A simplified run time analysis of the univariate marginal distribution algorithm on LeadingOnes

Benjamin Doerr [a], Martin S. Krejca [b],*

[a] *Laboratoire d'Informatique (LIX), CNRS, École Polytechnique, Institut Polytechnique de Paris, Palaiseau, France*
[b] *Hasso Platter Institute, University of Potsdam, Potsdam, Germany*

ARTICLE INFO

ABSTRACT

With elementary means, we prove a stronger run time guarantee for the univariate marginal distribution algorithm (UMDA) optimizing the LeadingOnes benchmark function in the desirable regime with low genetic drift. If the population size is at least quasilinear, then, with high probability, the UMDA samples the optimum in a number of iterations that is linear in the problem size divided by the logarithm of the UMDA's selection rate. This improves over the previous guarantee, obtained by Dang and Lehre (2015) via the deep level-based population method, both in terms of the run time and by demonstrating further run time gains from small selection rates. Under similar assumptions, we prove a lower bound that matches our upper bound up to constant factors.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Estimation-of-distribution algorithms (EDAs) are randomized search heuristics that maintain a probabilistic model of the search space and refine it iteratively. In each iteration, the current model of an EDA is used to create some samples which, in turn, are used to adjust the model such that better solutions are more likely to be created in the following iteration. Thus, the model evolves over time into one that creates very good solutions. EDAs have been applied to real-world problems with great success [1].

Within the last few years, the theoretical analysis of EDAs has gained increasing interest (see, for example, the survey by Krejca and Witt [2]). One of the first papers in this period was by Dang and Lehre [3], who proved run time guarantees for the *univariate marginal distribution algorithm* (UMDA, [4]) when optimizing the two classical benchmark functions OneMax and LeadingOnes. While their run time bound for OneMax has been improved since then independently by Lehre and Nguyen [5] and Witt [6,7], the run time bound of $O(n^2 + n\lambda \log \lambda)$ is the best known result so far on LeadingOnes. Here, n is the problem dimension and λ is the offspring population size of the UMDA, which is required to be $\Omega(\log n)$ for the result to hold. In an extension of their result, Dang et al. [8] show the same run time bound but slightly improve a restriction on the population size by a factor of 13.

In this work, we improve in Theorem 5 the second term of this bound from $O(n\lambda \log \lambda)$ to $O\left(n \frac{\lambda}{\log(\lambda/\mu)}\right)$ when $\lambda \geq C\mu = \Omega(n \log n)$, where μ is the size of the subpopulation selected for the model update, and where C is a sufficiently large constant. In the regime of $\mu = \Omega(n \log n)$, the UMDA shows the generally desirable behavior of low genetic drift, that is, the

sampling frequencies stay in the middle range of, say, $(\frac{1}{4}, \frac{3}{4})$ until a sufficiently strong fitness signal moves them into the right direction.

Using the same proof method as for the upper bound, we prove a lower bound (see Theorem 6) that matches the upper bound for polynomial λ. This improves the previously best known lower bounds by Lehre and Nguyen [9] for the regime of $\mu = \Omega(n \log n)$. For the regime of $\mu = \Omega(\log n) \cap o(n \log n)$, the bound $\Omega\left(\frac{n\lambda}{\log(\lambda - \mu)}\right)$ by Lehre and Nguyen remains the best known lower bound. Additionally, Lehre and Nguyen prove a lower bound of $e^{\Omega(\mu)}$ for $\mu = \Omega(\log n)$ and $\lambda \leq C \cdot e\mu$, for an appropriately chosen constant C, which remains untouched by our result.

From our lower bound and the upper bound by Dang and Lehre [3], we observe that the UMDA generally performs better on LeadingOnes in the regime of strong genetic drift, as the regime $\lambda = \Omega(\log n) \cap O(n/\log n)$ results in a run time bound of $O(n^2)$, whereas the regime $\lambda = \Omega(n \log n)$ results in a bound of $\Omega\left(n \frac{\lambda}{\log(\lambda/\mu)}\right)$. Interestingly, for the regime of $\lambda = \omega(n/\log n) \cap o(n \log n)$, it remains open whether a run time of $O(n^2)$ is possible. We believe that this is a particularity of the LeadingOnes problem and that, in general, genetic drift in EDAs leads to a performance loss, since it may take long to move a frequency from the wrong boundary value back into the middle range. This has been rigorously shown by Lengler et al. [10] for the OneMax problem, and by Lehre and Nguyen [11] and Doerr and Krejca [12] for the DeceptiveLeadingBlocks problem.

Equally interesting as the improved run time guarantee is our elementary proof method. While it was truly surprising that Dang and Lehre [3] could use the level-based population method to analyze an EDA (which does not have a population that is transferred from one iteration to the next), this method is a highly advanced tool restricted to algorithms whose population follows a distribution entirely characterized by the previous population. In contrast to this, our proofs only use elementary arguments common in the analysis of evolutionary algorithms and are applicable for deriving upper and lower bounds. We are thus optimistic that our arguments can more easily be applied to other EDAs as well.

We note that both of our bounds do not require the fraction μ/λ to be constant, which is a common requirement of many other analyses of the UMDA [5,7,8,13] (although this is not always explicitly stated in the result). In particular, our bounds show that the gain from reducing the selection rate μ/λ (which often requires a costly increase of λ) is very small, namely, only logarithmic in $\frac{1}{\mu/\lambda}$.

Another advantage of our approach is that it gives run time guarantees that hold with high probability, whereas the level-based method, relying on drift arguments, can only give bounds on expected run times. Consequently, the result of Dang and Lehre [3] also concerns the expectation only. We believe that a result that holds with high probability is often more relevant, as has also been argued by Doerr [14]. A high-probability bound provides can be interpreted as a deterministic upper with a (very) low failure probability, which is more closely connected to classical run time analysis. In contrast, an upper bound on the expected run time only guarantees that the algorithm terminates almost surely, and it provides (without further deeper analysis) only rather weak tail bounds via Markov's inequality.

2. Preliminaries

We are concerned with the run time analysis of algorithms optimizing pseudo-Boolean functions, that is, functions $f : \{0, 1\}^n \to \mathbb{R}$, where $n \in \mathbb{N}$ denotes the dimension of the problem. Given a pseudo-Boolean function f and a bit string x, we refer to f as a *fitness function,* to x as an *individual,* and to $f(x)$ as the *fitness of x.*

For an $n \in \mathbb{N}$, we define $[n] = [1, n] \cap \mathbb{N}$. From now on, if not stated otherwise, the variable n always denotes a natural number. For a vector x of length n, we denote its component at position $i \in [n]$ via x_i.

We consider the optimization of the pseudo-Boolean function LeadingOnes: $\{0, 1\}^n \to \{0\} \cup [n]$, which states for a bit string of length n the longest prefix of leading 1s within that bit string. More formally, for all $x \in \{0, 1\}^n$,

$$\text{LeadingOnes}(x) = \sum_{i=1}^{n} \prod_{j=1}^{i} x_i \, .$$

Note that the all-1s bit string is the unique global optimum of LeadingOnes.

Our algorithm of interest is the UMDA (Algorithm 1) with parameters $\mu, \lambda \in \mathbb{N}$, $\mu \leq \lambda$. It maintains a vector p of probabilities (the *frequency vector*) of length n, whose components we call *frequencies,* and it updates this vector iteratively in the following way: first, λ individuals are created independently from another such that, for each individual $x \in \{0, 1\}^n$ and each position $i \in [n]$, it holds that x_i is 1 with probability p_i and 0 otherwise. Then, from these λ individuals, a subset of μ individuals with the highest fitness is chosen (breaking ties uniformly at random), and, for each position $i \in [n]$, the frequency p_i is set to the relative number of 1s at position i among the μ best individuals. Last, if a frequency p_i is below $\frac{1}{n}$, it is increased to $\frac{1}{n}$, and if it is above $1 - \frac{1}{n}$, it is decreased to $1 - \frac{1}{n}$. This circumvents frequencies from being stuck at the extremal values 0 or 1. We denote the frequency vector of iteration $t \in \mathbb{N}$ by $p^{(t)}$. Note that we start with iteration $t = 0$.

In the context of optimizing LeadingOnes, we say that a position $i \in [n]$ of $p^{(t)}$ is *critical* in iteration $t \in \mathbb{N}$ if and only if all of the frequencies at indices less than i are $1 - \frac{1}{n}$ and $p_i^{(t)}$ is less than $1 - \frac{1}{n}$. Intuitively, a critical frequency is the next one that needs to be set to $1 - \frac{1}{n}$ in order to optimize LeadingOnes efficiently.

When analyzing the run time of the UMDA optimizing a fitness function f, we are interested in the number T of fitness function evaluations until an optimum of f is sampled for the first time. Since the UMDA is a randomized algorithm, T is

ARTICLE IN PRESS

Algorithm 1: The UMDA [4] with parameters μ and λ, $\mu \leq \lambda$, maximizing a fitness function $f \colon \{0, 1\}^n \to \mathbb{R}$ with $n \geq 2$.

1 $t \leftarrow 0$;
2 $p^{(t)} \leftarrow (\frac{1}{2})_{i \in [n]}$;
3 **repeat** ▷ *iteration t*
4   **for** $i \in [\lambda]$ **do** $x^{(i)} \leftarrow$ individual sampled via $p^{(t)}$;
5   let $y^{(1)}, \ldots, y^{(\mu)}$ denote the μ best individuals out of $x^{(1)}, \ldots, x^{(\lambda)}$ (breaking ties uniformly at random);
6   **for** $i \in [n]$ **do** $p_i^{(t+1)} \leftarrow \frac{1}{\mu} \sum_{j=1}^{\mu} y_j^{(i)}$;
7   restrict $p^{(t+1)}$ to the interval $[\frac{1}{n}, 1 - \frac{1}{n}]$;
8 **until** termination criterion met;

---

a random variable, and we are interested in a bound on T that holds with high probability. Note that the run time T of the UMDA is at most λ times the number I of iterations until an optimum is sampled for the first time. Likewise, T is at least $(I - 1)\lambda + 1$.

In order to prove statements on random variables that hold with high probability, we use the following commonly known Chernoff bounds.

Theorem 1 (*Chernoff bound [15, Theorem 1.10.5]*). *Let $k \in \mathbb{N}$, $\delta \in [0, 1]$, and let X be the sum of k independent random variables, each taking values in $[0, 1]$. Then*

$$\Pr\left[X \leq (1 - \delta)\mathrm{E}[X]\right] \leq e^{-\frac{\delta^2 \mathrm{E}[X]}{2}} .$$

Theorem 2 (*Chernoff bound [15, Theorem 1.10.1]*). *Let $k \in \mathbb{N}$, $\delta \in [0, 1]$, and let X be the sum of k independent random variables, each taking values in $[0, 1]$. Then*

$$\Pr\left[X \geq (1 + \delta)\mathrm{E}[X]\right] \leq e^{-\frac{\delta^2 \mathrm{E}[X]}{3}} .$$

The next two theorems, recently proven by Doerr and Zheng [16], give upper bounds on the negative effect of genetic drift on the UMDA. The first result considers the optimization of fitness functions f that *weakly prefer* a 1 at a position $i \in [n]$, that is, for all bit strings $x, x' \in \{0, 1\}^n$ with $x_i = 1$, $x'_i = 0$, and $x_j = x'_j$ for all other positions $j \in [n] \setminus \{i\}$, it holds that $f(x) \geq f(x')$. In other words, having a 1 at position i always yields a fitness at least as good as when having a 0 at i. Note that LeadingOnes weakly prefers a 1 in all bit positions. The theorem states that the frequency at such a position i does not drop far below its initial value $\frac{1}{2}$ for a long time.

Theorem 3 (*[16, Theorem 7]*). *Consider the UMDA with parameters μ and λ optimizing a function f that weakly prefers a 1 at position $i \in [n]$. Then, for all $d > 0$ and all iterations $t \in \mathbb{N}$, we have*

$$\Pr\left[\forall t' \in [0..t]: p_i^{(t')} > \tfrac{1}{2} - d\right] \geq 1 - 2e^{-\frac{d^2 \mu}{2t}} .$$

The next theorem considers the case that there is no preference for a bit value at position $i \in [n]$, that is, for all bit strings $x, x' \in \{0, 1\}^n$ with $x_i = 1$, $x'_i = 0$, and $x_j = x'_j$ for all other positions $j \in [n] \setminus \{i\}$, it holds that $f(x) = f(x')$. Given this assumption, we call position i *neutral*.

Theorem 4 (*[16, Corollary 2]*). *Consider the UMDA with parameters μ and λ optimizing a function f such that position $i \in [n]$ is neutral. Then, for all $d > 0$ and all iterations $t \in \mathbb{N}$, we have*

$$\Pr\left[\forall t' \in [0..t]: p_i^{(t')} \in (\tfrac{1}{2} - d, \tfrac{1}{2} + d)\right] \geq 1 - 2e^{-\frac{d^2 \mu}{2t}} .$$

3. Upper bound

In the following, we present our simple and intuitive run time analysis for the upper bound of the UMDA optimizing LeadingOnes, which gives the following theorem.

Theorem 5. *Let $\delta \in (0, 1)$ be a constant, and let $\zeta = \frac{1-\delta}{4e}$. Consider the UMDA optimizing LeadingOnes with $\mu \geq 128n \ln n$ and $\lambda \geq \frac{\mu}{\zeta}$. Further, let $d = \lfloor \log_4(\zeta \frac{\lambda}{\mu}) \rfloor$. Then the UMDA samples the optimum after at most $\lambda(\lceil \frac{n}{d+1} \rceil + \lceil \frac{n}{n-1} e \ln n \rceil)$ fitness function evaluations with a probability of at least $1 - 5n^{-1}$.*

As discussed in the introduction, we only want to consider the regime with low genetic drift. Hence, we first argue that no frequency drops below $\frac{1}{4}$ before the optimum is sampled (Lemma 1). Then we show that, in this case, in each iteration, roughly $\log \frac{\lambda}{\mu}$ additional frequencies are set to $1 - \frac{1}{n}$. More specifically, if $i \in [n]$ is critical, then all frequencies at positions roughly up to $i + \log \frac{\lambda}{\mu}$ are set to $1 - \frac{1}{n}$ (Lemma 2). Thus, a total of roughly $\frac{n}{\log(\lambda/\mu)}$ iterations suffice to move all frequencies to $1 - \frac{1}{n}$. From such a state, the optimum is sampled with high probability after a logarithmic number of iterations.

We start by proving that the following parameter setting ensures that no frequency drops below the value $\frac{1}{4}$ within $2n$ iterations with high probability.

Lemma 1. *Consider the UMDA with $\lambda \geq \mu \geq 128n \ln n$. Assume that it optimizes a function that weakly prefers a 1 at all positions. Then, with a probability of at least $1 - 2n^{-1}$, each frequency will stay at a value of at least $\frac{1}{4}$ for the first $2n$ iterations.*

Proof. Consider an iteration $t \leq 2n$ as well as a position $i \in [n]$. By Theorem 3 with $d = \frac{1}{4}$, we see that the probability that p_i drops below $\frac{1}{4}$ within the first $t \leq 2n$ iterations is at most $2e^{-\mu/(32 \cdot t)} \leq 2e^{-\mu/(32 \cdot 2n)} \leq 2n^{-2}$, where we used our bound on μ. Applying a union bound over all n frequencies gives the claim. □

We now prove that a critical frequency, all its preceding frequencies, as well as roughly $\log \frac{\lambda}{\mu}$ following frequencies are set to $1 - \frac{1}{n}$ within a single iteration. That is, we increase roughly $1 + \log \frac{\lambda}{\mu}$ new frequencies to their maximum value.

Lemma 2. *Let $\delta \in (0, 1)$ be a constant, and let $\zeta = \frac{1-\delta}{4e}$. Consider the UMDA optimizing LEADINGONES with $\mu \geq 4 \frac{1-\delta}{\delta^2} \ln n$ and $\lambda \geq \frac{\mu}{\zeta}$. Furthermore, consider an iteration $t \in \mathbb{N}$ such that position $i \in [n]$ is critical and that, for all positions $j \geq i$, we have $p_j^{(t)} \geq \frac{1}{4}$. Let $d = \lfloor \log_4(\zeta \frac{\lambda}{\mu}) \rfloor$. Then, with a probability of at least $1 - n^{-2}$, for all positions $j \in [\min\{n, i + d\}]$, we have $p_j^{(t+1)} = 1 - \frac{1}{n}$.*

Proof. Note that $d \geq 0$ due to our assumption on λ. We look at the population of λ individuals that is sampled in iteration t and determine the number X of individuals that have at least $i' := \min\{n, i + d\}$ leading 1s. Since the frequencies at all positions less than i are at $1 - \frac{1}{n}$, the probability that all of these frequencies sample a 1 for a single individual is $(1 - \frac{1}{n})^{i-1} \geq (1 - \frac{1}{n})^{n-1} \geq \frac{1}{e}$. Further, since the probability to sample a 1 at positions at least i is at least $\frac{1}{4}$, we have $E[X] \geq \frac{\lambda}{e} \cdot 4^{-(1+d)} \geq \frac{\mu}{4e\zeta} \geq \frac{\mu}{1-\delta}$.

We now apply Theorem 1 in order to show that it is unlikely that fewer than μ individuals from iteration t have fewer than i' leading 1s. Using our bounds on μ and our estimate on $E[X]$ from above, we compute

$$\Pr[X < \mu] \leq \Pr[X \leq (1 - \delta)E[X]] \leq e^{-\frac{\delta^2 E[X]}{2}}$$
$$\leq e^{-\frac{\delta^2}{2(1-\delta)}\mu} \leq n^{-2}.$$

Thus, with a probability of at least $1 - n^{-2}$, at least μ individuals have at least i' leading 1s.

Since the UMDA is optimizing LEADINGONES, in this case, all of the selected top μ individuals have at least i' leading 1s, which results in all frequencies at positions in $[i']$ being set to $1 - \frac{1}{n}$, that is, for all $j \in [i']$, we have $p_j^{(t+1)} = 1 - \frac{1}{n}$. □

We now prove our main result.

Proof of Theorem 5. We prove that the UMDA samples the optimum after $\lceil \frac{n}{d+1} \rceil + \lceil \frac{n}{n-1} e \ln n \rceil$ iterations with a probability of at least $1 - 5n^{-1}$. Since it performs λ fitness function evaluations each iteration, the theorem follows.

Since LEADINGONES weakly prefers 1 at all positions, by Lemma 1 and $\mu \geq 128n\lceil \ln n \rceil$, no frequency drops below $\frac{1}{4}$ within $2n$ iterations with a probability of at least $1 - 2n^{-1}$.

Consider an iteration $t \leq n$ such that position $i \in [n]$ is critical. Note that $\mu \geq 4\lceil \frac{1-\delta}{\delta^2} \ln n \rceil$ for sufficiently large n. By Lemma 2, with a probability of at least $1 - n^{-2}$, for each frequency at position in $j \in [\min\{n, i+d\}]$, we have $p_j^{(t+1)} = 1 - \frac{1}{n}$. That is, $d + 1$ additional frequencies are set to $1 - \frac{1}{n}$. Applying a union bound for the first $2n$ iterations of the UMDA shows that all frequencies are at $1 - \frac{1}{n}$ after the first $\lceil \frac{n}{d+1} \rceil$ iterations and stay there for at least n additional iterations with a probability of at least $1 - 2n^{-1}$.

Consequently, after the first n iterations, the optimum is sampled in each iteration with a probability of $(1 - \frac{1}{n})^n \geq (1 - \frac{1}{n})\frac{1}{e}$. Thus, after $\lceil \frac{n}{n-1} e \ln n \rceil$ additional iterations, the optimum is sampled with a probability of at least $1 - (1 - \frac{n-1}{n} \frac{1}{e})^{\lceil \frac{n}{n-1} e \ln n \rceil} \geq 1 - n^{-1}$.

Overall, by applying a union bound over all failure probabilities, the UMDA needs at most $\lceil \frac{n}{d+1} \rceil + \lceil \frac{n}{n-1} e \ln n \rceil$ iterations to sample the optimum for the first time with a probability of at least $1 - 5n^{-1}$. □

We note that we stated explicit constants in the result above as we felt that this eases reading, but we did not try to optimize them. For example, a selection rate of at most some constant less than $\frac{1}{2e}$ can give the same run time guarantee when raising λ by a sufficiently large constant factor. A selection rate of at most some constant less than $\frac{1}{e}$ can also be tolerated. Now it takes a constant number of iterations to move a critical frequency to $1 - \frac{1}{n}$, so the run time guarantee increases by a constant factor.

4. Lower bound

Our main insight, which gave our sharper upper bound with a proof simpler than in previous works, was that the UMDA, when optimizing LeadingOnes in the regime of low genetic drift, makes a steady progress in each iteration: It sets the frequencies to the maximum value $1 - \frac{1}{n}$ in a left-to-right fashion, keeping the other frequencies close to the middle value of $\frac{1}{2}$. The increase of the number of frequencies at the maximum value, with a simple Chernoff bound argument, could be shown to be logarithmic in the reciprocal $\frac{1}{\mu/\lambda}$ of the selection rate.

In this section, we show that the same proof approach (with small modifications) can also be employed to show lower bounds, and in this case, a matching lower bound.

Theorem 6. *Let $\delta \in (0, 1)$ be a constant, and let $\zeta = \frac{3}{4}(1 + \delta)$. Consider the UMDA optimizing LeadingOnes with $\lambda \geq \mu \geq 64n \ln n$ and $\lambda \geq \frac{\mu}{\zeta}$. Further, let $d = \lceil \log_{4/3}(\zeta \frac{\lambda}{\mu}) \rceil$, and let $\xi = \lceil \log_{4/3}(n^2 \lambda) \rceil + 1$. Then the UMDA requires more than $\lambda \lfloor \frac{n-\xi}{d+1} \rfloor$ fitness function evaluations to sample the optimum with a probability of at least $1 - 4n^{-1}$.*

To prove a lower bound via the general idea laid out above, we need to show that frequencies that do not receive a fitness signal do not approach $1 - \frac{1}{n}$ due to genetic drift. Here we have to be slightly more careful than in our upper bound analysis, since now the fitness signal does move the frequencies into the undesired (from the view-point of lower bound proofs) direction. Consequently, we can employ the low-genetic drift argument only while we are sure that we do not receive a fitness signal (Lemma 3).

Using a Chernoff-type concentration argument (which in principle works similarly for upper and lower bounds), we show that at most roughly $\log \frac{\lambda}{\mu}$ frequencies above the critical position receive a fitness signal (and thus potentially leave the middle range), see Lemma 4.

Consequently, in the first $O(\frac{n}{\log \lambda/\mu})$ iterations, we have many frequencies that are far from the maximum value, and thus sampling the optimum is unlikely (Lemma 5). This yields our lower bound.

To make these arguments precise, we define when a frequency of the UMDA stops being neutral, that is, receives a fitness signal. To this end, we say that a position $i \in [n]$ is *selection-relevant* (with respect to LeadingOnes) in iteration $t \in \mathbb{N}$ if and only if the offspring population of the UMDA in iteration t has at least μ individuals with at least $i - 1$ leading 1s. Thus, with respect to selection, the bit value at position i decides whether an individual is selected for the update or not. We call the largest selection-relevant position in an iteration the *maximum* selection-relevant position. Note that all positions greater than the maximum selection-relevant position are neutral during this iteration.

Since, by the definition of a selection-relevant position i, all frequencies at positions less than i are set to $1 - \frac{1}{n}$, the critical position for the next iteration is i, too. Thus, bounding the progress of the selection-relevant position also bounds the overall progress of the UMDA on LeadingOnes.

We start by showing that each frequency stays in the interval $(\frac{1}{4}, \frac{3}{4})$ until its position becomes selection-relevant.

Lemma 3. *Consider the UMDA with $\lambda \geq \mu \geq 64n \ln n$. Further, for each position $i \in [n]$, let $t_i' \in \mathbb{N}$ denote the first iteration such that position i is selection-relevant, and let $t_i^{\text{sel}} = \min\{t_i', n\}$. Then, with a probability of at least $1 - 2n^{-1}$, within the first n iterations, for each position $i \in [n]$ and for each iteration $t \leq t_i^{\text{sel}}$, it holds that $p_i^{(t)} \in (\frac{1}{4}, \frac{3}{4})$.*

Proof. Consider a position $i \in [n]$. Note that, for all iterations $t \leq t_i^{\text{sel}}$, the frequency p_i is neutral. By Theorem 4 with $d = \frac{1}{4}$, we see that the probability that p_i leaves the interval $(\frac{1}{4}, \frac{3}{4})$ within the first $t_i^{\text{sel}} \leq n$ iterations is at most $2e^{-\mu/(32 \cdot t_i^{\text{sel}})} \leq 2e^{-\mu/(32 \cdot n)} \leq 2n^{-2}$, where we used our lower bound on μ.

Applying a union bound over all n frequencies yields that at least one frequency leaves the interval $(\frac{1}{4}, \frac{3}{4})$ within the first n iterations before being selection-relevant with a probability of at most $2n^{-1}$, which concludes the proof. □

We now show that the maximum selection-relevant position is only roughly $\log \frac{\lambda}{\mu}$ larger than the critical position during each iteration.

Lemma 4. *Let $\delta \in (0, 1)$ be a constant, and let $\zeta = \frac{3}{4}(1 + \delta)$. Consider the UMDA optimizing LeadingOnes with $\mu \geq 6\frac{1+\delta}{\delta^2} \ln n$ and $\lambda \geq \mu \cdot \max\{1, \frac{1}{\zeta}\}$. Furthermore, consider an iteration $t \in \mathbb{N}$ such that position $i \in [n]$ is critical and that, for all positions $j > i$, we have*

$p_j^{(t)} \leq \frac{3}{4}$. *Let* $d = \lceil \log_{4/3}(\zeta \frac{\lambda}{\mu}) \rceil$. *Then, with a probability of at least* $1 - n^{-2}$, *the maximum selection-relevant position for iteration* t *is at most* $\min\{n, i + d + 1\}$.

Proof. Note that $d \geq 0$ by our assumption on λ. Similar to the proof of Lemma 2, we consider the offspring population of λ individuals sampled in iteration t. Let X denote the number of individuals that have at least $i' := \min\{n, i + d + 1\}$ leading 1s. By assumption, all frequencies at positions greater than i are at most $\frac{3}{4}$. Thus, $\mathrm{E}[X] \leq \lambda(\frac{3}{4})^{1+d} = \lambda(\frac{4}{3})^{-(1+d)} \leq \frac{3}{4}\frac{\mu}{\zeta} \leq \frac{\mu}{1+\delta}$.

We now apply Theorem 2 in order to show that it is unlikely that at least μ individuals from iteration t have at least i' leading 1s. Using our bounds on μ and our estimate on $\mathrm{E}[X]$ from above, we compute

$$\Pr[X \geq \mu] \leq \Pr\left[X \geq (1 + \delta)\mathrm{E}[X]\right] \leq e^{-\frac{\delta^2 \mathrm{E}[X]}{3}}$$
$$\leq e^{-\frac{\delta^2 \mu}{3(1+\delta)}} \leq n^{-2}.$$

Thus, with a probability of at least $1 - n^{-2}$, fewer than μ individuals have at least i' leading 1s. This means that the maximum selection-relevant position in this iteration is in $[i']$. $\quad\square$

Before we prove our lower bound, we show that the UMDA does not sample the optimal solution of LeadingOnes with high probability while the critical position is at least logarithmically far away from the end.

Lemma 5. *Consider the UMDA optimizing* LeadingOnes *with* $\lambda \geq \mu$. *Further, consider an iteration* $t \in \mathbb{N}$ *and a position* $i \in [n]$ *such that, for all positions* $j > i$, *we have* $p_j^{(t)} \leq \frac{3}{4}$. *Then, with a probability of at least* $1 - \lambda(\frac{3}{4})^{n-i}$, *the UMDA does not sample the optimum in this iteration.*

Proof. By our assumption on the frequencies and on i, the probability that a single individual in the offspring population in iteration t is the all-1s string (that is, the optimum of LeadingOnes) is at most $(\frac{3}{4})^{n-i}$. Thus, the probability that none of the λ offspring is optimal is, by Bernoulli's inequality [15, Lemma 1.4.8], at least $(1 - (\frac{3}{4})^{n-i})^\lambda \geq 1 - \lambda(\frac{3}{4})^{n-i}$, as desired. $\quad\square$

We now prove our lower bound.

Proof of Theorem 6. We prove that the UMDA needs, with a probability of at least $1 - 4n^{-1}$, more than $\lfloor \frac{n-\xi}{d+1} \rfloor$ iterations until it samples the optimum. Since it performs λ fitness function evaluations each iteration, the theorem then follows.

In the following, we assume that all frequencies remain in the interval $(\frac{1}{4}, \frac{3}{4})$ for the first n iterations as long as they have never been selection-relevant. By Lemma 3, this happens with a probability of at least $1 - 2n^{-1}$.

We now prove by induction on the iteration index $t \in \mathbb{N}$ that, with a probability of at least $1 - (t+1)n^{-2}$, for each position $i > 1 + (t+1)(d+1)$, we have that position i is not selection-relevant up to iteration t.

For the base case $t = 0$, note that position $i = 1$ is critical and that all frequencies are $\frac{1}{2}$ and thus at most $\frac{3}{4}$. By Lemma 4, with a probability of at least $1 - n^{-2}$, the maximum selection-relevant position is at most $d + 2$. Thus, all positions greater than $d + 2$ are not selection-relevant up to iteration 0.

For the inductive step, we assume that our inductive hypothesis holds up to iteration t. Note that this means that, with a probability of at least $1 - (t+1)n^{-2}$, the maximum selection-relevant index is in $[1 + (t+1)(d+1)]$ and, thus, the critical position for iteration $t+1$ is in $[1 + (t+1)(d+1)]$. By Lemma 3, all frequencies at positions greater than $1 + (t+1)(d+1)$ are thus at most $\frac{3}{4}$.[1] Then, in iteration $t+1$, again by Lemma 4, with a probability of at least $1 - n^{-2}$, the maximum-selection relevant index is at most $1 + (t+1)(d+1) + d + 1 = 1 + (t+2)(d+1)$. Consequently, via a union bound on the error probabilities of the inductive hypothesis and the current iteration $t+1$, with a probability of at least $1 - (t+2)n^{-2}$, no position greater than $1 + (t+2)(d+1)$ is selection-relevant up to iteration $t+1$, which proves our claim.

We now assume that $n - \xi \geq 1$, as Theorem 6 is trivial otherwise. Our claim above then yields that, up to iteration $t' := \lfloor \frac{n-\xi}{d+1} \rfloor - 1$, with a probability of at least $1 - \frac{n-\xi}{d+1}n^{-2} \geq 1 - n^{-1}$, each position greater than $1 + n - \xi$ was never selection-relevant. This means that by Lemma 3, all such frequencies are at most $\frac{3}{4}$. By Lemma 5 with $i = 1 + n - \xi$, with a probability of at least $1 - \lambda(\frac{3}{4})^{n-i} = 1 - \lambda(\frac{3}{4})^{\xi-1} = 1 - n^{-2}$, the UMDA does not sample the optimum of LeadingOnes within a single iteration. Applying a union bound over the first $t' + 1 \leq n$ iterations, with a probability of at least $1 - n^{-1}$, the UMDA does not sample the optimum up to iteration t' (which are $t' + 1$ iterations).

Overall, by a union bound over all error probabilities, with a probability of at least $1 - 4n^{-1}$, the UMDA does not sample the optimum within the first $t' + 1 = \lfloor \frac{n-\xi}{d+1} \rfloor$ iterations, which concludes the proof. $\quad\square$

For the sake of completeness, we state the combined result of our upper and lower bound.

---

[1] Note that such frequencies are at most $\frac{3}{4}$ with a probability of 1, as we condition on this event throughout the proof, as stated at the beginning of the proof.

ARTICLE IN PRESS

Corollary 1 (*combining Theorems 5 and 6*). *Let C be a sufficiently large constant. Consider the UMDA optimizing* LeadingOnes *with* $\lambda \geq C\mu \geq 128 n \ln n$ *and with* λ *being bounded from above by a polynomial in n. With a probability of at least* $1 - 9n^{-1}$*, it samples the optimum after* $\Theta\left(\lambda \frac{n}{\log(\lambda/\mu)}\right)$ *fitness function evaluations.*

Proof. From the assumptions of Theorems 5 and 6, we take the stricter ones. The additive term $\lceil \frac{n}{n-1} e \ln n \rceil$ in Theorem 5 vanishes in asymptotic notation, and the term $n - \xi$ in Theorem 6 is $\Omega(n)$, due to λ being bounded from above by a polynomial in ξ. Taking the union bound over the failure probabilities of both theorems concludes the proof. □

5. Conclusion

We improved the best known upper bound for the run time of the UMDA on LeadingOnes for the case of $\mu \in \Omega(n \log n)$ from $O(n\lambda \log \lambda)$ to $O\left(n \frac{\lambda}{\log(\lambda/\mu)}\right)$. This result improves the previous best result both by removing an unnecessary $\log \lambda$ factor and, not discussed in previous works, by gaining a $\log(\lambda/\mu)$ factor and thus showing an advantage of using a low selection rate μ/λ. We obtained these results via a different proof method that avoids the technically deeper level-based method. Our arguments can also be employed for lower bounds. Overall, we provide a run time estimate for the UMDA on LeadingOnes that is tight up to constant factors and that holds with high probability.

We note that the general proof idea can be extended also to the parameter regime of $\mu \in o(n \log n)$ for the UMDA. We conjecture that a more general upper bound of the UMDA (with $\lambda \in \Omega(\log n)$) on LeadingOnes is

$$O\left(\lambda \left(n + \frac{n}{e^{\mu/n}}\left(\frac{n}{\lambda} + \log\min\{\mu, n\}\right)\right)\right).$$

Speaking in terms of iterations and thus ignoring the factor of λ, this expression can be explained as follows: the first term of n considers $O(n)$ frequencies that do not drop below constant values. Each of these frequencies is set to $1 - \frac{1}{n}$ within a constant number of iterations with high probability. Since $\lambda \in \Omega(\log n)$, frequencies at $1 - \frac{1}{n}$ do not drop until the optimum is sampled with high probability.

The second, more complicated term deals with frequencies that, pessimistically, reached the lower border $\frac{1}{n}$. There are $O(n/e^{\mu/n})$ of these frequencies, by the same argument as used in the proof of Lemma 1. The other factor is concerned with the time it takes a critical frequency to be increased to $1 - \frac{1}{n}$ with high probability. Here, a case distinction needs to be made with respect to whether $\mu \geq n$. The inverse of the maximum of μ and n determines the step size in which a critical frequency can be increased. The term $\log\min\{\mu, n\}$ stems from the multiplicative up-drift [17] of such a frequency in order to reach a constant value. Afterward, it is set to $1 - \frac{1}{n}$ within a constant number of iterations (as the first $O(n)$ frequencies). Last, the term n/λ is only important if $\lambda \in o(n)$ and denotes the waiting time for a critical frequency to sample at least one 1 with λ tries (given that the prefix consists of only 1s).

Last, we are positive that our proof technique is applicable to a greater class of EDAs and to functions that have a hierarchical structure similar to that of LeadingOnes. In order to transfer the proof of the upper bound to other univariate EDAs, only Lemmas 1 and 2 need to be adjusted to the specific algorithm, which should work similarly for other EDAs too. For the lower bound, Lemmas 3 to 5 need to be changed.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] M. Pelikan, M. Hauschild, F.G. Lobo, Estimation of distribution algorithms, in: Springer Handbook of Computational Intelligence, Springer, 2015, pp. 899–928.

[2] M.S. Krejca, C. Witt, Theory of estimation-of-distribution algorithms, in: Theory of Evolutionary Computation—Recent Developments in Discrete Optimization, Springer, 2020, pp. 405–442, also available at http://arxiv.org/abs/1806.05392.

[3] D. Dang, P.K. Lehre, Simplified runtime analysis of estimation of distribution algorithms, in: Proc. of GECCO '15, 2015, pp. 513–518.

[4] H. Mühlenbein, G. Paaß, From recombination of genes to the estimation of distributions I. Binary parameters, in: Proc. of PPSN '96, 1996, pp. 178–187.

[5] P.K. Lehre, P.T.H. Nguyen, Improved runtime bounds for the univariate marginal distribution algorithm via anti-concentration, in: Proc. of GECCO '17, 2017, pp. 1383–1390.

[6] C. Witt, Upper bounds on the runtime of the univariate marginal distribution algorithm on OneMax, in: Proc. of GECCO '17, 2017, pp. 1415–1422.

[7] C. Witt, Upper bounds on the running time of the univariate marginal distribution algorithm on OneMax, Algorithmica 81 (2019) 632–667, https://doi.org/10.1007/s00453-018-0463-0.

[8] D. Dang, P.K. Lehre, P.T.H. Nguyen, Level-based analysis of the univariate marginal distribution algorithm, Algorithmica 81 (2019) 668–702, https://doi.org/10.1007/s00453-018-0507-5.

[9] P.K. Lehre, P.T.H. Nguyen, Runtime analysis of the univariate marginal distribution algorithm under low selective pressure and prior noise, in: Proc. of GECCO '19, 2019, pp. 1497–1505.

[10] J. Lengler, D. Sudholt, C. Witt, Medium step sizes are harmful for the compact genetic algorithm, in: Proc. of GECCO '18, 2018, pp. 1499–1506.

[11] P.K. Lehre, P.T.H. Nguyen, On the limitations of the univariate marginal distribution algorithm to deception and where bivariate EDAs might help, in: Proc. of FOGA '19, 2019, pp. 154–168.

[12] B. Doerr, M.S. Krejca, The univariate marginal distribution algorithm copes well with deception and epistasis, in: Proc. of EvoCOP '20, vol. 12102, 2020, pp. 51–66.

[13] M.S. Krejca, C. Witt, Lower bounds on the run time of the univariate marginal distribution algorithm on onemax, Theor. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004.

[14] B. Doerr, A tight runtime analysis for the cGA on jump functions: EDAs can cross fitness valleys at no extra cost, in: Proc. of GECCO '19, 2019, pp. 1488–1496.

[15] B. Doerr, Probabilistic tools for the analysis of randomized optimization heuristics, in: Theory of Evolutionary Computation—Recent Developments in Discrete Optimization, Springer International Publishing, 2020, pp. 1–87, also available at https://arxiv.org/abs/1801.06733.

[16] B. Doerr, W. Zheng, Sharp bounds for genetic drift in estimation of distribution algorithms, IEEE Trans. Evol. Comput. (2020), https://doi.org/10.1109/TEVC.2020.2987361, early access.

[17] B. Doerr, T. Kötzing, Multiplicative up-drift, in: Proc. of GECCO '19, 2019, pp. 1470–1478.