x_i^*)\right)$$

$$> 1 - e^{-(1-\frac{1}{M})^n G_2^{t-\hat{t}_{i-1}-1}\hat{p}_{\hat{t}_{i-1},i}(x_i^*)N\delta^2/2}$$

$$> 1 - e^{-(1-\frac{1}{M})^n \hat{p}_{\hat{t}_{i-1},i}(x_i^*)N\delta^2/2}$$

$$> 1 - e^{-(1-\frac{1}{M})^n N\delta^2/2e}$$

holds, where we utilize the facts that $\hat{p}_{\hat{t}_{i-1},i}(x_i^*) > \frac{1}{e}$ holds with an overwhelming probability (the consequence of genetic drift. Original analysis can be found before (27), and $G_2 > 1$ (which ensures that $\hat{p}_{t,i}(x_i^*)$ is mono-increasing when the time index t satisfies $\hat{t}_{i-1} < t \leq \hat{t}_i$). As a consequence of the above inequality, similar to Table III in the proof of Theorem 2, we obtain the probability mentioned in Item 4

$$\left(1 - e^{-(1-\frac{1}{M})^n N\delta^2/2e}\right)^{\hat{t}_i - \hat{t}_{i-1}}$$

$$= \left(1 - e^{-e^{-1/\epsilon(n)}\omega(n^{2+\alpha}\log n)\delta^2/2e}\right)^{\hat{t}_i - \hat{t}_{i-1}}.$$

Now combining the probabilities mentioned in Items 1, 2, 3 and 4 together, we can obtain that t_i is upper bounded by $\hat{t}_i$ at least with the probability of

$$\left(1 - n^{-e^{-1/\epsilon(n)}\omega(n^{2+\alpha})\delta^2/2e}\right)^{2\hat{t}_i}$$

$$\cdot \prod_{k=0}^{i-1}\left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \omega(1)}\right)^{2\hat{t}_k}.$$

As a result, t_{n-1} is bounded by $\hat{t}_{n-1}$ with the overwhelming probability of

$$\left(1 - n^{-e^{-1/\epsilon(n)}\omega(n^{2+\alpha})\delta^2/2e}\right)^{2\hat{t}_{n-1}} \cdot \prod_{k=0}^{n-2}\left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \omega(1)}\right)^{2\hat{t}_k}.$$

When all the marginal probabilities $p_{\cdot,i}(x_i^*)$ ($i \neq n$) have reached $1 - \frac{1}{M}$, the marginal probability $p_{\cdot,n}(\bar{x}_n^*)$ will become smaller and smaller and the probability of finding the optimum becomes larger and larger.

Now we consider the $(n + 1)$th stage, in which two events hold: 1) $\hat{p}_{\hat{t}_{n-1},n}(x_n^*) \geq \frac{1}{M}$ holds; 2) $\forall t > \hat{t}_{n-1}$, $t \prec Poly(n)$, $\forall j \leq n-1 : p_{t,j}(x_j^*) = 1 - \frac{1}{M}$ holds with an overwhelming probability (38). Thus, there is no genetic drift to be taken into account. Meanwhile, the probability of generating the optimum in one sampling of a generation, conditional on the above two events, is at least $(1 - \frac{1}{M})^{n-1}\frac{1}{M} = e^{-(n-1)/n\epsilon(n)}\frac{1}{M}$, which implies that if the above two events both happen (which is true in the $(n+1)$th stage), then the optimum will be found within $M \ln^2 n$ extra samplings (which generates $M \ln^2 n$ new individuals) with the overwhelming probability $1 - \left(\frac{1}{e}\right)^{\omega(\ln n)}$. Consequently, after the first n stages, at most $\frac{M}{N}\ln^2 n$ generations can guarantee the emergence of the optimum with an overwhelming probability.

TABLE X

BOUNDING $p_{t,i}(x_i^*)$ FROM BELOW WITH AN OVERWHELMING PROBABILITY

$$\mathbb{P}\left(p_{t,i}(x_i^*) \geq \min\left\{ (1-\delta)\frac{N}{M} p_{t-1,i}(x_i^*) p_{t-1,n}(\bar{x}_n^*) \prod_{j=1}^{i-1} p_{t-1,j}(x_j^*), 1-\frac{1}{M} \right\} \right.$$

$$\left. \mid p_{t-1,i}(x_i^*), p_{t,n}(\bar{x}_n^*) = 1 - \frac{1}{M}, \forall j \leq i-1 : p_{t-1,j}(x_j^*) = 1 - \frac{1}{M} \right)$$

$$> \mathbb{P}\left(p_{t,i}(x_i^*) \geq \min\left\{ (1-\delta)\frac{N}{M}(1-\frac{1}{M})^n p_{t-1,i}(x_i^*), 1-\frac{1}{M} \right\} \mid p_{t-1,i}(x_i^*) \right)$$

$$> 1 - e^{-(1-\frac{1}{M})^n p_{t-1,i}(x_i^*) N\delta^2/2}$$

TABLE XI

CALCULATION OF PROBABILITY THAT T_n' IS UPPER BOUNDED BY $\hat{T}_n'$

$$\mathbb{P}\left(T_n' \leq \hat{T}_n' \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right) \tag{41}$$

$$> \mathbb{P}\left(p_{\hat{T}_n'-1,1}(1) \geq \frac{M}{N(1-\delta)} \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right)\left(1 - e^{-\frac{\hat{p}_{0,1}(1)N}{2}\delta^2} \right) \tag{42}$$

$$> \mathbb{P}\left(p_{\hat{T}_n'-1,1}(1) \geq \hat{p}_{\hat{T}_n'-1,1}(1) = G^{\hat{T}_n'-1} p_{0,n}(\bar{x}_n^*) > \frac{M}{N(1-\delta)} \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right)\left(1 - e^{-\frac{\hat{p}_{0,1}(1)N}{2}\delta^2} \right)$$

$$> \mathbb{P}\left(p_{\hat{T}_n'-1,1}(1) \geq \hat{p}_{\hat{T}_n'-1,n}(\bar{x}_n^*) \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*), \hat{p}_{\hat{T}_n'-1,n}(\bar{x}_n^*) > \frac{M}{N(1-\delta)} \right)$$

$$\mathbb{P}\left(\hat{p}_{\hat{T}_n'-1,n}(\bar{x}_n^*) > \frac{M}{N(1-\delta)} \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right)\left(1 - e^{-p_{0,n}(\bar{x}_n^*)N\delta^2/2} \right)$$

Hence, the first hitting time τ is upper bounded by a deterministic value $\bar{\tau}$.

$$\tau < \bar{\tau} = \frac{\left(\ln \frac{e(M-1)}{N} - \ln(1-\delta) \right)n\epsilon(n) + n}{\epsilon(n)\ln(1-\delta) + \epsilon(n)\ln\left(\frac{N}{M}\right) - 1}$$
$$+ \frac{M}{N}\ln^2 n + 2n$$

with the overwhelming probability at least

$$\left(1 - n^{-e^{-1/\epsilon(n)}\omega(n^{2+\alpha})\delta^2/2e} \right)^{2\bar{\tau}}$$
$$\cdot \left(1 - n^{-\left(1-(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \omega(1)} \right)^{2(n-1)\bar{\tau}}$$
$$\cdot \left(1 - \left(\frac{1}{e}\right)^{\omega(\ln n)} \right).$$

∎

The results in this section show that margins can avoid misleading convergence and leave some chances to the UMDA$_M$ to find the global optimum. However, UMDA$_M$ cannot converge to the global optimum completely anymore, i.e., the CT becomes infinite. This is an interesting case where the FHT is bounded polynomially in the problem size, but the CT is infinite, and it demonstrates that FHT is a more appropriate measure for EDAs time complexity than CT. It is noteworthy that the idea of margins is quite similar to the Laplace correction [2], which was also proposed to prevent the marginal probabilities from premature convergence. However, since our purpose here is to demonstrate the influence of forbidding a marginal probability to be 0 or 1, the slight difference between relaxation and Laplace correction is not investigated.

VII. CONCLUSION

In this paper, we utilized the FHT to measure the time complexity of EDAs. Based on the FHT measure, we proposed a classification of problem hardness for EDAs and the corresponding probability conditions. This is the first time the general issues related to the time complexity of EDAs were discussed theoretically. After that, a new approach to analyzing the FHT for EDAs with finite population was introduced. Using this approach, we investigated the time complexity of UMDAs as examples.

In this paper, UMDAs were analyzed in depth on two problems: LEADINGONES [37] and BVLEADINGONES. Both of the problems are unimodal. The latter was derived from the former, and inherited the domino convergence property of the former. For the original UMDA, LEADINGONES is shown to be EDA-easy, and BVLEADINGONES is shown to be EDA-hard. Comparing the theoretical results of EDAs with those of the EAs', although the first result is similar to EAs', i.e., LEADINGONES is easy, it should be noted that the general case does not hold. That is, a problem that is easy for the EAs could be hard for EDAs, e.g., the BVLEADINGONES problem. However, it is still an open issue to analyze problems that are hard for the EAs but easy for the EDAs.

If the UMDA is further relaxed by margins, BVLEADIN-GONES will no longer be EDA-hard. Our analysis shows that the margin is helpful for UMDA to avoid wrong convergence and thus significantly increases the performance of UMDA on BVLEADINGONES. This is the first rigorous time complexity evidence that supports the efficacy of relaxations (corrections) of EDAs.

Finally, although we only analyze UMDAs, our approach has the potential for analyzing other EDAs with the finite populations. The general idea is to find a way to simplify the EDAs and then estimate the probability that this simplification holds. However, since different EDAs may have different characteristics, more work needs to be done for the generalization of our approach.

APPENDIX

Proof of Lemma 6. According to Chernoff bounds, we have

$$\mathbb{P}\left(p_{t,n}(\bar{x}_n^*) \geq (1-\delta)\frac{p_{t-1,n}(\bar{x}_n^*)N}{M} \mid p_{t-1,n}(\bar{x}_n^*) \leq \frac{M}{N(1-\delta)} \right)$$
$$> 1 - e^{-p_{t-1,n}(\bar{x}_n^*)N\delta^2/2}, \forall t \leq U$$

where $\delta \in (\max\{0, 1-\frac{2M}{N}\}, 1-\frac{M}{N})$ is a positive constant. Since no global optimum is generated before the Uth generation, we have

$$\hat{p}_{t,n}(\bar{x}_n^*) = G^t p_{0,n}(\bar{x}_n^*), \forall t \leq U$$

where $G = (1-\delta)\frac{N}{M}$, and $\hat{p}_{t,n}(\bar{x}_n^*)$ is deterministic given the initial value $p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) = \frac{1}{2}$. Furthermore, setting $t = U$ in the above equation, by calculation we obtain that

$$\hat{p}_{U,n}(\bar{x}_n^*) = 1.$$

Let $\hat{T}_n'$ denote the minimal t for $\hat{p}_{t,n}(\bar{x}_n^*)$ to reach 1, then the above equation implies $\hat{T}_n' \leq U$. We study the probability that the random variable $p_{t,n}(\bar{x}_n^*)$ is larger than $\hat{p}_{t,n}(\bar{x}_n^*)$. Similar to Table III, $\forall t \leq \hat{T}_n'$ we obtain

$$\mathbb{P}\left(p_{t,n}(\bar{x}_n^*) \geq \hat{p}_{t,n}(\bar{x}_n^*) \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right)$$
$$> \left(1 - e^{-p_{0,n}(\bar{x}_n^*)N\delta^2/2} \right)^t.$$

By inequalities in Table XI, we estimate the probability that T_n' is bounded by $\hat{T}_n'$, where in (42) the factor $\left(1 - e^{-p_{0,n}(\bar{x}_n^*)N\delta^2/2} \right)$ is added since we apply Chernoff bounds at the end of the $(\hat{T}_n' - 1)$th generation. We then consider the following item:

$$\mathbb{P}\left(\hat{p}_{\hat{T}_n'-1,n}(\bar{x}_n^*) > \frac{M}{N(1-\delta)} \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right)$$
$$= \mathbb{P}\left(\hat{p}_{\hat{T}_n'-1,n}(\bar{x}_n^*) > \frac{M}{N(1-\delta)} \right).$$

According to the definition of $\hat{T}_n'$, and noting that $\hat{p}_{\hat{T}_n'-1,n}(\bar{x}_n^*) > \frac{M}{N(1-\delta)}$ is deterministic, we know the probability

above is 1. Thus, we continue to estimate the corresponding probability mentioned in (41)

$$\mathbb{P}\left(T_n' \leq \hat{T}_n' \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right)$$
$$> \mathbb{P}\left(p_{\hat{T}_n'-1,n}(\bar{x}_n^*) \geq \hat{p}_{\hat{T}_n'-1,n}(\bar{x}_n^*) \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right)\left(1 - e^{-\frac{\hat{p}_{0,n}(1)N}{2}\delta^2} \right)$$
$$> \left(1 - e^{-\frac{\hat{p}_{0,n}(1)N}{2}\delta^2} \right)^{\hat{T}_n'}.$$

Since $\hat{T}_n' \leq U$, we further get

$$\mathbb{P}\left(T_n' \leq U \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right)$$
$$> \mathbb{P}\left(T_n' \leq \hat{T}_n' \mid p_{0,n}(\bar{x}_n^*) = \hat{p}_{0,n}(\bar{x}_n^*) \right)$$
$$> \left(1 - e^{-\frac{\hat{p}_{0,n}(1)N}{2}\delta^2} \right)^{U}.$$

The analysis above tells us, the probability that the marginal probability converges before the Uth generation ($T_n < U$) is at least $\left(1 - e^{-\frac{N}{4}\delta^2} \right)^{U}$. Since $N = \omega(n^{2+\alpha}\log n)$, $M = \beta N$ ($\beta \in (0,1)$ is a constant) and U is polynomial in the problem size n, this probability is overwhelming. Hence, we have proven the lemma.

REFERENCES

[1] S. Baluja, "Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning," Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-CS-94-163, 1994.
[2] B. Cestnik, "Estimating probabilities: A crucial task in machine learning," in *Proc. Eur. Conf. Artif. Intell.*, 1990, pp. 147–149.
[3] T. Chen, K. Tang, G. Chen, and X. Yao, "On the analysis of average time complexity of estimation of distribution algorithms," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, 2007, pp. 453–460.
[4] T. Chen, J. He, G. Sun, G. Chen, and X. Yao, "A new approach to analyzing average time complexity of population-based evolutionary algorithms on unimodal problems," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 5, pp. 1092–1106, Oct. 2009.
[5] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*. New York: McGraw-Hill, 2001.
[6] J. F. Crow and M. Kimura, *An Introduction of Population Genetics Theory*. New York: Harper and Row, 1970.
[7] S. Droste, T. Jansen, and I. Wegener, "On the analysis of the (1+1) evolutionary algorithm," *Theor. Comput. Sci.*, vol. 276, nos. 1–2, pp. 51–81, Apr. 2002.
[8] S. Droste, "A Rigorous analysis of the compact genetic algorithm for linear functions," *Natural Comput.*, vol. 5, no. 3, pp. 257–283, 2006.
[9] S. Droste, T. Jansen, and I. Wegener, "Upper and lower bounds for randomized search heuristics in black-box optimization," *Theor. Comput. Syst.*, vol. 39, no. 4, pp. 525–544, 2006.
[10] C. González, A. Ramírez, J. A. Lozano, and P. Larrañaga, "Average time complexity of estimation of distribution algorithms," in *Proc. 8th Int. Work Conf. Artif. Neural Netw. (IWANN)*, LNCS 3512. 2005, pp. 42–49.
[11] C. González, J. A. Lozano, and P. Larrañaga, "Analyzing the PBIL algorithm by means of discrete dynamical systems," *Complex Syst.*, vol. 12, no. 4, pp. 465–479, 2000.

[12] C. González, J. A. Lozano, and P. Larrañaga, "Mathematical modelling of discrete estimation of distribution algorithms," in *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*, P. Larrañaga and J. A. Lozano, Eds. Norwell, MA: Kluwer, 2002, pp. 147–163.

[13] C. González, "Contributions on theoretical aspects of estimation of distribution algorithms," Doctoral dissertation, Dept. Comput. Sci. Artif. Intell., Univ. Basque Country, Donostia, San Sebastián, Spain, 2005.

[14] G. R. Harik, F. G. Lobo, and D. E. Goldberg, "The compact genetic algorithm," in *Proc. IEEE Int. Conf. Evol. Comput.*, 1998, pp. 523–528.

[15] J. He and L. Kang, "On the convergence rate of genetic algorithms," *Theor. Comput. Sci.*, vol. 229, nos. 1–2, pp. 23–39, Nov. 1999.

[16] J. He and X. Yao, "Drift analysis and average time complexity of evolutionary algorithms," *Artif. Intell.*, vol. 127, no. 1, pp. 57–85, Mar. 2001.

[17] J. He and X. Yao, "Toward an analytic framework for analysing the computation time of evolutionary algorithms," *Artif. Intell.*, vol. 145, nos. 1–2, pp. 59–97, Apr. 2003.

[18] J. He and X. Yao, "A study of drift analysis for estimating computation time of evolutionary algorithms," *Natural Comput.*, vol. 3, no. 1, pp. 21–35, 2004.

[19] J. He, C. Reeves, and X. Yao, "A discussion on posterior and prior measures of problem difficulties," in *Proc. Parallel Problem Solving Nature 9th Workshop Evol. Algor. Bridging Theory Practice*, 2006.

[20] J. He, C. Reeves, C. Witt, and X. Yao, "A note on problem difficulty measures in black-box optimization: Classification, realizations and predictability," *Evol. Comput.*, vol. 15, no. 4, pp. 435–444, 2007.

[21] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Am. Statist. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.

[22] J. Horn, D. E. Goldberg, and K. Deb, "Long path problems," in *Proc. 3rd Parallel Problem Solving Nature*, LNCS 886. 1994, pp. 149–158.

[23] T. Jansen and I. Wegener, "Evolutionary algorithms: How to cope with plateaus of constant fitness and when to reject strings of the same fitness," *IEEE Trans. Evol. Comput.*, vol. 5, no. 6, pp. 589–599, Dec. 2001.

[24] T. Jansen, K. A. D. Jong, and I. Wegener, "On the choice of the offspring population size in evolutionary algorithms," *Evol. Comput.*, vol. 13, no. 4, pp. 413–440, 2005.

[25] P. Larrañaga and J. A. Lozano, *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*. Norwell, MA: Kluwer, 2001.

[26] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge, MA: Cambridge University Press, 1995.

[27] H. Mühlenbein, "The equation for response to selection and its use for prediction," *Evol. Comput.*, vol. 5, no. 3, pp. 303–346, 1997.

[28] H. Mühlenbein and G. Paaß, "From recombination of genes to the estimation of distribution, I: Binary parameters," in *Proc. 4th Parallel Problem Solving Nature*, LNCS 1411. 1996, pp. 178–187.

[29] H. Mühlenbein and T. Mahnig, "Evolutionary optimization and the estimation of search distributions with applications to graph bipartitioning," *Int. J. Approx. Reasoning*, vol. 31, no. 3, pp. 157–192, Nov. 2002.

[30] H. Mühlenbein, T. Mahnig, and A. Ochoa, "Schemata, distributions and graphical models in evolutionary optimization," *J. Heuristics*, vol. 5, pp. 215–247, Jul. 1999.

[31] H. Mühlenbein and D. Schlierkamp-Voosen, "Predictive models for the Breeder Genetic Algorithm, I: Continuous parameter optimization," *Evol. Comput.*, vol. 1, no. 1, pp. 25–49, 1993.

[32] M. Pelikan, K. Sastry, and D. E. Goldberg, "Evolutionary algorithms + graphical models = scalable black-box optimization," Illinois Genetic Algorithms Lab., Univ. Illinois, Urbana-Champaign, IlliGAL Rep. 2001029, 2001.

[33] M. Pelikan, K. Sastry, and D. E. Goldberg, "Scalability of the Bayesian optimization algorithm," *Int. J. Approx. Reasoning*, vol. 31, no. 3, pp. 221–258, Nov. 2002.

[34] J. A. Rice, *Mathematical Statistics and Data Analysis*. Belmont, CA: Duxbury Press, 1994.

[35] R. Rastegar and M. R. Meybodi, "A study on global convergence time complexity of estimation of distribution algorithms," in *Proc. Rough Sets Fuzzy Sets Data Mining Granular Comput. (RSFDGrC)*, LNAI 3641. 2005, pp. 441–450.

[36] M. Rudnick, "Genetic algorithms and fitness variance with an application to the automated design of artificial neural networks," Doctoral dissertation, Oregon Graduate Instit. Sci. Technol., Beaverton, 1992.

[37] G. Rudolph, "Finite Markov chain results in evolutionary computation: A tour d'horizon," *Fundamenta Informaticae*, vol. 35, nos. 1–4, pp. 67–89, Aug. 1998.

[38] R. J. Serfling, "Probability inequalities for the sum in sampling without replacement," *Ann. Statist.*, vol. 2, no. 1, pp. 39–48, 1974.

[39] E. R. Sheinerman, *Invitation to Dynamical Systems*. Upper Saddle River, NJ: Prentice-Hall, 1996.

[40] T. Stützle and H. H. Hoos, "MAX-MIN ant system," *Future Generation Comput. Syst.*, vol. 16, no. 9, pp. 889–914, 2000.

[41] D. Thierens, D. E. Goldberg, and A. G. Pereira, "Domino convergence, drift, and the temporal-salience structure of problems," in *Proc. IEEE Int. Conf. Evol. Comput.*, 1998, pp. 535–540.

[42] I. Wegener, "Simulated annealing beats metropolis in combinatorial optimization," in *Proc. 32nd Int. Colloq. Automata Languages Programming (ICALP)*, 2005, pp. 589–601.

[43] Y. Yu and Z.-H. Zhou, "A new approach to estimating the expected first hitting time of evolutionary algorithms," *Artif. Intell.*, vol. 172, no. 15, pp. 1809–1832, Oct. 2008.

[44] Q. Zhang, "On stability of fixed points of limit models of univariate marginal distribution algorithm and factorized distribution algorithm," *IEEE Trans. Evol. Comput.*, vol. 8, no. 1, pp. 80–93, Feb. 2004.

[45] Q. Zhang and H. Mühlenbein, "On the convergence of a class of estimation of distribution algorithms," *IEEE Trans. Evol. Comput.*, vol. 8, no. 2, pp. 127–136, Apr. 2004.

Tianshi Chen (S'07) received the B.S. degree in mathematics from the Special Class for the Gifted Young, University of Science and Technology of China (USTC), Hefei, Anhui, China, in 2005. He is currently working toward the Ph.D. degree in computer science from the Nature Inspired Computation and Applications Laboratory, School of Computer Science and Technology, USTC.

His research interests include theoretical aspects of evolutionary algorithms, various real-world applications of evolutionary algorithms, and theoretical aspects of parallel computing.

Ke Tang (S'05–M'07) received the B.E. degree from the Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2002, and the Ph.D. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, in 2007.

Since 2007, he has been an Associate Professor with the Nature Inspired Computation and Applications Laboratory, School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, China. He is the coauthor of more than 30 refereed publications. His major research interests include machine learning, pattern analysis, evolutionary computation, data mining, metaheuristic algorithms, and real-world applications.

Dr. Tang is an Editorial Board Member of three international journals and the Chair of the IEEE Task Force on Large Scale Global Optimization.

Guoliang Chen received the B.S. degree from Xian Jiaotong University, Xian, China, in 1961.

Since 1973, he has been with the University of Science and Technology of China, Hefei, Anhui, China, where he is currently the Academic Committee Chair of the Nature Inspired Computation and Applications Laboratory, a Professor with the School of Computer Science and Technology, and the Director of the School of Software Engineering. From 1981 to 1983, he was a Visiting Scholar with Purdue University, West Lafayette, IN. He is currently also the Director of the National High Performance Computing Center, Hefei, Anhui, China. He has published nine books and more than 200 research papers. His research interests include parallel algorithms, computer architecture, computer networks, and computational intelligence.

Prof. Chen is an Academician of the Chinese Academy of Sciences. He was the recipient of the National Excellent Teaching Award of China in 2003.

Xin Yao (M'91–SM'96–F'03) received the B.S. degree from the University of Science and Technology of China (USTC), Hefei, Anhui, China, in 1982, the M.S. degree from the North China Institute of Computing Technology, Beijing, China, in 1985, and the Ph.D. degree from USTC, in 1990, all in computer science.

From 1985 to 1990, he was an Associate Lecturer and Lecturer with USTC, while working toward the Ph.D. degree in simulated annealing and evolutionary algorithms. In 1990, he was a Postdoctoral Fellow with the Computer Sciences Laboratory, Australian National University, Canberra, Australia, where he continued his work on simulated annealing and evolutionary algorithms. In 1991, he was with the Knowledge-Based Systems Group, Commonwealth Scientific and Industrial Research Organization Division of Building, Construction and Engineering, Melbourne, Australia, where he worked primarily on an industrial project on automatic inspection of sewage pipes. In 1992, he returned to Canberra to take up a Lectureship with the School of Computer Science, University College, University of New South Wales, Australian Defense Force Academy, Sydney, Australia, where he was later promoted to a Senior Lecturer and Associate Professor. Attracted by the English weather, he moved to the University of Birmingham, Edgbaston, Birmingham, U.K., where he became a Professor (Chair) of computer science on April 1, 1999. He is currently the Director of the Center of Excellence for Research in Computational Intelligence and Applications, School of Computer Science, University of Birmingham. He is currently also a Changjiang (Visiting) Chair Professor (Cheung Kong Scholar) with the Nature Inspired Computation and Applications Laboratory, School of Computer Science and Technology, USTC. He has given more than 50 invited keynote and plenary speeches at conferences and workshops worldwide. He has more than 300 refereed publications. His major research interests include evolutionary artificial neural networks, automatic modularization of machine-learning systems, evolutionary optimization, constraint-handling techniques, computational time complexity of evolutionary algorithms, coevolution, iterated prisoner's dilemma, data mining, and real-world applications.

Dr. Yao was the Editor-in-Chief of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION from 2003 to 2008, an Associate Editor or Editorial Board Member of 12 other journals, and the Editor of the World Scientific Book Series on Advances in Natural Computation. He was the recipient of the President's Award for Outstanding Thesis by the Chinese Academy of Sciences for his Ph.D. work on simulated annealing and evolutionary algorithms in 1989. He was the recipient of the 2001 IEEE Donald G. Fink Prize Paper Award for his work on evolutionary artificial neural networks.