Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax

Martin S. Krejca [a,*], Carsten Witt [b,*]

[a] Hasso Platter Institute, University of Potsdam, Postdam, Germany
[b] DTU Compute, Technical University of Denmark, Kongens Lyngby, Denmark

ARTICLE INFO

Article history:
Received 30 October 2017
Received in revised form 19 April 2018
Accepted 1 June 2018
Available online xxxx

Keywords:
Estimation-of-distribution algorithm
Run time analysis
Lower bound

ABSTRACT

The Univariate Marginal Distribution Algorithm (UMDA) – a popular estimation-of-distribution algorithm – is studied from a run time perspective. On the classical OneMax benchmark function on bit strings of length n, a lower bound of $\Omega(\lambda + \mu\sqrt{n} + n\log n)$, where μ and λ are algorithm-specific parameters, on its expected run time is proved. This is the first direct lower bound on the run time of UMDA. It is stronger than the bounds that follow from general black-box complexity theory and is matched by the run time of many evolutionary algorithms. The results are obtained through advanced analyses of the stochastic change of the frequencies of bit values maintained by the algorithm, including carefully designed potential functions. These techniques may prove useful in advancing the field of run time analysis for estimation-of-distribution algorithms in general.

1. Introduction

Traditional algorithms in the field of Evolutionary Computation optimize problems by sampling a certain amount of solutions from the problem domain, the so-called *population*, and transforming them, such that the new population is closer to an optimum. *Estimation-of-distribution algorithms* (EDAs; [1]) have a very similar approach but do not store an explicit population of sample solutions. Instead, they store a probability distribution over the problem domain and update it via an algorithm-specific rule that learns from samples drawn from said distribution.

Although many different variants of EDAs (cf. [2]) and many different domains are possible, theoretical analyses of EDAs in discrete search spaces often consider run times over $\{0,1\}^n$. Further, the focus is on EDAs that store a Poisson binomial distribution, i.e., EDAs that store a probability vector p of n independent probabilities, each component p_i denoting the probability that a sampled bit string will have a 1 at position i.

The first theoretical analysis in this setting was conducted by Droste [3], who analyzed the *compact Genetic Algorithm* (cGA) – an EDA that only samples two solutions each iteration – on linear functions. Papers considering other EDAs, like, e.g., an iteration-best *Ant Colony Optimization* (ACO) algorithm by Neumann et al. [4] followed, where the pheromone vector represents the probability vector of an EDA.

Recently, the interest in the theoretical analysis of EDAs has increased [5–10]. Most of these works derive upper bounds for a specific EDA on the popular OneMax function, which counts the number of 1s in a bit string and is considered to be one of the easiest functions with a unique optimum for most EAs [11,12]. The only exceptions are Friedrich et al. [6], who

* Corresponding author.
E-mail addresses: martin.krejca@hpi.de (M.S. Krejca), cawi@dtu.dk (C. Witt).

https://doi.org/10.1016/j.tcs.2018.06.004
0304-3975/

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

look at general properties of EDAs, Sudholt and Witt [9], who derive *lower bounds* on OneMax for the aforementioned cGA and an iteration-best ACO, and Dang and Lehre [5], who focus on general methods for upper bounds.

In this paper, we follow the ideas of Sudholt and Witt [9] and derive a lower bound of $\Omega(n \log n)$ for the *Univariate Marginal Distribution Algorithm* (UMDA; [13]) on OneMax, which is a typical lower bound for many evolutionary algorithms on this function. UMDA is an EDA that samples λ solutions each iteration, selects $\mu < \lambda$ best solutions, and then sets p_i to the relative occurrence of 1s among these μ individuals. The algorithm has already been analyzed some years ago for several artificially designed example functions [14–17]. However, none of these papers consider the standard benchmark function for theory: the OneMax function. In fact, the run time analysis of UMDA on the simple OneMax function has turned out to be rather challenging; the first such result, showing an upper bound of $O(n \log n \log\log n)$ on its expected run time for certain settings of μ and λ, was not published until 2015 [5]. Specific lower bounds for UMDA were to date missing; the previous best result $\Omega(n/\log n)$ on the expected run time followed from general black box complexity theory [18] and did not shed light on the working principles of UMDA.

Recently, two matching upper bounds of $O(n \log n)$ of UMDA on OneMax have been proved independently from one another [8,10] for certain cases of μ and λ. Our results match almost all of these cases, providing a tight run time bound of $\Theta(n \log n)$.

The concepts of the proofs in this paper are based on the prior work from Sudholt and Witt [9]. However, analyzing UMDA is much more difficult than analyzing cGA or iteration-best ACO, since the update of the latter algorithms is bounded by an algorithm-specific parameter and the algorithms only have up to three distinct successor states for each value p_i. UMDA, on the other hand, can change each of its p_i to any value x/μ with a certain probability, where $x \in \{0, \dots, \mu\}$, due to the nature of its update rule. This makes analyzing UMDA far more involved, because every single update has to be bounded probabilistically. Further, the simple update rules for cGA and iteration-best ACO allow for a distinction into two cases that determines whether a value p_i will increase or decrease; a fact that was heavily exploited in the analyses in [9]. For UMDA, no such simple case distinction can be made.

This paper is structured as follows: in Section 2, we shortly introduce the setting we are going to analyze and go into detail about UMDA's update rule, that is, we explain and analyze a property of the algorithm that leads to the lower bound when optimizing OneMax.

Then in Section 3, we state our main result – a run time bound of $\Omega(n \log n)$ (Theorem 6) – and prove it step by step. The rough outline of the proof follows the one presented in [9]. However, we think that our style of presentation is more accessible, due to dissecting our proof into smaller (and often independent) lemmas.

In Section 4, we relax the condition of Theorem 6 with respect to the dependency of μ to λ and also prove a bound of $\Omega(n \log n)$ (Theorem 20). Our result holds for values of $\mu \leq c \log n$, for a sufficiently small constant c. This includes the case $\mu = 1$, for which no matching upper bound has explicitly been proved up to date.[1]

Finally, we conclude and discuss our results and future work in the Conclusions section.

We think that parts of our results (especially the detailed analysis of the selection process in Section 2.2) can also be used when analyzing UMDA on other functions than OneMax.

This version is an extension of our prior lower-bound analysis of UMDA [20] in the way that we also consider the case of $\mu \leq c \log n$ (for a sufficiently small constant c), independent of $\lambda > \mu$.

2. Preliminaries

We consider the *Univariate Marginal Distribution Algorithm* (UMDA [13]; Algorithm 1) maximizing the pseudo-Boolean function OneMax, where, for all $x \in \{0,1\}^n$,

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

Note that the function's unique maximum is the all-ones bit string. However, a more general version can be defined by choosing an arbitrary optimum $a \in \{0,1\}^n$ and defining, for all $x \in \{0,1\}^n$,

$$\text{OneMax}_a(x) = n - d_{\text{H}}(x,a) \,,$$

where $d_{\text{H}}(x,a)$ denotes the Hamming distance of the bit strings x and a. Note that OneMax_{1^n} is equivalent to the original definition of OneMax. Our analyses hold true for any function OneMax_a, with $a \in \{0,1\}^n$, due to symmetry of UMDA's update rule.

We call bit strings *individuals* and their respective OneMax values *fitness*.

UMDA does not store an explicit population but does so implicitly, which makes it an *estimation-of-distribution algorithm* (EDA). For each of the n different bit positions, it stores a rational number p_i, which we call *frequency*, determining

[1] Note that for this case, UMDA (with frequency borders) basically is a $(1,\lambda)$ EA with standard bit mutation, for which matching upper bounds have been proved [19].

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

Algorithm 1: Univariate Marginal Distribution Algorithm (UMDA).

$t \leftarrow 0$;
$p_{1,t} \leftarrow p_{2,t} \leftarrow \cdots \leftarrow p_{n,t} \leftarrow \frac{1}{2}$;
while termination criterion not met **do**
  $P_t \leftarrow \emptyset$;
  for $j \in \{1, \dots, \lambda\}$ **do**
    for $i \in \{1, \dots, n\}$ **do**
      $x_{i,t}^{(j)} \leftarrow 1$ with prob. $p_{i,t}$, $x_{i,t}^{(j)} \leftarrow 0$ with prob. $1 - p_{i,t}$;
    $P_t \leftarrow P_t \cup \{x_t^{(j)}\}$;
  Sort individuals in P descending by fitness, breaking ties uniformly at random;
  for $i \in \{1, \dots, n\}$ **do**
    $p_{i,t+1} \leftarrow \frac{1}{\mu} \sum_{j=1}^{\mu} x_{i,t}^{(j)}$;
    Restrict $p_{i,t+1}$ to be within $\left[\frac{1}{n}, 1 - \frac{1}{n}\right]$;
  $t \leftarrow t + 1$;

how likely it is that a hypothetical individual would have a 1 at this position. In other words, UMDA stores a probability distribution over $\{0, 1\}^n$. The starting distribution is the uniform distribution.

In each iteration, UMDA samples λ individuals such that each individual has a 1 at position i ($i \in \{1, \dots, n\}$) with probability p_i, independent of all of the other frequencies. Thus, individuals are sampled such that their number of 1s follows a Poisson binomial distribution with probability vector $(p_i)_{i \in \{1,\dots,n\}}$.

After sampling λ individuals, μ of them with highest fitness are chosen, breaking ties uniformly at random (so-called *selection*). Then, for each position, the respective frequency is set to the relative occurrence of 1s in this position. That is, if x of the chosen μ best individuals have a 1 at position i, the frequency p_i will be updated to x/μ for the next iteration. Note that such an update allows large jumps like, e.g., from $(\mu - 1)/\mu$ to $1/\mu$, spanning almost the entire interval of a frequency!

If a frequency reaches either 0 or 1, it cannot change anymore, since then all bits at this position will be either 0 or 1. To prevent UMDA from getting stuck in such a way, we narrow the interval of possible frequencies down to $[1/n, 1 - 1/n]$. This way, there is always a chance of sampling 0s and 1s for each position. This is a common approach used by other EDAs as well, such as cGA or ACO algorithms (mentioned in the introduction).

Overall, we are interested in a lower bound of UMDA's expected number of *function evaluations* on OneMax until the optimum is sampled. Note that this is at least the expected number of iterations until the optimum is sampled (minus one) times λ, as we do not necessarily have to evaluate all λ offspring in the last iteration.

In all of our calculations except Section 4, we always assume that $\lambda = (1 + \beta)\mu$, for some constant $\beta > 0$. Of course, we could also choose $\lambda = \omega(\mu)$ but then each iteration would be even more expensive. Choosing $\lambda = \Theta(\mu)$ lets us basically focus on the minimal number of function evaluations per iteration, as μ of them are at least needed to make an update.

Given two random variables X and Y, we say that X *dominates* Y, written as $X \succeq Y$, if, for all x, $\Pr(X \geq x) \geq \Pr(Y \geq x)$.

2.1. Selecting individuals

In order to optimize a function efficiently, UMDA needs to evolve its frequencies toward the right direction, making it more likely to sample an optimum. In the setting of OneMax, this means that each frequency should be increased (toward a value of $1 - 1/n$). This is where selection comes into play.

By selecting μ best individuals every iteration w.r.t. their fitness, we hope that many of them have correctly set bits at each position, such that the respective frequencies increase. However, even in the simple case of OneMax, where a 1 is always better than a 0, there is a flaw in the update process that prevents UMDA from optimizing OneMax too fast. To see why this flaw occurs, consider an arbitrary position j in the following.

When selecting individuals for an update to p_j, UMDA does so by always considering the fitness of each *entire* individual. That is, although each frequency is independently updated from the others, selection is done w.r.t. *all* positions at once. Thus, when looking at position j, it can happen that we have many 0s, because the individuals chosen for the update may have many 1s in their remaining positions, which can lead to a decrease of p_j.

Since having a 1 at a position is always better than a 0 when considering OneMax, the selection is biased, pushing for more 1s at each position. However, this bias is not necessarily too large: Consider that for each individual each bit but bit j has already been sampled. When looking at selection w.r.t. only $n - 1$ bits in each individual, some individuals may already be so good that they are determined to be chosen for selection, whereas others may be so bad that they definitely cannot be chosen for selection, regardless of the outcome of bit j.

Consider the fitness of all individuals sampled during one iteration of UMDA w.r.t. $n - 1$ bits, i.e., all bits but bit j. We call each of these n different fitness values (from 0 to $n - 1$) a *level*. Assume that the individuals are sorted decreasingly by their level; each individual having a unique index. Let w^+ be the level of the individual with rank μ, and let w^- be the level of the individual with rank $\mu + 1$. Since bit j has not been considered so far, its value can potentially increase each individual's level by 1. Now assume that $w^+ = w^- + 1$. Then, individuals from level w^- can end up with the same fitness

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

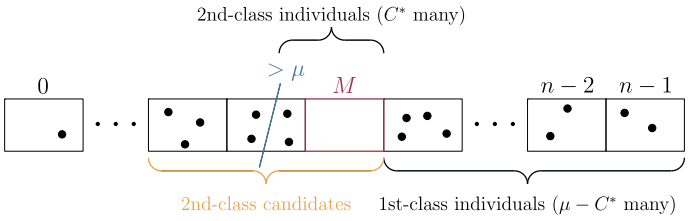

Fig. 1. An exemplary visualization of the different definitions we need. The boxes depict all of the n levels, the numbers above show their respective fitness, and the dots symbolize individuals in these levels. The line cutting through level $M-1$ marks the point where more than μ individuals have been sampled when starting from the top. In that level, not all individuals are going to be selected. Further, the individuals from the level below can be selected (as their fitness can still increase by one when sampling the last bit), and individuals from the level above can be not selected. Hence, the individuals in those levels are 2nd-class candidates. The individuals in higher levels will always be selected, thus they are 1st-class individuals. Out of the 2nd-class candidates, those individuals that are chosen during selection are the 2nd-class individuals (in this example, those would be two individuals, i.e., $C^* = 2$). Last, M depicts the cut level, i.e., the topmost level such that the number of sampled individuals is greater than μ when including the next (lower) level.

as individuals from level w^+, once bit j has been sampled. Thus, individuals from level w^+ were still *prone to selection*. This means that the outcome of bis j can influence whether the individual is being selected or not.

Among the μ individuals chosen during selection, we distinguish between two different types: 1st-class and 2nd-class individuals. 1st-class individuals are those which are chosen during selection no matter which value bit j has. The remaining of the μ individuals are the 2nd-class individuals; they had to compete with other individuals for selection. Therefore, their bit value j is biased toward 1 compared to 1st-class individuals. Note that 2nd-class individuals can only exist if $w^+ \leq w^- + 1$, since in this case, individuals from level w^- can still be as good as individuals from level w^+ after sampling bit j.

Let X_t be the number of 1s at position j of the μ selected individuals in iteration t of UMDA, and let C^* denote the number of 2nd-class individuals in iteration $t+1$. Note that the number of 1s of 1st-class individuals during iteration $t+1$ follows a binomial distribution with success probability X_t/μ. Since we have $\mu - C^*$ 1st-class individuals, the distribution of the number of 1s of these follows $\mathrm{Bin}(\mu - C^*, X_t/\mu)$. Note that the actual frequency in iteration $t+1$ might be set to either $1/n$ or $1-1/n$ if the number of 1s in the μ selected individuals is too close to 0 or μ, respectively. We will be able to ignore this fact in our forthcoming analyses since all considerations are stopped when a frequency drops below $1/n$ or exceeds $1-1/n$.

2.2. The number of 2nd-class individuals

As in the previous section, consider again a bit position j. In this section, we again speak of levels as defined in the previous section. Those definitions as well as the following ones are also depicted in Fig. 1. Level $n-1$ is the topmost, and level 0 is the bottommost. For all $i \in \{0, \dots, n-1\}$, let C_i denote the cardinality of level i, i.e., the number of individuals in level i during an arbitrary iteration of UMDA, and let $C_{\geq i} = \sum_{a=i}^{n-1} C_a$.

Let M denote the index of the first level from the top such that the number of sampled individuals is greater than μ when including the following level, i.e.,

$$M = \max\{i \mid C_{\geq i-1} > \mu\} .$$

Note that M can never be 0, and only if $M = n-1$, C_M can be greater than μ. Note further that C_M can be 0.

Due to the definition of M, if $M \neq n-1$, level $M-1$ contains the individual of rank $\mu+1$, as described in the previous section. Thus, levels M, $M-1$, and $M-2$ contain all of the individuals that are prone to selection (if such exist at all). Hence, individuals in levels at least $M+1$ are definitely 1st-class individuals. 2nd-class individuals, if any, have to come from level M, $M-1$, or $M-2$. We call the individuals from these three levels *2nd-class candidates*. Note that the actual number of 2nd-class individuals is bounded from above by $\mu - C_{\geq M+1} = \mu - C_{\geq M} + C_M$, since exactly μ individuals are selected.

Since the 2nd-class individuals are the only ones that are prone to selection and thus the only ones that actively help in progressing a single frequency toward $1-1/n$, it is of utmost importance to understand the distribution of $C^* := \mu - C_{\geq M+1}$, that is, the biased impact to an update as introduced in Section 2.1. Moreover, we will also need a bound on the number of 2nd-class candidates.

Before we get to analyzing the 2nd-class individuals, we introduce several auxiliary statements. We start with a very useful lemma on conditional binomially distributed random variables.

Lemma 1. *Let X be a binomially distributed random variable with arbitrary parameters. Then for any $x, y \geq 0$, it holds*

$$\Pr(X \geq x + y \mid X \geq x) \leq \Pr(X \geq y) .$$

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

Proof. Let n and p be the parameters of the underlying binomial distribution. Given $x \geq 0$, we define the random variable $Y_x := X - x$. Conditioning on $X \geq x$, we have $Y_x \sim \mathrm{Bin}(k, p)$ for $0 \leq k \leq n - x$ and therefore $Y_x \preceq X$. Hence, $\Pr(X \geq x + y \mid X \geq x) = \Pr(Y_x \geq y \mid X \geq x) \leq \Pr(X \geq y)$. □

Moreover, we are going to use a corollary that is based on Lemma 8 from [9], the proof of which can be seen in [21, Lemma 9]. Also, the idea behind the corollary is given in [21] but not presented as an independent statement.

Lemma 2. *Let S be the sum of m independent Poisson trials with probabilities $p_i \in [1/6, 5/6]$ for all $i \in \{1, \dots, m\}$. Then, for all $0 \leq s \leq m$, $\Pr(S = s) = O(1/\sqrt{m})$.*

Corollary 3. *Let X be the sum of n independent Poisson trials with probabilities p_i, $i \in \{1, \dots, n\}$. Further, let $\Theta(n)$ many p_i-s be within $[1/6, 5/6]$. Then, for all $0 \leq x \leq n$, $\Pr(X = x) = O(1/\sqrt{n})$.*

Proof. Let $m = \Theta(n)$ denote the number of p_i-s that are within $[1/6, 5/6]$. When sampling X, assume w.l.o.g. that the first m trials are the ones with $p_i \in [1/6, 5/6]$. Let S denote the sum of these trials, and let Y denote the sum of the remaining $n - m$ trials. Since the trials are independent, we get $\Pr(X = x) = \sum_{s=0}^{x} \Pr(S = s) \Pr(Y = x - s)$.

We can upper-bound $\Pr(S = s) = O(1/\sqrt{m}) = O(1/\sqrt{n})$ by using Lemma 2 and $m = \Theta(n)$. Thus, we have $\Pr(X = x) = O(1/\sqrt{n}) \sum_{s=0}^{x} \Pr(Y = x - s)$. Bounding the sum by 1 concludes the proof. □

The corollary lets us easily get upper bounds for the probability that a sampled individual has a certain (and arbitrary) fitness (w.r.t. either all n positions or all positions but j). In order to apply it, we have to make sure that $\Theta(n)$ frequencies are still within $[1/6, 5/6]$. Thus, we assume from now on that this assumption holds. In Section 3.2, we will go into detail and prove under which circumstances this assumption holds.

Note that all statements from now on regarding a specific position j hold regardless of the bits at any other of the $\Theta(n)$ positions that do not stay within $[1/6, 5/6]$. This means that the statements are even true if the bits at those other positions are chosen by an adversary.

We are now ready to analyze C^* and the number of 2nd-class candidates.

Lemma 4. *Consider UMDA with $\lambda = (1 + \beta)\mu$ optimizing* OneMax*, and let $\widetilde{Z}$ be a random variable that takes values in $\{1, \dots, \lambda\}$ only with probability at most $2e^{-(\varepsilon^2/(3+3\varepsilon))\mu} = e^{-\Omega(\mu)}$ and is 0 otherwise, where $\varepsilon > 0$ is a constant such that $\varepsilon < 1 - 1/(1 + \beta)$. If there are $\Theta(n)$ frequencies in $[1/6, 5/6]$, then the distribution of C^* is stochastically dominated by $\mathrm{Bin}\big(\lambda, O(1/\sqrt{n})\big) + \widetilde{Z}$ and the distribution of $C_M + C_{M-1} + C_{M-2}$ is stochastically dominated by $1 + \mathrm{Bin}\big(\lambda, O(1/\sqrt{n})\big) + \widetilde{Z}$.*

Proof. The proof carefully investigates and then reformulates the stochastic process generating the λ individuals (before selection), restricted to $n - 1$ bits. Each individual is sampled by a Poisson binomial distribution for a vector of probabilities $p' = (p'_1, \dots, p'_{n-1})$ obtained from the frequencies of UMDA by leaving one entry out. By counting its number of 1s, each of the λ individuals then falls into some level i, where $0 \leq i \leq n - 1$, with some probability q_i depending on the vector p'. Since the individuals are created independently, the number of individuals in level i is binomially distributed with parameters λ and q_i.

Next, we take an alternative view on the process of putting individuals into levels, using the principle of deferred decisions. We imagine that the process first samples all individuals in level 0 (through λ trials, all of which either hit the level or not), then (using the trials which did not hit level 0) all individuals in level 1, and so on, up to level $n - 1$.

The number of individuals C_0 in level 0 is still binomially distributed with parameters λ and q_0. However, after all individuals in level 0 have been sampled, the distribution changes. We have $\lambda - C_0$ trials left, each of which can hit one of the levels 1 to $n - 1$. In particular, such a trial will hit level 1 with probability $q_1/(1 - q_0)$, by the definition of conditional probability, since level 0 is excluded. This holds independently for all of the $\lambda - C_0$ trials so that C_1 follows a binomial distribution with parameters $\lambda - C_0$ and $q_1/(1 - q_0)$. Inductively, also all C_i for $i > 1$ are binomially distributed; e.g., C_{n-1} is distributed with parameters $\lambda - C_{n-2} - \cdots - C_0$ and 1. Note that this model of the sampling process can also be applied for any other permutation of the levels; we will make use of this fact.

Analyzing the number of 2nd-class individuals. We first focus on $C^* = \mu - C_{\geq M+1}$ and will later use bounds on its distribution to analyze $C_M + C_{M-1} + C_{M-2}$. Formally, by applying the law of total probability, the distribution of C^* looks as follows for $k \in \{0, \dots, \lambda\}$:

$$\Pr(C^* \geq k) = \sum_{i=1}^{n-1} \Pr(M = i) \cdot \Pr(\mu - C_{\geq i+1} \geq k \mid M = i) \,. \tag{1}$$

We will bound the terms of the sum differently with respect to the index i. First, we look into a particular value i^* such that $\Pr(M \geq i^*)$ is exponentially unlikely, and then make a case distinction via i^*.

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

Let X be the number of 1s in a single individual sampled (without conditioning on certain levels being hit). Choose i^* such that $\Pr(X \geq i^* - 1) \leq 1/\big((1+\varepsilon)(1+\beta)\big)$ and $\Pr(X \geq i^* - 1) \geq 1/\big((1+\varepsilon)(1+\beta)\big) - O(1/\sqrt{n})$. Such an i^* must exist, since every level is hit with probability $O(1/\sqrt{n})$ when sampling an individual, according to Corollary 3. Clearly, we also have $i^* \leq n-1$.

A crucial observation is that $\Pr(M \geq i^*) = e^{-\Omega(\mu)}$, since the expected number of individuals sampled with at least $i^* - 1$ 1s is at most $\lambda/\big((1+\varepsilon)(1+\beta)\big) = \mu/(1+\varepsilon)$, and the probability of sampling at least $(1+\varepsilon)\cdot\mu/(1+\varepsilon) = \mu$ is at most $e^{-\varepsilon^2\cdot\mu/(3(1+\varepsilon))} = e^{-\Omega(\mu)}$ by Chernoff bounds. Note that we have considered level $i^* - 1$ since $C_{\geq i^*-1} < \mu$ implies $M < i^*$.

In Equation (1), considering the partial sum for all $i \geq i^*$, we therefore immediately estimate

$$\sum_{i=i^*}^{n-1} \Pr(M = i) \cdot \Pr(\mu - C_{\geq i+1} \geq k \mid M = i) \leq \Pr(M \geq i^*) \leq e^{-\Omega(\mu)} ,$$

as shown just before.

For the terms with $i < i^*$ (in particular, the case $i = n-1$ is excluded), we take a view on the final expression in Equation (1) and focus on $\Pr(\mu - C_{\geq i+1} \geq k \mid M = i)$, in which we want to reformulate the underlying event appropriately. Here we note that

$$\{\mu - C_{\geq i+1} \geq k\} \cap \{M = i\}$$

is equivalent to

$$\{C_{\leq i} \geq \lambda - \mu + k\} \cap \{M = i\} ,$$

where $C_{\leq i} = \sum_{j=0}^{i} C_j$, and, using the definition of M, this is also equivalent to

$$\{C_{\leq i} \geq \lambda - \mu + k\} \cap \{C_{\leq i-2} < \lambda - \mu\} \cap \{C_{\leq i-1} \geq \lambda - \mu\} .$$

We now take the above-mentioned view on the stochastic process and assume that levels 0 to $i-2$ have been sampled and a number of experiments in a binomial distribution is carried out to determine the individuals from level $i-1$. Hence, given some $C_{\leq i-2} = a < \lambda - \mu$, our event is equivalent to that the event

$$E^* := \big\{C_i + C_{i-1} \geq (\lambda - \mu - a) + k\big\} \cap \{C_{i-1} \geq \lambda - \mu - a\}$$

happens.

Recall from our model above that C_{i-1} follows a binomial distribution with $\lambda - a$ trials and with a certain success probability s; similarly, C_i follows a binomial distribution with parameters $\lambda - a - C_{i-1}$ and s'. As we are interested in a cumulative distribution, we may pessimistically upper-bound the total number of trials for C_{i-1} by λ. Regarding s, note that it denotes the probability to sample an individual with $i-1$ 1s, given that it cannot have less than $i-1$ 1s. Note further that $\Pr(X \geq i^* - 1)$, where X again denotes the level of the individual sampled in a trial, is a lower bound for all probabilities $\Pr(X \geq i - 1)$, since $i < i^*$. To upper-bound s, we use Corollary 3, which tells us that the unconditional probability to hit a level is in $O(1/\sqrt{n})$, regardless of which level is hit. However, we have to condition on the event that certain levels (namely $0, \dots, i-2$, where $i < i^*$) cannot be hit anymore. We pessimistically exclude even some more levels than possible, more precisely, we exclude the levels from 0 up to $i^* - 2$. This means that we condition on $\Pr(X \geq i^* - 1)$. By the definition of conditional probability, the probability of $O(1/\sqrt{n})$ from Corollary 3 thus gets increased by a factor of $1/\Pr(X \geq i^* - 1)$, which is constant. Hence, C_{i-1} is stochastically dominated by a binomial distribution with parameters λ and $O(1/\sqrt{n})$.

Similarly, assuming that also level $i-1$ has been sampled, C_i is dominated by a binomial distribution with parameters $\lambda - C_{i-1}$ and $O(1/\sqrt{n})$.

To finally bound $\Pr(E^*)$ from above, which involves a condition on the outcome on C_{i-1}, we apply Lemma 1, where we let $X := C_{i-1}$ and $x = \lambda - \mu - a$ as well as $y = k$. Since we have bounded C_{i-1} (without the condition on $C_{i-1} \geq x$) by a binomial distribution with success probability $O(1/\sqrt{n})$, we get from the lemma that $\Pr(C_{i-1} - x \geq k \mid C_{i-1} \geq x) \leq \Pr\big(\mathrm{Bin}(\lambda, O(1/\sqrt{n})) \geq k\big)$. Note that the right-hand side is a bound independent of $C_0, \dots, C_{i-1}$. With respect to C_i, we do not consider an additional condition on its outcome but use the result $\Pr(C_i \geq k) \leq \Pr\big(\mathrm{Bin}(\lambda - C_{i-1}, O(1/\sqrt{n})) \geq k\big)$ derived in the last paragraph directly. Hence, both $C_{i-1} - x$, conditioned on $C_{i-1} \geq x$, and C_i have been bounded by binomial distributions with second parameter $O(1/\sqrt{n})$. In E^*, we are confronted with the sum of these two random variables and study the distribution of the sum. Together, $\Pr(E^*) \leq \Pr\big(\mathrm{Bin}(\lambda, O(1/\sqrt{n})) \geq k\big)$, since we consider at most λ trials. Pulling this term in front of the sum over i for the terms $i < i^*$ in (1) and estimating this sum with 1 (since we sum over mutually disjoint events) leaves us with an additional term of $\Pr\big(\mathrm{Bin}(\lambda, O(1/\sqrt{n})) \geq k\big)$ for $\Pr(\mu - C_{\geq M+1} \geq k)$. This proves the lemma's statement on the distribution of C^*.

Analyzing the number of 2nd-class candidates. We are left with analyzing $C^{**} := C_M + C_{M-1} + C_{M-2}$. We handle the very unlikely case $M = n-1$, whose probability is upper-bounded by $\Pr(M \geq i^*)$, separately and cover it by adding the random variable $\widetilde{Z}$ to our result. By a symmetrical argument to the above, for some index i^{**} such that $\Pr(X < i^{**}) = 1 - 1/\big((1 -$

$\varepsilon)(1+\beta)\big) + O(1/\sqrt{n})$), we obtain that $M \leq i^{**}$ also happens with probability at most $e^{-\varepsilon^2\cdot\mu/(2(1-\varepsilon))} \leq e^{-\varepsilon^2\cdot\mu/(3+3\varepsilon))}$, for $\varepsilon < 1 - 1/(1+\beta)$. This unlikely case is also included in $\widetilde{Z}$. From now on, we assume $i^{**} < M < n-1$. We note that by definition of M, we then have $C_{\geq M} \leq \mu$ and $C_{\geq M-1} \geq \mu+1$. Hence, $C_{M-1} \geq 1$ such that we have to investigate the distribution of C^{**} conditional on $C_{M-1} \geq 1 + (\mu - C_{\geq M})$.

We take the same view on the stochastic process as above but imagine now that the levels are sampled in the order from $n-1$ down to 0. Conditioning on that levels $n-1,\dots,M+1$ have been sampled, levels M, $M-1$ and $M-2$ are still hit with probability $O(1/\sqrt{n})$ each, since $\Pr(X < i^{**})$ is a constant. Therefore, the distribution of C_M is bounded according to

$$\Pr(C_M \geq k) \leq \Pr\big(\mathrm{Bin}(\lambda - C_{\geq M+1}, O(1/\sqrt{n})) \geq k\big)\,.$$

To analyze C_{M-1}, we recall that we have to condition on $C_{M-1} \geq 1 + (\mu - C_{\geq M})$. Hence, we can use Lemma 1 similarly as above and get

$$\Pr\big(C_{M-1} \geq 1 + (\mu - C_{\geq M}) + k \;\big|\; C_{M-1} \geq 1 + (\mu - C_{\geq M})\big) \leq \Pr\big(\mathrm{Bin}(\lambda - C_{\geq M}, O(1/\sqrt{n})) \geq k\big)\,.$$

Note that the right-hand side of the inequality is independent of C^*. Applying the argumentation once more for level $M-2$ (where no conditions on the size exist), we get $\Pr(C_{M-2} \geq k) \leq \Pr\big(\mathrm{Bin}(\lambda - C_{\geq M-1}, O(1/\sqrt{n})) \geq k\big)$. Using our stochastic bound on C^* from above, we altogether obtain that C^{**} is stochastically dominated by the sum of 1, three binomially distributed random variables with a total number of λ trials and success probability $O(1/\sqrt{n})$ each, and the variable $\widetilde{Z}$. □

Now that we understand how C^* is distributed, we can look at the distribution of both the 1st- and 2nd-class individuals. We even can take a finer-grained view on the number of 1s contributed by them.

Lemma 5. *Consider UMDA optimizing* OneMax. *Consider further that $\Theta(n)$ frequencies are within $[1/6, 5/6]$ and that we are in iteration t. Let j be any position, and let X_{t-1} denote the number of* 1s *at position j in iteration $t-1$.*

The distribution $Z_{1,t}$ of the number of 1s *of* 1*st-class individuals is stochastically dominated by $\mathrm{Bin}(\mu, X_{t-1}/\mu)$, and the distribution $Z_{2,t}$ of the number of* 1s *of* 2*nd-class individuals is stochastically dominated by C^*, where C^* is distributed as seen in Lemma 4. In particular, the expected value of $Z_{2,t}$ is at most $O(\mu/\sqrt{n}) + e^{-\Omega(\mu)}$.*

Further the expected value of $Z_{2,t}$, given X_{t-1}, is at most $O\big(X_{t-1}/\mu + X_{t-1}/\sqrt{n}\big) + e^{-\Omega(\mu)}$.

Proof. The distribution of $Z_{1,t}$ has already been described in Section 2.1 as $\mathrm{Bin}(\mu - C^*, X_{t-1}/\mu)$, which is dominated by $\mathrm{Bin}(\mu, X_{t-1}/\mu)$. We also know that the number of 2nd-class individuals is bounded from above by C^*, and their number of 1s is trivially bounded by this cardinality too. The first statement on the expected value of $Z_{2,t}$ follows by taking the expected value of the binomial distribution and noting that $E(\widetilde{Z}) \leq \lambda e^{-\Omega(\mu)} = e^{-\Omega(\mu)}$, using $\lambda = O(\mu)$.

To show the second statement on the expected value of $Z_{2,t}$, we recall our definition of 2nd-class candidates from above. These candidates have not been subject to selection yet. Each of these candidates samples a 1 at position j independently of the others with probability X_{t-1}/μ, so the expected total number of 1s in 2nd-class candidates is the expected number of candidates multiplied with X_{t-1}/μ, by Wald's identity. By Lemma 4, there is an expected number of at most $1 + O(\mu/\sqrt{n}) + e^{-\Omega(\mu)}$ of candidates, using again $\lambda = O(\mu)$. Since the 2nd-class individuals are only selected from the candidates, the claim on the expected value of $Z_{2,t}$ follows. □

3. Lower bound on OneMax

In the following, we derive a lower bound on UMDA's run time on OneMax. First, we state the main theorem.

Theorem 6. *Let $\lambda = (1+\beta)\mu$ for some constant $\beta > 0$. Then the expected optimization time of UMDA on* OneMax *is $\Omega(\mu\sqrt{n} + n\log n)$.*

To prove the theorem, we will distinguish between different cases for λ: small, medium, and large. We will cover the lemmas we use to prove the different cases in different sections. The first and the last case are fairly easy to prove, hence we discuss them first, leaving the second case of medium λ – the most difficult one – to be discussed last.

In each of the following sections, we will introduce the basic idea behind each of the proofs.

3.1. Small population sizes

In this section, we consider a population size of $\lambda \leq (1-c_1)\log_2 n$, for some constant $c_1 > 0$. If the population size is that small, the probability that a frequency reaches $1/n$ is rather high, and thus the probability to sample the optimum will be quite small.

If enough frequencies drop to $1/n$, we can bound the expected number of fitness evaluations until we sample the optimum by $\Omega(n\log n)$. The following lemma and its proof closely follow [21, Lemma 13].

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

Lemma 7. *Assume that $\Omega(n^{c_1})$ frequencies, $c_1 > 0$ being a constant, are at $1/n$. Then UMDA will need with high probability and in expectation still $\Omega(n \log n)$ function evaluations to optimize any function with a unique global optimum.*

Proof. Due to symmetry, we can w.l.o.g. assume that the global optimum is the all-ones string.

We look at $(c_2 n \ln n)/(2\lambda)$ iterations, where $c_2 < c_1$ is a positive constant, and show that it is very unlikely to sample the all-ones string during that time. Note that this translates to $\Omega(n \log n)$ function evaluations until the optimum is sampled, as UMDA samples λ offspring every iteration.

Consider a single position with frequency at $1/n$. The probability that this position never samples a 1 during our time of $(c_2 n \ln n)/(2\lambda)$ iterations is at least

$$\left(\left(1 - \frac{1}{n}\right)^{\lambda}\right)^{\frac{c_2 n \ln n}{2\lambda}} = \left(1 - \frac{1}{n}\right)^{\frac{c_2 n \ln n}{2}} \geq \big(1 - o(1)\big) e^{-\frac{c_2}{2} \ln n} \geq n^{-c_2}$$

if n is large enough.

Given $\Omega(n^{c_1})$ frequencies at $1/n$, the probability that all of these positions sample at least one 1 during $(c_2 n \ln n)/(2\lambda)$ iterations is at most

$$\big(1 - n^{-c_2}\big)^{\Omega(n^{c_1})} \leq e^{-\Omega(n^{c_1 - c_2})},$$

which is exponentially small in n, since $c_1 > c_2$, due to our assumptions.

Hence, with high probability, UMDA will need at least $\Omega(n \log n)$ function evaluations to find the optimum.

Since the expected value of function evaluations is finite (due to the bound of $1 - 1/n$ and $1/n$ for the frequencies) and it is $\Omega(n \log n)$ with high probability, it follows that the expected number of fitness evaluations is $\Omega(n \log n)$ as well. □

We can now prove our lower bound for small population sizes.

Theorem 8. *Let $\lambda \leq (1 - c_1) \log_2 n$ for some arbitrarily small constant $c_1 > 0$. Then UMDA will need with high probability and in expectation $\Omega(n \log n)$ function evaluations to optimize any function with a unique global optimum.*

Proof. Due to symmetry, we can w.l.o.g. assume that the global optimum is the all-ones string. We consider an arbitrary position i and study the first iteration of UMDA. The probability that all λ bits at position i are sampled as 0 equals $2^{-\lambda} \geq n^{-(1-c_1)}$. In this case, the frequency of the position is set to $1/n$. The expected number of such positions is n^{c_1}, and by Chernoff bounds, with high probability $\Omega(n^{c_1})$ such positions exist (noting that c_1 is a positive constant by assumption).

Applying Lemma 7 yields the result, since we already have $\Omega(n^{c_1})$ frequencies at $1/n$ after a single iteration of UMDA with high probability. □

3.2. Large population sizes

Here we are going to show that a population size of $\lambda = \Omega(\sqrt{n} \log n)$ leads to a run time of $\Omega(n \log n)$. To prove this, we first show that it is unlikely that too many frequencies leave the interval $[1/6, 5/6]$ quickly in this scenario. Thus, it is also unlikely to sample the optimum.

We start by proving that a single frequency does not leave $[1/6, 5/6]$ too quickly, for $\mu = \omega(1)$. We make use of Corollary 3 and the lemmas following from it, all of which make use of the lemmas we prove here themselves. At the end of this section, we will discuss why this seemingly contradictory approach is feasible.

Lemma 9. *Consider an arbitrary frequency of UMDA with $\lambda = \omega(1)$ optimizing* OneMax. *During the first at least $\gamma \cdot \min\{\mu, \sqrt{n}\}$ iterations, for a sufficiently small constant γ, this frequency will not leave $[1/6, 5/6]$ with a probability of at least a constant greater than 0.*

Proof. We consider the expected change of an arbitrary position's frequency p_t over time t. Let X_t, again, denote the number of 1s of the μ selected individuals. Note that $p_{t+1} = X_t/\mu$.

Due to Lemma 5, we know that X_t is the sum of two random variables $Z_{1,t}$ and $Z_{2,t}$, where $Z_{1,t} \prec \mathrm{Bin}(\mu, X_{t-1}/\mu)$ corresponds to the number of 1s due to the 1st-class individuals, and $Z_{2,t} \prec \mathrm{Bin}\big(\lambda, O(1/\sqrt{n})\big) + \widetilde{Z}_t$ corresponds to the 2nd-class individuals' number of 1s, pessimistically assuming that each 2nd-class individual contributes a 1.

First, we are going to upper-bound the probability of p_t reaching 5/6 during $\gamma \cdot \min\{\mu, \sqrt{n}\}$ iterations. Then, we do the same for reaching 1/6. Taking the converse probability of a union bound over both cases then yields the result.

The probability of reaching 5/6. Since $Z_{1,t}$ is dominated by a martingale which we want to account for in the process, we analyze $\phi_{t+1} := (X_t/\mu)^2$, with $\phi_0 = (1/2)^2$. Note that the square function is injective in this case because both X_t and μ are nonnegative. The original process of p_t reaching 5/6 translates into the new process p_t^2 reaching $(5/6)^2$.

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

We bound the expected change during one step:

$$\begin{aligned}\mathrm{E}(\phi_{t+1}-\phi_t \mid \phi_t) &= \frac{1}{\mu^2}\left(\mathrm{E}(X_t^2 \mid \phi_t) - X_{t-1}^2\right)\\ &= \frac{1}{\mu^2}\left(\mathrm{E}\big((Z_{1,t}+Z_{2,t})^2 \mid \phi_t\big) - X_{t-1}^2\right)\\ &= \frac{1}{\mu^2}\big(\mathrm{E}(Z_{1,t}^2 \mid \phi_t) + \mathrm{E}(Z_{2,t}^2 \mid \phi_t) + 2\mathrm{E}(Z_{1,t}\cdot Z_{2,t} \mid \phi_t) - X_{t-1}^2\big)\,.\end{aligned}$$

As discussed before, we will look at the dominating distributions of $Z_{1,t}$ and $Z_{2,t}$. Further, note that $Z_{1,t}$ and $Z_{2,t}$ are not independent, but their dominating distributions are.

We calculate the different terms separately:

$$\mathrm{E}(Z_{1,t}^2 \mid \phi_t) \le \mu\frac{X_{t-1}}{\mu}\left(1-\frac{X_{t-1}}{\mu}\right) + \left(\mu\frac{X_{t-1}}{\mu}\right)^2 \le X_{t-1} + X_{t-1}^2\,,$$

i.e., the second moment of a binomially distributed random variable, as seen by noting that $\mathrm{E}(Z_{1,t}^2 \mid \phi_t) = \mathrm{Var}(Z_{1,t} \mid \phi_t) + \mathrm{E}(Z_{1,t} \mid \phi_t)^2$.

For $Z_{2,t}$, let $Z_t^* \sim \mathrm{Bin}\big(\lambda, \mathrm{O}(1/\sqrt{n})\big)$, and recall that $\widetilde{Z}$ is a random variable that takes values in $\{1,\dots,\lambda\}$ with probability $\mathrm{e}^{-\Omega(\mu)}$ and is 0 otherwise. Using, again, the second moment of a binomially distributed random variable, we get

$$\begin{aligned}\mathrm{E}(Z_{2,t}^2 \mid \phi_t) &\le \mathrm{E}\big((Z_t^*)^2 \mid \phi_t\big) + \mathrm{E}\big((\widetilde{Z}_t)^2 \mid \phi_t\big) + 2\mathrm{E}(Z_t^* \mid \phi_t)\mathrm{E}(\widetilde{Z}_t \mid \phi_t)\\ &\le \mathrm{O}\left(\frac{\mu}{\sqrt{n}}\right) + \mathrm{O}\left(\frac{\mu^2}{n}\right) + \mu^2\mathrm{e}^{-\Omega(\mu)} + \mathrm{O}\left(\frac{\mu^2}{\sqrt{n}}\mathrm{e}^{-\Omega(\mu)}\right)\\ &\le \max\left\{\mathrm{O}\left(\frac{\mu}{\sqrt{n}}\right), \mathrm{O}\left(\frac{\mu^2}{n}\right), \mu^2\mathrm{e}^{-\Omega(\mu)}\right\}\,,\end{aligned}$$

because the term $\mathrm{O}\big(\mu^2/(\sqrt{n}\mathrm{e}^{\Omega(\mu)})\big)$ is always dominated by another term. Note that $\mathrm{O}(\mu/\sqrt{n})$ dominates if $\mu = \mathrm{o}(\sqrt{n})$ and if $\mu \ge c\ln n$ for a sufficiently large constant $c > 0$. For $\mu = \Omega(\sqrt{n})$, the term $\mathrm{O}(\mu^2/n)$ dominates. In the remaining cases (when μ is logarithmic), the term $\mu^2\mathrm{e}^{-\Omega(\mu)}$ dominates.

For the first moment of $Z_{2,t}$, we can get a similar bound:

$$\mathrm{E}(Z_{2,t} \mid \phi_t) \le \max\left\{\mathrm{O}\left(\frac{\mu}{\sqrt{n}}\right), \mu\mathrm{e}^{-\Omega(\mu)}\right\}\,,$$

where the term $\mu\mathrm{e}^{-\Omega(\mu)}$ only dominates if $\mu \le c\ln n$ for a sufficiently small constant $c > 0$.

Using our prior calculations and independence of the dominating distributions, we can bound

$$2\mathrm{E}\big(Z_{1,t}\cdot Z_{2,t} \mid \phi_t\big) \le X_{t-1}\cdot\max\left\{\mathrm{O}\left(\frac{\mu}{\sqrt{n}}\right), \mu\mathrm{e}^{-\Omega(\mu)}\right\}\,.$$

Thus, we get

$$\begin{aligned}\mathrm{E}(\phi_{t+1}-\phi_t \mid \phi_t) &\le \frac{1}{\mu^2}\Bigg(X_{t-1} + X_{t-1}^2 + \max\left\{\mathrm{O}\left(\frac{\mu}{\sqrt{n}}\right), \mathrm{O}\left(\frac{\mu^2}{n}\right), \mu^2\mathrm{e}^{-\Omega(\mu)}\right\}\\ &\qquad + X_{t-1}\cdot\max\left\{\mathrm{O}\left(\frac{\mu}{\sqrt{n}}\right), \mu\mathrm{e}^{-\Omega(\mu)}\right\} - X_{t-1}^2\Bigg)\\ &\le \frac{1}{\mu^2}\Bigg(\max\left\{\mathrm{O}\left(\frac{\mu}{\sqrt{n}}\right), \mathrm{O}\left(\frac{\mu^2}{n}\right), \mu^2\mathrm{e}^{-\Omega(\mu)}\right\}\\ &\qquad + X_{t-1}\left(1+\max\left\{\mathrm{O}\left(\frac{\mu}{\sqrt{n}}\right), \mu\mathrm{e}^{-\Omega(\mu)}\right\}\right)\Bigg)\\ &\overset{X_{t-1}\le\mu}{\le} \frac{1}{\mu^2}\mu\left(1+\max\left\{\mathrm{O}\left(\frac{\mu}{\sqrt{n}}\right), \mu\mathrm{e}^{-\Omega(\mu)}\right\}\right)\\ &\le \mathrm{O}\left(\max\left\{\frac{1}{\mu}, \frac{1}{\sqrt{n}}\right\}\right)\,.\end{aligned}$$

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

Let P_T describe the Markov process $p_t^2 = \phi_t$ starting at $(1/2)^2$ and then progressing by $\phi_{t+1} - \phi_t$ for T iterations. Due to our bounds, we get

$$\mathrm{E}(P_T) = \left(\frac{1}{2}\right)^2 + \sum_{t=0}^{T-1} \mathrm{E}(\phi_{t+1} - \phi_t \mid \phi_t) \leq \frac{1}{4} + \zeta T \cdot \max\left\{\frac{1}{\mu}, \frac{1}{\sqrt{n}}\right\},$$

for a sufficiently large constant ζ.

Using Markov's inequality gives us, for $k > 1$,

$$\Pr\left(P_T \geq k\left(\frac{1}{4} + \zeta T \cdot \max\left\{\frac{1}{\mu}, \frac{1}{\sqrt{n}}\right\}\right)\right) \leq \Pr\big(P_T \geq k\mathrm{E}(P_T)\big) \leq \frac{1}{k}.$$

We want that $(5/6)^2 \geq k(1/4 + \zeta T \cdot \max\{1/\mu, 1/\sqrt{n}\})$, since then $\Pr\big(P_T \geq (5/6)^2\big)$ is upper-bounded by $\Pr\big(P_T \geq k(1/4 + \zeta T \cdot \max\{1/\mu, 1/\sqrt{n}\})\big) \leq 1/k$, which we want to be less than $1/2$ in order to apply a meaningful union bound over both cases at the end of this proof. Hence, assume $k > 2$. We get

$$\left(\frac{5}{6}\right)^2 \geq k\left(\frac{1}{4} + \zeta T \cdot \max\left\{\frac{1}{\mu}, \frac{1}{\sqrt{n}}\right\}\right) \Leftrightarrow T \leq \left(\frac{25}{36k} - \frac{1}{4}\right)\frac{\min\{\mu, \sqrt{n}\}}{\zeta},$$

which is positive as long as $k < 25/9$. Thus, we can bound $k \in (2, 25/9)$.

Therefore, if $T \leq \gamma \cdot \min\{\mu, \sqrt{n}\}$, for a constant γ sufficiently small, then the probability of an arbitrary frequency exceeding $5/6$ is at most a constant less than $1/2$ (for $k \in (2, 25/9)$).

The probability of reaching 1/6. We now analyze how likely it is that p_t hits $1/6$ in a similar amount of time. For this case, we define a slightly different potential $\phi'_{t+1} := (1 - X_t/\mu)^2 = 1 - 2X_t/\mu + (X_t/\mu)^2$, i.e., we mirror the process at $1/2$ and then use the same potential as before.

Looking at the difference during one step, we see that

$$\begin{aligned}\phi'_{t+1} - \phi'_t &= 1 - 2\frac{X_t}{\mu} + \left(\frac{X_t}{\mu}\right)^2 - 1 + 2\frac{X_{t-1}}{\mu} - \left(\frac{X_{t-1}}{\mu}\right)^2 \\ &= \frac{2}{\mu}(X_{t-1} - X_t) + \phi_{t+1} - \phi_t,\end{aligned}$$

where we only have to determine the expected value of $X_{t-1} - X_t$, because we already analyzed $\phi_{t+1} - \phi_t$ before.

Considering just the 1st-class individuals, it holds that $\mathrm{E}(X_t) = \mathrm{E}(X_{t-1})$, because we then have a martingale. But due to the elitist selection of UMDA, actually $\mathrm{E}(X_t) \geq \mathrm{E}(X_{t-1})$ holds, because of the bias of the 2nd-class individuals, which prefer 1s over 0s. Thus, $\mathrm{E}(X_{t-1} - X_t \mid \phi'_t) \leq 0$, and we get

$$\mathrm{E}(\phi'_{t+1} - \phi'_t \mid \phi'_t) \leq \mathrm{E}(\phi_{t+1} - \phi_t \mid \phi_t),$$

which we already analyzed.

Hence, we can argue analogously as before and get, again, a probability of at most a constant less than $1/2$ to reach $1/6$ during at most $\gamma \cdot \min\{\mu, \sqrt{n}\}$ iterations.

Taking a union bound over both cases finishes the proof. □

We now expand the case from a single frequency to all frequencies.

Lemma 10. *During the first at least $\gamma \cdot \min\{\mu, \sqrt{n}\}$ iterations of UMDA optimizing* OneMax, *for a sufficiently small constant γ, $\Theta(n)$ frequencies stay in the interval $[1/6, 5/6]$ with at least constant probability.*

Proof. We look at $T \leq \gamma \cdot \min\{\mu, \sqrt{n}\}$ iterations. Thus, the probability for a single frequency to leave $[1/6, 5/6]$ is at most a constant $c < 1$, according to Lemma 9. In expectation, there are at most cn frequencies outside of $[1/6, 5/6]$, and due to Markov's inequality, the probability that there are at least $(1+\delta)cn$ such frequencies, for a constant $\delta > 0$ with $(1+\delta)c < 1$, is at most $1/(1+\delta)$. This means that with at least constant probability, at least $\big(1 - c(1+\delta)\big)n = \Theta(n)$ frequencies are still within $[1/6, 5/6]$. □

Note that the proof of Lemma 9 relies on Corollary 3, and the proof of Corollary 3 also relies on Lemma 9. Formally, this cyclic dependency can be solved by proving both propositions in conjunction via induction over the number of iterations up to $\gamma \cdot \min\{\mu, \sqrt{n}\}$, for a sufficiently small constant γ. For the base case, all frequencies are at $1/2 \in [1/6, 5/6]$, and both propositions hold. For the inductive step, assuming that $t < \gamma \cdot \min\{\mu, \sqrt{n}\}$, we already know that both propositions hold up to iteration t. Thus, the requirements for the proofs of Corollary 3 and Lemma 9 are fulfilled, and the proofs themselves pass.

We now prove an easy lower bound.

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

Corollary 11. *Consider UMDA optimizing* OneMax *with* $\mu = \Omega(\sqrt{n}\log n)$. *Its run time is then in* $\Omega(n\log n)$ *in expectation and with at least constant probability.*

Proof. Since we assume $\mu = \Omega(\sqrt{n}\log n)$, Lemma 10 yields that within at most $\gamma \cdot \min\{\mu, \sqrt{n}\} = \gamma\sqrt{n}$ iterations, γ sufficiently small, at least $\Theta(n)$ frequencies are at most 5/6 with probability $\Omega(1)$. Hence, assuming this to happen, the probability to sample the optimum is at most $(5/6)^{\Theta(n)} \leq e^{-\Theta(n)}$, and, thus, the expected run time is in $\gamma\sqrt{n}\lambda = \Omega(n\log n)$. □

3.3. Medium population sizes

In this section, we consider the remaining population sizes of $\mu = O(\sqrt{n}\log n)$ (and $\mu = \Omega(\log n)$), where we recall that $\lambda = (1+\beta)\mu$. Basically, we lower-bound the probability that a single frequency hits $1/n$. To do so, we analyze the one-step change of the number of 1s at the frequency's position and approximate it via a normal distribution. For this, we are going to use a general form of the central limit theorem (CLT), along with a bound on the approximation error.

Lemma 12 *(CLT with Lyapunov condition, Berry–Esseen inequality [22, p. 544]). Let* $X_1, \dots, X_m$ *be a sequence of independent random variables, each with finite expected value* μ_i *and variance* σ_i^2. *Define*

$$s_m^2 := \sum_{i=1}^{m} \sigma_i^2 \quad \text{and} \quad C_m := \frac{1}{s_m} \sum_{i=1}^{m} (X_i - \mu_i) \,.$$

If there exists a $\delta > 0$ *such that*

$$\lim_{m\to\infty} \frac{1}{s_m^{2+\delta}} \sum_{i=1}^{m} \mathrm{E}\left(|X_i - \mu_i|^{2+\delta}\right) = 0$$

(assuming all the moments of order $2+\delta$ *to be defined), then* C_m *converges in distribution to a standard normally distributed random variable.*

Moreover, the approximation error is bounded as follows: for all $x \in \mathbb{R}$,

$$|\Pr(C_m \leq x) - \Phi(x)| \leq C \cdot \frac{\sum_{i=1}^{m} \mathrm{E}\left(|X_i - \mu_i|^3\right)}{s_m^3} \,,$$

where C *is an absolute constant and* $\Phi(x)$ *denotes the cumulative distribution function of the standard normal distribution.*

In order to make use of Lemma 12, we need to study the stochastic process on the X_t values (which, again, denotes the number of 1s of an arbitrary position) and determine the accumulated expectations and variances of every single one-step change. Using the notation from Lemma 5, we note that the X_t value in expectation changes very little from one step to the next since $\mathrm{E}(Z_{1,t}) = X_{t-1}$ and also $\mathrm{E}(Z_{2,t})$ is close to X_{t-1}. However, considerable variances are responsible for changes of the X_t value, and it turns out that the variances are heavily dependent on the current state. We get $\mathrm{Var}(Z_{1,t}) = X_{t-1}(1 - X_{t-1}/\mu)$, i.e., if $X_{t-1} \leq \mu/2$, then the 1st-class individuals are responsible for a typical deviation of $\sqrt{X_{t-1}}$. This dependency of $\mathrm{Var}(Z_{1,t})$ on X_{t-1} makes a direct application of Lemma 12 difficult.

In order to make the CLT applicable, we define a potential function that transforms X_{t-1} such that the expected difference between two points in time is still close to 0, but the variance is independent of the state. This potential function is inspired by the approach used in [9] in order to analyze two very simple EDAs. Since the standard deviation of $Z_{1,t}$ is $\Theta(\sqrt{X_{t-1}})$, we work with a potential function whose slope at point X_{t-1} is $\Theta(1/\sqrt{X_{t-1}})$, so that the dependency of the variance on the state cancels out.

We proceed with the formal definition. Let g denote the potential function, defined over $\{0, \dots, \mu\}$. Our definition is simpler than the one from Sudholt and Witt [9], as we do not need g to be centrally symmetric around $\mu/2$. We define

$$g(x) := \sqrt{\mu} \sum_{j=x}^{\mu-1} \frac{1}{\sqrt{j+1}} \,.$$

We will often use the following bounds on the change of potential. For $0 \leq y < x \leq \mu$, we get

$$g(y) - g(x) = \sqrt{\mu} \sum_{j=y}^{x-1} \frac{1}{\sqrt{j+1}} \leq \sqrt{\mu} \frac{x-y}{\sqrt{y+1}} \,, \text{ and} \tag{2}$$

$$g(y) - g(x) = \sqrt{\mu} \sum_{j=y}^{x-1} \frac{1}{\sqrt{j+1}} \geq \sqrt{\mu} \frac{x-y}{\sqrt{x+1}} \,. \tag{3}$$

Let $\Delta_t = g(X_{t+1}) - g(X_t)$.

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

3.3.1. Bounding the expected change of potential

We start by bounding the expected value of Δ_t and see that also the transformed process moves very little in expectation (however, its variance will be large, as shown in the following subsection). Because of the Lyapunov condition, which we will address in Section 3.3.3, we do so in both directions.

Lemma 13. *Let $\mu = O(\sqrt{n}\log n)$. Then, for all t and all $X_t \in \{1, \dots, \mu - 1\}$,*

$$E(\Delta_t \mid X_t) \geq -\left(e^{-\Omega(\mu)} + O\left(\frac{X_t}{\mu} + \frac{X_t}{\sqrt{n}}\right)\right)\sqrt{\frac{\mu}{X_t + 1}} \text{ and}$$

$$E(\Delta_t \mid X_t) \leq 111\sqrt{\frac{\mu}{X_t}}.$$

Proof. We abbreviate $X_t = x$. Further, we always condition on x without denoting this explicitly.

The lower bound. First, we derive the lower bound. We have $E(\Delta_t) = E\big(g(X_{t+1})\big) - g(x)$. Because g is convex we get by Jensen's inequality that $E\big(g(X_{t+1})\big) - g(x) \geq g\big(E(X_{t+1})\big) - g(x) \geq g\big(x + e^{-\Omega(\mu)} + O(x/\mu + x/\sqrt{n})\big) - g(x)$, where we used that

$$E(X_{t+1}) \leq x + e^{-\Omega(\mu)} + O\left(\frac{x}{\mu} + \frac{x}{\sqrt{n}}\right),$$

which follows from Lemma 5 by studying the expected number of 1s contributed by the two classes of individuals.

Applying (2), gives us the desired result of

$$g\left(x + e^{-\Omega(\mu)} + O\left(\frac{x}{\mu} + \frac{x}{\sqrt{n}}\right)\right) - g(x) \geq -\left(e^{-\Omega(\mu)} + O\left(\frac{x}{\mu} + \frac{x}{\sqrt{n}}\right)\right)\sqrt{\frac{\mu}{x+1}}.$$

The upper bound. The upper bound will be shown by ignoring 2nd-class individuals, since they are biased toward increasing x and, therefore, decreasing Δ_t. Hence, we now assume that X_{t+1} follows a binomial distribution with parameters μ and x/μ, i.e., $E(X_{t+1} - x) = 0$. In a delicate analysis, we will estimate how much $E(\Delta_t)$ is shifted away from 0 due to the nonlinearity of the potential function. We use the inequalities

$$g(i) \leq g(x) + \frac{\sqrt{\mu}(x-i)}{\sqrt{i+1}} \quad \text{for } i < x, \text{ and}$$

$$g(i) \leq g(x) + \frac{\sqrt{\mu}(x-i)}{\sqrt{i+1}} \quad \text{for } i > x,$$

which are just rearrangements of (2) and (3), noting that $x - i$ is negative in the second inequality.

$$\begin{aligned}
E(\Delta_t) &= \sum_{i=0}^{\mu} \big(g(i) - g(x)\big) \Pr(X_{t+1} = i) \\
&\leq \sum_{i=0}^{x-1} \left(g(x) + \frac{\sqrt{\mu}(x-i)}{\sqrt{i+1}} - g(x)\right) \Pr(X_{t+1} = i) + \sum_{i=x+1}^{\mu} \left(g(x) + \frac{\sqrt{\mu}(x-i)}{\sqrt{i+1}} - g(x)\right) \Pr(X_{t+1} = i) \\
&= \sum_{i=0}^{\infty} \left(\frac{\sqrt{\mu}(x-i)}{\sqrt{i+1}} \Pr(X_{t+1} = i)\right).
\end{aligned}$$

We now split the set of possible outcomes of i into intervals of length $\sqrt{x}$. More precisely $I_k := \{\lceil x - (k+1)\sqrt{x}\rceil, \dots, \lfloor x - k\sqrt{x}\rfloor\}$ for $k \in \mathbb{N}_0$. The points in these intervals are all less than or equal to x. To cover the outcomes above x when considering some $i \in I_k$, we consider the points i and $2x - i$ together, exploiting that they are mirrors of each other of distance $x - i$ to x, more formally $x - i = -(x - (2x - i))$. Plugging in i and $2x - i$ for $i \in I_k$ and summing over all $k \geq 0$, we obtain

$$\begin{aligned}
E(\Delta_t) &\leq \sum_{k=0}^{\infty} \sum_{i \in I_k} \left(\frac{\sqrt{\mu}(x-i)}{\sqrt{i+1}} \Pr(X_{t+1} = i) + \frac{\sqrt{\mu}(i-x)}{\sqrt{2x-i+1}} \Pr(X_{t+1} = 2x-i)\right) \\
&\leq \sum_{k=0}^{\infty} \sum_{i \in I_k} \left(\frac{\sqrt{\mu}(x-i)}{\sqrt{x-(k+1)\sqrt{x}+1}} \Pr(X_{t+1} = i) - \frac{\sqrt{\mu}(x-i)}{\sqrt{x+(k+1)\sqrt{x}+1}} \Pr(X_{t+1} = 2x-i)\right),
\end{aligned}$$

where the last inequality used that the choice $i = x - (k+1)\sqrt{x}$ maximizes both the positive and the negative term in the inner sum.

We take special care of intervals where $x - (k+1)\sqrt{x} \leq x/2$ (i.e., $k \geq \sqrt{x}/2 - 1$) and handle them directly. The maximum increase in potential is observed when $X_{t+1} = 0$ and equals

$$\sqrt{\mu}\sum_{j=0}^{x-1}\frac{1}{\sqrt{j+1}} \leq \sqrt{\mu}\left(1 + \int_1^x \frac{1}{\sqrt{z}}\,\mathrm{d}z\right) = \sqrt{\mu}(1 + 2\sqrt{x} - 2\sqrt{1}) \leq \sqrt{4\mu x}\,.$$

By Chernoff bounds, the probability of $X_{t+1} \leq x/2$ is at most $\mathrm{e}^{-x/24}$. Hence, the intervals of index at least $k_{\max} := \sqrt{x}/2 - 1$ contribute only a term of $S^* := \sqrt{4\mu x}\mathrm{e}^{-x/24} \leq 100\sqrt{\mu/x}$ to $\mathrm{E}(\Delta_t)$.[2]

For smaller k, we argue more precisely. Since

$$\frac{\sqrt{x+(k+1)\sqrt{x}+1}}{\sqrt{x-(k+1)\sqrt{x}+1}} = 1 + \frac{\sqrt{x+(k+1)\sqrt{x}+1} - \sqrt{x-(k+1)\sqrt{x}+1}}{\sqrt{x-(k+1)\sqrt{x}+1}} \leq 1 + \frac{\frac{2(k+1)\sqrt{x}}{2\sqrt{x-(k+1)\sqrt{x}+1}}}{\sqrt{x-(k+1)\sqrt{x}+1}} = 1 + \frac{(k+1)\sqrt{x}}{x-(k+1)\sqrt{x}+1}$$

(where the last inequality follows from $a - b \leq (a^2 - b^2)/2b$ for $a \geq b > 0$), we have

$$\mathrm{E}(\Delta_t) \leq \sum_{k=0}^{k_{\max}}\sum_{i\in I_k}\left(\frac{\sqrt{\mu}(x-i)}{\sqrt{x+(k+1)\sqrt{x}+1}}\left(1 + \frac{(k+1)\sqrt{x}}{x-(k+1)\sqrt{x}+1}\right)\Pr(X_{t+1}=i) - \frac{\sqrt{\mu}(x-i)}{\sqrt{x+(k+1)\sqrt{x}+1}}\Pr(X_{t+1}=2x-i)\right) + S^*\,. \tag{4}$$

We now look more closely into the inner sum and work with the abbreviation

$$E_k^* := \sum_{i\in I_k}\Big((x-i)\cdot\Pr(X_{t+1}=i) - (x-i)\Pr(X_{t+1}=2x-i)\Big)\,.$$

Coming back to (4), this enables us to estimate the inner sum for arbitrary k:

$$\begin{aligned}
&\sum_{i\in I_k}\left(\frac{\sqrt{\mu}(x-i)}{\sqrt{x+(k+1)\sqrt{x}+1}}\left(1 + \frac{(k+1)\sqrt{x}}{x-(k+1)\sqrt{x}+1}\right)\Pr(X_{t+1}=i) - \frac{\sqrt{\mu}(x-i)}{\sqrt{x+(k+1)\sqrt{x}+1}}\Pr(X_{t+1}=2x-i)\right)\\
&= E_k^*\cdot\frac{\sqrt{\mu}}{\sqrt{x+(k+1)\sqrt{x}+1}} + \sum_{i\in I_k}\frac{\sqrt{\mu}(x-i)}{\sqrt{x+(k+1)\sqrt{x}+1}}\frac{(k+1)\sqrt{x}}{x-(k+1)\sqrt{x}+1}\Pr(X_{t+1}=i)\\
&\leq \frac{E_k^*\sqrt{\mu}}{\sqrt{x+(k+1)\sqrt{x}+1}} + \sum_{i\in I_k}\frac{\sqrt{\mu}(x-i)(k+1)}{x-(k+1)\sqrt{x}+1}\Pr(X_{t+1}=i)\,,
\end{aligned}$$

where the last inequality estimated $\sqrt{x}/\sqrt{x+(k+1)\sqrt{x}+1} \leq 1$. Since $k \leq k_{\max}$, i.e., $(k+1)\sqrt{x} \leq \sqrt{x}/2$, the last bound is easily bounded from above by

$$\frac{E_k^*\sqrt{\mu}}{\sqrt{x+(k+1)\sqrt{x}+1}} + \sum_{i\in I_k}\frac{\sqrt{\mu}(x-i)(k+1)}{\frac{x}{2}}\Pr(X_{t+1}=i)\,.$$

We proceed by bounding the sum over I_k, noting that we have $\Pr(X_{t+1}\in I_k) \leq \Pr(X_{t+1} \leq x - k\sqrt{x}) \leq \mathrm{e}^{-k^2/3}$ by Chernoff bounds. Hence, since $x - i \leq (k+1)\sqrt{x}$ for $i \in I_k$, we get

$$\sum_{i\in I_k}\frac{\sqrt{\mu}(x-i)(k+1)}{\frac{x}{2}} \leq \frac{2\sqrt{\mu}}{x}\sum_{i\in I_k}(k+1)\sqrt{x}\Pr(X_{t+1}=i)$$

[2] The inequality $2x\mathrm{e}^{-x/24} \leq 100/\sqrt{x}$ for $x \geq 1$ can be checked using elementary calculus.

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

$$\leq \frac{2\sqrt{\mu}(k+1)}{\sqrt{x}} \sum_{i \in I_k} \Pr(X_{t+1} = i)$$

$$\leq \frac{2\sqrt{\mu}(k+1)\mathrm{e}^{-\frac{k^2}{3}}}{\sqrt{x}} .$$

Altogether, we have obtained from (4) the simpler inequality

$$\mathrm{E}(\Delta_t) \leq \sum_{k=0}^{k_{\max}} \left(\frac{E_k^* \sqrt{\mu}}{\sqrt{x + (k+1)\sqrt{x} + 1}} + \frac{2\sqrt{\mu}(k+1)\mathrm{e}^{-\frac{k^2}{3}}}{\sqrt{x}} \right) + S^*, \tag{5}$$

which we will bound further. The idea is to exploit that

$$\sum_{k \geq 0} E_k^* = 0 , \tag{6}$$

which is a consequence of $\mathrm{E}(X_{t+1}) = x$ since

$$\begin{aligned} 0 &= \mathrm{E}(X_{t+1}) - x \\ &= \sum_{k \geq 0} \sum_{i \in I_k} \Big((i - x) \cdot \Pr(X_{t+1} = i) + ((2x - i) - x) \Pr(X_{t+1} = 2x - i) \Big) \\ &= \sum_{k \geq 0} E_k^* . \end{aligned}$$

Using similar calculations as above, we manipulate the sum

$$\sum_{k \geq 0} \frac{E_k^* \sqrt{\mu}}{\sqrt{x + (k+1)\sqrt{x} + 1}} ,$$

stemming from the upper bound (5), and recognize that it equals

$$\begin{aligned} &\sum_{k \geq 0} \frac{E_k^* \sqrt{\mu}}{\sqrt{x + \sqrt{x} + 1}} \cdot \left(1 + \frac{\sqrt{x + \sqrt{x} + 1} - \sqrt{x + (k+1)\sqrt{x} + 1}}{\sqrt{x + (k+1)\sqrt{x} + 1}} \right) \\ &\leq \sum_{\substack{k \geq 0 \\ E_k^* < 0}} \frac{E_k^* \sqrt{\mu}}{\sqrt{x + \sqrt{x} + 1}} \left(1 - \frac{k\sqrt{x}}{2\sqrt{x + (k+1)\sqrt{x} + 1}\sqrt{x + \sqrt{x} + 1}} \right) + \sum_{\substack{k \geq 0 \\ E_k^* \geq 0}} \frac{E_k^* \sqrt{\mu}}{\sqrt{x + \sqrt{x} + 1}} \cdot 1 , \end{aligned}$$

where we again used $a - b \leq (a^2 - b^2)/2b$ for $a \geq b > 0$.

Similarly as above, we get, using Chernoff bounds,

$$E_k^* \geq \sum_{i = x + k\sqrt{x}}^{x + (k+1)\sqrt{x}} (x - i) \Pr(X_{t+1} = i) \geq -2(k+1)\mathrm{e}^{-\frac{k^2}{3}} \sqrt{x} .$$

Combining this with (6), we arrive at the inequality

$$\begin{aligned} &\sum_{k \geq 0} \frac{E_k^* \sqrt{\mu}}{\sqrt{x + (k+1)\sqrt{x} + 1}} \\ &\leq \sum_{\substack{k \geq 0 \\ E_k^* < 0}} \frac{E_k^* \sqrt{\mu}}{\sqrt{x + \sqrt{x} + 1}} \left(- \frac{k\sqrt{x}}{2\sqrt{x + (k+1)\sqrt{x} + 1}\sqrt{x + \sqrt{x} + 1}} \right) \\ &\leq \sum_{k \geq 0} \frac{2(k+1)\mathrm{e}^{-\frac{k^2}{3}} \sqrt{x}\sqrt{\mu}}{\sqrt{x + \sqrt{x} + 1}} \cdot \frac{k\sqrt{x}}{2\sqrt{x + (k+1)\sqrt{x} + 1}\sqrt{x + \sqrt{x} + 1}} , \end{aligned}$$

which is at most $\sum_{k \geq 0} \big(k(k+1)\mathrm{e}^{-k^2/3} \sqrt{\mu} \big) / \sqrt{x}$.

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

Substituting this into (5), we finally obtain

$$\begin{aligned}\mathrm{E}(\Delta_t) &\le \sum_{k\ge 0}\left(\frac{2(k+1)\mathrm{e}^{-\frac{k^2}{3}}\sqrt{\mu}}{\sqrt{x}} + \frac{k(k+1)\mathrm{e}^{-\frac{k^2}{3}}\sqrt{\mu}}{\sqrt{x}}\right) + S^* \\ &\le 11\frac{\sqrt{\mu}}{\sqrt{x}} + \frac{100\sqrt{\mu}}{\sqrt{x}} = 111\frac{\sqrt{\mu}}{\sqrt{x}}\,,\end{aligned}$$

where the bound $\sum_{k=0}^{\infty}(2(k+1)+k(k+1))e^{-k^2/3} \le 11$ was obtained numerically. This finally proves the upper bound on $\mathrm{E}(\Delta_t)$. □

3.3.2. Lower bound on the variance of the potential change

Before we analyze the variance of Δ_t, we introduce a lemma that we are going to use.

Lemma 14 *([23, Lemma 6]). Let $X \sim \mathrm{Bin}(\mu, r/\mu)$ with $r \in [1, \mu]$, let $\ell = \min\{r, \mu - r\}$, and let $\zeta > 0$ be an arbitrary constant. Then $\Pr\big(X \ge \mathrm{E}(X) + \zeta\sqrt{\ell}\big) = \Omega(1)$. Note that if $r \le \mu/2$, we get $\Pr\big(X \ge \mathrm{E}(X) + \zeta\sqrt{\mathrm{E}(X)}\big) = \Omega(1)$.*

In [23], the lemma is only stated for $\zeta = 1$. However, introducing the constant factor does not change the lemmas's proof at all.

With Lemma 14 in place, we now lower-bound the variance of Δ_t. Note that the following lemma only applies up to $X_t \le (5/6)\mu$, which will be guaranteed in its application.

Lemma 15. *Let $\mu = \omega(1)$ and $\mu = \mathrm{O}(\sqrt{n}\log n)$. Then, for all t and $X_t \in \{1, \dots, (5/6)\mu\}$,*

$$\mathrm{Var}(\Delta_t \mid X_t) = \Omega(\mu)\,.$$

Proof. Again, we abbreviate $X_t = x$ and always condition on x without denoting so. Let $E^* := -\big(1 + \gamma(x/\sqrt{n}+1)\big)\cdot\sqrt{\mu/(x+1)}$ be a lower bound on $\mathrm{E}(\Delta_t)$ from Lemma 13, where we pessimistically estimated $\mathrm{e}^{-\Omega(\mu)} \le 1$, $x/\mu \le 1$ because $x \le \mu$, and where γ is a sufficiently large constant that captures the implicit constant in the O-notation. We estimate

$$\begin{aligned}\mathrm{Var}(\Delta_t) &= \mathrm{E}\Big(\big(\Delta_t - \mathrm{E}(\Delta_t)\big)^2\Big) \\ &\ge \mathrm{E}\Big(\big(\Delta_t - \mathrm{E}(\Delta_t)\big)^2 \cdot \mathbb{1}\{\Delta_t < E^*\}\Big) \\ &\ge \mathrm{E}\big((\Delta_t - E^*)^2 \cdot \mathbb{1}\{\Delta_t < E^*\}\big)\,.\end{aligned}$$

Note that we can ignore 2nd-class individuals, as they would only increase X_{t+1} even further, leading to a greater difference of Δ_t and E^*.

We derive a sufficient condition for $\Delta_t < E^*$. For this, we introduce the constant ζ and claim that $g(x+\zeta\sqrt{x}) \le g(x) + E^*$ if ζ is sufficiently large. This claim is equivalent to $g(x) - g(x+\zeta\sqrt{x}) \ge -E^*$.

We lower-bound the left-hand side as follows, assuming that ζ is sufficiently large and using Inequality (3):

$$\begin{aligned}g(x) - g(x+\zeta\sqrt{x}) &\ge \sqrt{\mu}\cdot\frac{\zeta\sqrt{x}}{\sqrt{x+\zeta\sqrt{x}+1}} \\ &\ge \sqrt{\mu}\cdot\frac{\zeta\sqrt{x}}{\sqrt{2\zeta x}} \\ &= \sqrt{\frac{\mu\zeta}{2}}\,,\end{aligned}$$

and we want this to be at least $-E^*$.

The inequality $\sqrt{\mu\zeta/2} \ge -E^*$ is equivalent to

$$\sqrt{\frac{\zeta}{2}}\cdot\sqrt{x+1} - 1 \ge \gamma\left(\frac{x}{\sqrt{n}} + 1\right)\,.$$

We prove this inequality by lower-bounding the left-hand side as follows if ζ is sufficiently large:

$$\sqrt{\frac{\zeta}{2}}\cdot\sqrt{x+1} - 1 \ge \frac{\sqrt{\zeta x}}{2}\,.$$

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004

It is now evident that $\sqrt{\zeta x}/2 \geq \gamma(x/\sqrt{n}+1) \Leftrightarrow \sqrt{\zeta}/2 \geq \gamma(\sqrt{x/n}+1/\sqrt{x})$ holds (for $x \neq 0$) if ζ is sufficiently large, i.e., if $\zeta \geq (4\gamma)^2$, because $x \leq \mu$ and we assume $\mu = \mathrm{O}(\sqrt{n}\log n)$, thus, $\sqrt{x/n}+1/\sqrt{x} \leq 1+\mathrm{o}(1)$. For $x=0$, the inequality trivially holds.

Using the inequality derived above, we get:

$$\begin{aligned}\Delta_t < E^* &\Leftrightarrow g(X_{t+1}) - g(x) < E^* \Leftrightarrow g(X_{t+1}) < g(x) + E^* \\ &\Leftarrow g(X_{t+1}) < g(x+\zeta\sqrt{x}) \Leftrightarrow X_{t+1} > x + \zeta\sqrt{x}\,,\end{aligned}$$

where we used the definition of g and that it is a decreasing function.

We proceed by estimating the expected value. First, we see that, assuming $X_{t+1} > x+\zeta\sqrt{x}$,

$$\begin{aligned}\Delta_t - E^* &= g(X_{t+1}) - \big(g(x) + E^*\big) \\ &\leq g(X_{t+1}) - g(x+\zeta\sqrt{x}) \\ &= -\sqrt{\mu} \sum_{j=x+\zeta\sqrt{x}}^{X_{t+1}-1} \frac{1}{\sqrt{j+1}}\,,\end{aligned}$$

by using the same bounds as before. Note that we derive an upper bound of $\Delta_t - E^*$, because we only consider $\Delta_t < E^*$, i.e., $\Delta_t - E^* < 0$. Thus, its square gets minimized for an upper bound.

Since $X_{t+1} > x+\zeta\sqrt{x}$ implies $\Delta_t < E^*$, we get

$$\begin{aligned}\mathrm{E}\left(\big(\Delta_t - E^*\big)^2 \cdot \mathbb{1}\{\Delta_t < E^*\}\right) &\geq \mathrm{E}\left(\big(\Delta_t - E^*\big)^2 \cdot \mathbb{1}\{X_{t+1} > x+\zeta\sqrt{x}\}\right) \\ &\geq \mathrm{E}\Big(\big(g(X_{t+1}) - g(x+\zeta\sqrt{x})\big) \cdot \mathbb{1}\{X_{t+1} > x+\zeta\sqrt{x}\}\Big)^2 \\ &= \left(\sum_{i=0}^{\mu}(-\sqrt{\mu}) \sum_{j=x+\zeta\sqrt{x}}^{i-1} \frac{1}{\sqrt{j+1}} \cdot \mathbb{1}\{i > x+\zeta\sqrt{x}\}\Pr(X_{t+1}=i)\right)^2 \\ &= \mu\left(\sum_{i=x+\zeta\sqrt{x}+1}^{\mu} \sum_{j=x+\zeta\sqrt{x}}^{i-1} \frac{1}{\sqrt{j+1}} \Pr(X_{t+1}=i)\right)^2,\end{aligned}$$

where the second inequality is due to Jensen's inequality.

We now derive a lower bound for the inner sum. Using Inequality (3), we get

$$\sum_{j=x+\zeta\sqrt{x}}^{i-1} \frac{1}{\sqrt{j+1}} \geq \frac{i-x-\zeta\sqrt{x}}{\sqrt{i}}\,.$$

Substituting this back into the expectation gives us

$$\begin{aligned}\mu\left(\sum_{i=x+\zeta\sqrt{x}+1}^{\mu} \sum_{j=x+\zeta\sqrt{x}}^{i-1} \frac{1}{\sqrt{j+1}} \Pr(X_{t+1}=i)\right)^2 &\geq \mu\left(\sum_{i=x+\zeta\sqrt{x}+1}^{\mu} \frac{i-x-\zeta\sqrt{x}}{\sqrt{i}} \Pr(X_{t+1}=i)\right)^2 \\ &\geq \mu\left(\sum_{i=x+2\zeta\sqrt{x}+1}^{\mu} \frac{i-x-\zeta\sqrt{x}}{\sqrt{i}} \Pr(X_{t+1}=i)\right)^2,\end{aligned}$$

where we narrowed the range for i. In this new range, $(i-x-\zeta\sqrt{x})/\sqrt{i}$ is monotonically increasing with respect to i and hence minimal for $i = x+2\zeta\sqrt{x}+1$:

$$\begin{aligned}\frac{x+2\zeta\sqrt{x}+1-x-\zeta\sqrt{x}}{\sqrt{x+2\zeta\sqrt{x}+1}} &= \frac{\zeta\sqrt{x}+1}{\sqrt{x+2\zeta\sqrt{x}+1}} \\ &\geq \frac{\zeta\sqrt{x}+1}{\sqrt{3\zeta x}} \\ &= \sqrt{\frac{\zeta}{3}} + \frac{1}{\sqrt{3\zeta x}} \\ &= \Omega(1)\,.\end{aligned}$$

Hence, we finally have

$$\begin{aligned}\mathrm{Var}(\Delta) &\geq \Omega(\mu)\left(\sum_{i=x+2\zeta\sqrt{x}+1}^{\mu}\Pr(X_{t+1}=i)\right)^2\\ &\geq \Omega(\mu)\Pr(X_{t+1}\geq x+2\zeta\sqrt{x}+1)^2 \geq \Omega(\mu)\,.\end{aligned}$$

The last inequality used Lemma 14 to lower-bound the probability. The lemma can be used immediately for $x \leq \mu/2$. Otherwise, we still have $x \leq (5/6)\mu$ by assumption. Then Lemma 14 gives us a bound on $\Pr(X_{t+1} \geq x + \zeta\sqrt{\mu - x})$, which only changes everything by a constant factor, since it holds that $\sqrt{x}/\sqrt{\mu - x} \leq \sqrt{(5\mu/6)/(\mu/6)} = \mathrm{O}(1)$. □

3.3.3. Establishing the Lyapunov condition

To establish the Lyapunov condition w.r.t. the sequence Δ_t, it is by Lemma 12 crucial to bound the individual variances and the $(2+\delta)$-th central absolute moment. The variances have already been studied in Lemma 15. Using $\delta = 1$, we are left with the analysis of the third central moment. This is dealt with in the following lemma.

Lemma 16. *If $\mu = \omega(1)$ and $\mu = \mathrm{O}(\sqrt{n}\log n)$, then*

$$\mathrm{E}\big(|\Delta_t - \mathrm{E}(\Delta_t)|^3 \mid X_t\big) = \mathrm{O}(\mu^{3/2})\,.$$

Proof. We bound $\mathrm{E}\big(|\Delta_t - \mathrm{E}(\Delta_t)|^3 \mid X_t\big)$ by

$$\mathrm{E}\Big(\big(|\Delta_t| + |\mathrm{E}(\Delta_t)|\big)^3 \,\Big|\, X_t\Big)\,,$$

aiming at reusing the bounds on $\mathrm{E}(\Delta_t \mid X_t)$ we know from Lemma 13.

To treat the binomial expression raised to the third power, we use the simple bound

$$(a+b)^3 = a^3 + 3ab^2 + 3a^2b + b^3 \leq 4a^3 + 4b^3$$

for $a, b \geq 0$.

Thus,

$$\mathrm{E}\big(|\Delta_t - \mathrm{E}(\Delta_t)|^3 \mid X_t\big) \leq 4\mathrm{E}(|\Delta_t|^3 \mid X_t) + 4\,|\mathrm{E}(\Delta_t \mid X_t)|^3\,,$$

and we already have the bounds $-\mathrm{O}(\sqrt{\mu}) \leq \mathrm{E}(\Delta_t \mid X_t) = \mathrm{O}(\sqrt{\mu})$, which follow from Lemma 13 for all $X_t \in \{1, \dots, \mu - 1\}$ and $x = \mathrm{O}(\sqrt{n}\log n)$.

The main task left is to bound $\mathrm{E}(|\Delta_t|^3 \mid X_t)$. We claim that $\mathrm{E}(|\Delta_t|^3 \mid X_t) = \mathrm{O}\big(\mu^{3/2}\big)$. To show this, we assume an arbitrary X_t value. To bound the third moment, we analyze the distribution of $g(X_{t+1}) - g(X_t)$. We recall from Lemma 5 that X_{t+1} (i.e., the new value before applying the potential function) is given by the sum of two distributions, both of which are binomial or almost binomial; more precisely, $X_{t+1} = Z_{1,t+1} + Z(C^*)$, where $Z_{1,t+1}$ is the number of 1s sampled through 1st-class individuals in iteration $t+1$, C^* is the number of 2nd-class individuals, and $Z(C^*)$ is the number of 1s sampled by them. We note, using Lemmas 4 and 5, that $Z(C^*) \prec C^* \prec \mathrm{Bin}(\lambda, c/\sqrt{n}) + \widetilde{Z}$, for some constant $c > 0$, and $\widetilde{Z}$ takes some value from $1, \dots, \lambda$ only with probability at most $\mathrm{e}^{-\Omega(\mu)}$. Moreover, $Z_{1,t+1} \sim \mathrm{Bin}(\mu - C^*, X_t/\mu)$.

To overestimate $|\Delta_t| = |g(X_{t+1}) - g(X_t)|$, we observe that

$$\begin{aligned}|g(X_{t+1}) - g(X_t)| = {}& \big|g\big(Z_{1,t+1} + Z(C^*)\big) - g(X_t)\big| \cdot \mathbb{1}\{Z_{1,t+1} + Z(C^*) < X_t\}\\ &+ \big|g\big(Z_{1,t+1} + Z(C^*)\big) - g(X_t)\big| \cdot \mathbb{1}\{Z_{1,t+1} + Z(C^*) \geq X_t\}\,.\end{aligned}$$

Hence, to bound $|\Delta_t|$, it is enough to take the maximum of the two values

- $\Psi_1 := \Big|g\Big(\mathrm{Bin}\big(\mu, \frac{X_t}{\mu}\big)\Big) - g(X_t)\Big|$ and
- $\Psi_2 := \Big|g\Big(\mathrm{Bin}\big(\mu, \frac{X_t}{\mu}\big) + \mathrm{Bin}\big(\lambda, \frac{c}{\sqrt{n}}\big) + \widetilde{Z}\Big) - g(X_t)\Big|$

and analyze it. The first expression covers the case that $Z_{1,t+1} + Z(C^*) < X_t$. Then, we transform C^* random variables whose success probability is greater than X_t/μ (since 2nd-class individuals are biased toward 1s) into variables with success probability exactly X_t/μ, which increases the probability of $Z_{1,t+1} + Z(C^*)$ being less than X_t. On the other hand, if $Z_{1,t+1} + Z(C^*) \geq X_t$, we get an even larger value by including C^* additional experiments.

Bounding Ψ_1. We claim that $\mathrm{E}(|\Psi_1|^3 \mid X_t) = \mathrm{O}(\mu^{3/2})$. To show this, we proceed similarly as in computing the first moment of Δ_t and define intervals of length $\sqrt{x}$, where $x := X_t$ (hereinafter, we implicitly condition on this outcome). More precisely

$I_k := \{\lceil x-(k+1)\sqrt{x}\rceil, \dots, \lfloor x-k\sqrt{x}\rfloor\}$ for $k \in \mathbb{Z}$, i.e., also negative indices are allowed, leading to intervals lying above x. We get

$$\begin{aligned}\mathrm{E}(|\Psi_1|^3 \mid x) &\le \sum_{k=0}^{\infty} \sum_{i \in I_k \cup I_{-k}} \left(\frac{\sqrt{\mu}(|i-x|)}{x-(k+1)\sqrt{x}}\right)^3 \Pr(|\Psi_1| = |i|) \\ &\le \sum_{k=0}^{\infty} \left(\frac{\sqrt{\mu}(k+1)\sqrt{x}}{x-(k+1)\sqrt{x}}\right)^3 \Pr(|\Psi_1| \ge k\sqrt{x}) \,,\end{aligned}$$

by applying (2) to bound $g(x)-g(y)$ for $y<x$. Note that for $k \le \sqrt{x}$, we have by Chernoff bounds that $\Pr(X_{t+1} \in I_k) \le \Pr(X_{t+1} \le x-k\sqrt{x}) \le \mathrm{e}^{-k^2/3}$ and $\Pr(X_{t+1} \in I_{-k}) \le \Pr(X_{t+1} \ge x+k\sqrt{x}) \le \mathrm{e}^{-k^2/4}$. Moreover, $\Pr(X_{t+1} \le x/2) \le \mathrm{e}^{-x/24}$. Using the standard form of Chernoff bounds, we also bound the probability $\Pr\big(X_{t+1} \ge (1+j/2)x\big) \le \big(\mathrm{e}^{j/2}/(1+j/2)^{1+j/2}\big)^x \le \mathrm{e}^{-jx/10}$ for $j \ge 1$.

Using these different estimates while distinguishing between $k \le \sqrt{x}/2-1$ and $k \ge \sqrt{x}/2$, we get for $x \ge 1$ that

$$\begin{aligned}\mathrm{E}(|\Psi_1|^3 \mid x) &\le \sum_{k=0}^{\frac{\sqrt{x}}{2}-1} \left(\frac{\sqrt{\mu}(k+1)\sqrt{x}}{\frac{x}{2}}\right)^3 2\mathrm{e}^{-\frac{k^2}{4}} + \big(g(0)-g(x)\big)^3 \Pr\left(X_{t+1} \le \frac{x}{2}\right) \\ &\quad + \sum_{j=1}^{\infty} \left(g(x) - g\left(x\left(1+\frac{j}{2}\right)\right)\right)^3 \Pr\left(X_{t+1} \ge x + j\frac{x}{2}\right) \\ &\le \mathrm{O}\left(\mu^{\frac{3}{2}}\right) + (x\sqrt{\mu})^3 \mathrm{e}^{-\frac{x}{24}} + \sum_{j=1}^{\infty} \left(j\frac{x}{2}\sqrt{\mu}\right)^3 \mathrm{e}^{-j\frac{x}{10}} \\ &= \mathrm{O}\left(\mu^{\frac{3}{2}}\right) \,,\end{aligned}$$

where we use the trivial bound $g(x)-g(y) \le \sqrt{\mu}\,|x-y|$ and pessimistically assume $X_{t+1}=0$ in the case $X_{t+1} \le x/2$.

Bounding Ψ_2. With respect to Ψ_2, we observe that

$$\Psi_2 \prec \left|g\left(\mathrm{Bin}\left(\mu, \frac{x}{\mu}\right)\right) - g(x)\right| + \mathrm{O}(\mu)\Pr(\widetilde{Z} \ne 0) + \left(g(0) - g\left(\mathrm{Bin}\left(\lambda, \frac{c}{\sqrt{n}}\right)\right)\right)$$

by using $g(x+a+b)-g(x) = \big(g(x+a)-g(x)\big) + \big(g(x+a+b)-g(x+a)\big)$, for arbitrary $a, b \in \mathbb{R}$, and pessimistically estimating the contribution of $Z(C^*)$ to occur at point 0, where the potential function is steepest. Moreover, we pessimistically assume that the event $\widetilde{Z} \ne 0$ leads to the maximum possible change of g-value, which is $g(0)-g(\mu) = \mathrm{O}(\mu)$. Hence,

$$\mathrm{E}(|\Psi_2|^3 \mid x) \le 4\mathrm{E}\left(\left|g\left(\mathrm{Bin}\left(\mu, \frac{x}{\mu}\right)\right) - g(x)\right|^3\right) + 4\mathrm{E}\left(\left(g(0) - g\left(\mathrm{Bin}\left(\lambda, \frac{c}{\sqrt{n}}\right)\right)\right)^3\right) + \mathrm{O}(\mu^3)\cdot\Pr(\widetilde{Z} \ne 0) \,. \tag{7}$$

We recall that $\Pr(\widetilde{Z} \ne 0) \le \mathrm{e}^{-\Omega(\mu)}$, so that

$$\mathrm{O}(\mu^3)\cdot\Pr(\widetilde{Z} \ne 0) = \mathrm{O}(\mu^3)\cdot\mathrm{e}^{-\Omega(\mu)} = \mathrm{o}(1) = \mathrm{O}(\mu^{3/2})$$

for $\mu = \omega(1)$. Hence, the last term from Lemma 7 has already been bounded as desired, and we only have to show bounds on the first two terms of inequality (7).

We recognize that the first term of (7) is $\mathrm{O}(\mu^{3/2})$ since, up to constant factors, it is the same as $\mathrm{E}(|\Psi_1|^3 \mid X_t)$. Hence, we are left with the claim

$$\mathrm{E}\left(\left(g(0) - g\left(\mathrm{Bin}\left(\lambda, \frac{c}{\sqrt{n}}\right)\right)\right)^3\right) = \mathrm{O}\left(\mu^{\frac{3}{2}}\right) \,.$$

In order to show this, we let $Z \sim \mathrm{Bin}(\lambda, c/\sqrt{n})$ and consider different definitions of the intervals I_k, $k \ge 0$, that Z can fall into. The definition of intervals distinguishes two cases.

Case 1: $\lambda \ge \sqrt{n}/(2ec)$. As the derivative of $-g$ is at most $\sqrt{\mu}$, it suffices to prove the stronger claim

$$\sqrt{\mu}\cdot\mathrm{E}\left(\mathrm{Bin}\left(\lambda, \frac{c}{\sqrt{n}}\right)^3\right) = \mathrm{O}\left(\mu^{\frac{3}{2}}\right) \,.$$

We define $I_0 := [0, 2ec\lambda/\sqrt{n}]$ and $I_k := [(1+k)ec\lambda/\sqrt{n}, (2+k)ec\lambda/\sqrt{n}]$ for $k \geq 1$. Then (similar to the analysis of $\mathrm{E}(|\Psi_1|^3 \mid x)$), we get

$$\mathrm{E}\left(\mathrm{Bin}\left(\lambda, \frac{c}{\sqrt{n}}\right)^3\right) \leq \left(\frac{2ec\lambda}{\sqrt{n}}\right)^3 + \sum_{k=1}^{\infty}\left(\frac{(2+k)ec\lambda}{\sqrt{n}}\right)^3 \Pr(Z \in I_k) .$$

We use the Chernoff bound $\Pr(X \geq t) \leq 2^{-t}$ for $t \geq 2\mathrm{e}\mathrm{E}(X)$. This gives us $\Pr(Z \in I_k) \leq \mathrm{e}^{-(2+k)\mathrm{e}\lambda/\sqrt{n}} \leq \mathrm{e}^{-k/2}$ by our assumption on λ. We get

$$\begin{aligned}\mathrm{E}\left(\mathrm{Bin}\left(\lambda, \frac{c}{\sqrt{n}}\right)^3\right) &\leq \mathrm{O}\left(\frac{\lambda^3}{n^{\frac{3}{2}}}\right) + \mathrm{O}\left(\frac{\lambda^3}{n^{\frac{3}{2}}}\right)\sum_{k=1}^{\infty}(2+k)^3\mathrm{e}^{-\frac{k}{2}} \\ &= \mathrm{O}\left(\frac{\lambda^3}{n^{\frac{3}{2}}}\right) \\ &= \mathrm{O}\left(\frac{\mu^3}{n^{\frac{3}{2}}}\right) ,\end{aligned}$$

hence $\sqrt{\mu} \cdot \mathrm{E}\left(\mathrm{Bin}(\lambda, c/\sqrt{n})^3\right) = \mathrm{O}(\mu^{7/2}/n^{3/2})$. Since $\mu = \mathrm{O}(\sqrt{n}\log n)$ by assumption of the lemma, the bound is at most $\mathrm{O}\left(n^{1/4}(\log n)^{7/2}\right)$, and this is clearly $\mathrm{O}(\mu^{3/2})$, since $\mu = \Omega(\sqrt{n})$ in this case.

Case 2: $\lambda < \sqrt{n}/2\mathrm{e}$. Then $I_k := [k, k+1]$ for $k \geq 0$. We note that $\mathrm{E}(Z) = \mathrm{O}(1)$ since $\mu = \mathrm{O}(\lambda) = \mathrm{O}(\sqrt{n})$. Hence, by Chernoff bounds for $k > \mathrm{E}(Z)$, $\Pr(Z \geq k) = \mathrm{e}^{-\alpha k}$ for some constant $\alpha > 0$. We get

$$\mathrm{E}\left(\left(g(0) - g(Z)\right)^3\right) \leq (\sqrt{\mu})^3 \cdot \mathrm{E}(Z^3) \leq \mu^{3/2} \cdot \mathrm{E}(Z)^3 + \sum_{k > \mathrm{E}(Z)}^{\infty} (\mu k)^3 2^{-\alpha k} .$$

Thus, using $\mu = \mathrm{O}(\sqrt{n})$,

$$\begin{aligned}\mathrm{E}\left(\left(g(0) - g(Z)\right)^3\right) &\leq \mathrm{O}\left((\sqrt{\mu})^3\right) + (\sqrt{\mu})^3 \sum_{k=1}^{\infty} k^3 2^{-\alpha k} \\ &= \mathrm{O}\left(\mu^{\frac{3}{2}}\right) ,\end{aligned}$$

which completes the proof. □

Using Lemmas 15 and 16, we now establish the Lyapunov condition, assuming $X_t \leq (5/6)\mu$ for all $t \geq 0$. Using Lemma 12, we get for $s_t^2 := \sum_{j=0}^{t-1} \mathrm{Var}(\Delta_j \mid X_j)$ that

$$\frac{1}{s_t^3}\sum_{j=0}^{t-1} \mathrm{E}\left(|\Delta_j - \mathrm{E}(\Delta_j)|^3 \mid X_j\right) = \mathrm{O}\left(\frac{\mu^{1.5}t}{\mu^{1.5}t^{1.5}}\right) = \mathrm{O}\left(\frac{1}{\sqrt{t}}\right) ,$$

which is $\mathrm{o}(1)$ for $t = \omega(1)$. The sum of the Δ_j can then be approximated as stated in the following lemma.

Lemma 17. *Let $Y_t := \sum_{j=0}^{t-1} \Delta_j$ and $t = \omega(1)$. Then*

$$\frac{Y_t - \mathrm{E}(Y_t \mid X_0)}{\sqrt{\sum_{j=0}^{t-1} \mathrm{Var}(\Delta_j \mid X_j)}}$$

converges in distribution to $\mathrm{N}(0, 1)$. *The absolute error of this approximation is* $\mathrm{O}(1/\sqrt{t})$.

3.3.4. Likelihood of a frequency getting very small

We will now apply Lemma 17 in order to prove how likely it is for a single frequency to either get close to $1/n$ or exceed $5/6$. For this, we will use the following estimates for $\Phi(x)$. More precise formulas exist, but they do not yield any benefit in our analysis.

Lemma 18 *([24, p. 175]). For any $x > 0$,*

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}} \leq 1 - \Phi(x) \leq \frac{1}{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}} ,$$

and for $x < 0$,

$$\left(\frac{-1}{x} - \frac{-1}{x^3}\right)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}} \leq \Phi(x) \leq \frac{-1}{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\,.$$

Lemma 19. *Consider a bit of UMDA on* OneMax *and let p_t be its frequency in iteration t. We say that the process breaks a border at time t if $\min\{p_t, 1-p_t\} \leq 1/n$. Given $s < 0$ and any starting state $p_0 \leq 5/6$, let T_s be the smallest t such that $p_t - p_0 \leq s$ holds or a border is broken.*

Assume that $\Theta(n)$ other frequencies stay within $[1/6, 5/6]$ until time T_s. Choosing $0 < \alpha < 1$, where $1/\alpha = \mathrm{o}(\mu)$ and $\alpha = \mathrm{O}(\sqrt{n}/\mu)$, and $-1 < s < 0$ constant, we then have for some constant $\kappa > 0$ that

$$\begin{aligned}&\Pr\left(T_s \leq \alpha s^2\mu \text{ or } p_t \text{ exceeds } \tfrac{5}{6} \text{ before } T_s\right)\\ &\geq \left(\frac{(|s|\,\alpha)^{\frac{1}{2}}}{\kappa} - \frac{(|s|\,\alpha)^{\frac{3}{2}}}{\kappa^3}\right)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{\kappa^2}{2|s|\alpha}} - \mathrm{O}\left(\frac{1}{\sqrt{\alpha\mu}}\right).\end{aligned}$$

Proof. Throughout the analysis, we assume $X_t \leq (5/6)\mu$, since all considerations are stopped when the frequency exceeds 5/6, i.e., when $X_t \geq (5/6)\mu$. By Lemma 13, we have $\mathrm{E}(\Delta_j \mid X_j) \geq -\sqrt{\mu/(X_j+1)}\big(\mathrm{e}^{-\Omega(\mu)} + \gamma_1(X_j/\sqrt{n} + X_j/\mu)\big)$ for all $j \geq 0$ and $1 \leq X_j \leq \mu - 1$, where $\gamma_1 > 0$ is a sufficiently large constant. Moreover, according to Lemma 15, $\mathrm{Var}(\Delta_j \mid X_j) \geq c\mu$ for some constant $c > 0$. Since the Lyapunov condition has been established for $Y_t \coloneqq \sum_{j=0}^{t-1}\Delta_j$ in Lemma 17, we know that $\big(Y_t - \mathrm{E}(Y_t \mid X_0)\big)/s_t$ converges in distribution to $\mathrm{N}(0,1)$ if $t = \omega(1)$. The lemma chooses $t = \alpha s^2\mu$, which is $\omega(1)$ since $\alpha = \omega(1/\mu)$ by assumption.

For $s_t^2 \coloneqq \sum_{j=0}^{t-1}\mathrm{Var}(\Delta_j \mid X_j)$, we obtain $s_t^2 \geq \alpha s^2 c\mu^2$. Hence, recalling that $s < 0$ is assumed, we get $s_t \geq \sqrt{\alpha c}\,|s|\,\mu$. The next task is to bound $\mathrm{E}(Y_t)$. Using our bound on $\mathrm{E}(\Delta_j \mid X_j)$ and recalling that $0 \leq X_t \leq (5/6)\mu$ and $\mu = \omega(1)$, we have

$$\begin{aligned}\mathrm{E}(\Delta_t \mid X_t) &\geq -\left(\mathrm{e}^{-\Omega(\mu)}\sqrt{\frac{\mu}{1}} + \gamma_1\frac{\frac{5}{6}\mu}{\sqrt{\frac{5}{6}\mu+1}}\left(\frac{\sqrt{\mu}}{\sqrt{n}} + \frac{1}{\sqrt{\mu}}\right)\right)\\ &\geq -\left(\mathrm{O}(1) + \gamma_2\frac{\mu}{\sqrt{n}}\right),\end{aligned}$$

for some constant $\gamma_2 > 0$.

This implies $\mathrm{E}(Y_t) \geq -t\big(\mathrm{O}(1) + \gamma_2\mu/\sqrt{n}\big) = -\alpha s^2\mu\big(\mathrm{O}(1) + \gamma_2\mu/\sqrt{n}\big)$. Therefore,

$$\frac{\mathrm{E}(Y_t)}{s_t} \geq -\frac{(\alpha s^2\mu)\big(\mathrm{O}(1) + \gamma_2\frac{\mu}{\sqrt{n}}\big)}{\sqrt{\alpha c}\,|s|\,\mu} \geq -\gamma_3\sqrt{\frac{1}{c\alpha}}\,,$$

for some constant $\gamma_3 > 0$ depending on α, using the assumptions $|s| \leq 1$ along with both $\alpha \leq 1$ and $\alpha = \mathrm{O}(\sqrt{n}/\mu)$.

To bound $\Pr(Y_t \geq r)$ for arbitrary r, we note that

$$Y_t \geq r \iff \frac{Y_t}{s_t} - \frac{\mathrm{E}(Y_t \mid X_0)}{s_t} \geq \frac{r}{s_t} - \frac{\mathrm{E}(Y_t \mid X_0)}{s_t}\,,$$

and recall that the distribution of $Y_t/s_t - \mathrm{E}(Y_t \mid X_0)/s_t$ converges to $\mathrm{N}(0,1)$ with absolute error $\mathrm{O}(1/\sqrt{t})$. Hence,

$$\Pr(Y_t \geq r) \geq 1 - \Phi\left(\frac{r}{\sqrt{c\alpha}\,|s|\,\mu} + \gamma_3\sqrt{\frac{1}{c\alpha}}\right) - \mathrm{O}\left(\frac{1}{\sqrt{t}}\right) \tag{8}$$

for any r such that the argument of Φ is positive, where Φ denotes the cumulative distribution function of the standard normal distribution.

We focus on the event E^* that $Y_t \geq 2\mu\sqrt{|s|}$, recalling that $s < 0$ and $X_t \geq X_0 \Leftrightarrow Y_t \leq Y_0$. Note that E^* means $g(X_t) - g(X_0) \geq 2\mu\sqrt{|s|}$, and this implies an upper bound on the negative $X_t - X_0$ as follows: function g is steepest at point 0, and from the definition for any $y \geq 1$,

$$\begin{aligned}g(y) - g(0) &\leq \sum_{j=0}^{y-1}\sqrt{\frac{\mu}{j+1}}\\ &\leq \sqrt{\mu}\left(1 + \int_1^y \frac{1}{\sqrt{j}}\,\mathrm{d}j\right)\end{aligned}$$

$$= \sqrt{\mu}(1 + 2\sqrt{y} - 2\sqrt{1})$$
$$\leq 2\sqrt{y\mu}\,.$$

Thus, the event $g(X_t) - g(X_0) \geq a$ for $a > 0$ is only possible if $X_t \leq X_0 - a^2/(4\mu)$. In other words, event E^* implies $X_t - X_0 \leq s\mu$, which is equivalent to $p_t - p_0 \leq s$. Hence, in order to complete the proof, we only need a lower bound on the probability of E^*. Setting $r := 2\mu\sqrt{|s|}$ in (8), we bound the argument of Φ according to

$$\frac{r}{\sqrt{c\alpha}\,|s|\,\mu} + \frac{\gamma_3}{\sqrt{c\alpha}} \leq \frac{2}{\sqrt{c\,|s|\,\alpha}} + \frac{\gamma_3}{\sqrt{c\alpha}} \leq \frac{\gamma_4}{\sqrt{c\,|s|\,\alpha}}\,,$$

for some constant $\gamma_4 > 0$, since $|s| \leq 1$.

By Lemma 18,

$$1 - \Phi\left(\frac{\gamma_4}{\sqrt{c\,|s|\,\alpha}}\right) \geq \left(\frac{\sqrt{c\,|s|\,\alpha}}{\gamma_4} - \frac{(\sqrt{c\,|s|\,\alpha})^3}{\gamma_4^3}\right)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{\gamma_4^2}{2cs\alpha}}$$
$$=: p(\alpha, s)\,,$$

which means that the frequency changes by s (which is negative) until iteration $\alpha s^2\mu$ with probability at least $p(\alpha, s) - \mathrm{O}(1/\sqrt{t}) = p(\alpha, s) - \mathrm{O}(1/\sqrt{\alpha\mu})$, where the last term stems from the bound on the absolute error of the approximation by the Normal distribution. Choosing $\kappa := \gamma_4/\sqrt{c}$ in the statement of the lemma completes the proof. □

3.4. Proof of the lower bound

Finally, we put all previous lemmas together to prove our main theorem: Theorem 6.

Proof of Theorem 6. As outlined above, we distinguish between three regimes for λ. The case of small λ ($\lambda < (1 - c_1)\log_2 n$) is covered by Theorem 8, noting that $\Omega(n \log n)$ dominates the lower bound for the considered range of μ. The case of large λ ($\mu = \Omega(\sqrt{n}\log n)$) is covered by Corollary 11. We are left with the medium case ($\mu = \Omega(\log n)$ and $\mu = \mathrm{o}(\sqrt{n}\log n)$), which is the most challenging one to prove.

In the following, we consider a phase consisting of $T := s^2\gamma \cdot \min\{\mu, \sqrt{n}\}$ iterations, for the constant $\gamma > 0$ from Lemma 10; without loss of generality, $\gamma < 1$ is assumed. We conceptually split individuals (i.e., bit strings) of UMDA into two substrings of length $n/2$ each and apply Lemma 10 w.r.t. the first half of the bits. In the following, we condition on the event that $\Theta(n)$ frequencies from the first half are within the interval $[1/6, 5/6]$ throughout the phase.

We show next that some frequencies from the second half are likely to walk down to the lower border. Let j be an arbitrary position from the second half. First, we apply Lemma 9. Hence, p_j does not exceed $5/6$ within the phase with probability $\Omega(1)$. In the following, we condition on this event.

We then revisit bit j and apply Lemma 19 in order to show that, under this condition, the random walk on its frequency p_j achieves a negative displacement. Note that the event of not exceeding a certain positive displacement (more precisely, the displacement of $5/6 - 1/2 = 1/3$) is positively correlated with the event of reaching a given negative displacement (formally, the state of the conditioned stochastic process is always stochastically smaller than of the unconditioned process). We can therefore apply Lemma 19 for a negative displacement of $s := -5/6$ within T iterations. Note that the condition of the lemma that demands $\Theta(n)$ frequencies to be within $[1/6, 5/6]$ is satisfied by our assumption concerning the first half of the bits. Choosing $\alpha = T/(s^2\mu)$, we get $1/\alpha = \mathrm{o}(\log n)$ (since $\mu = \mathrm{o}(\sqrt{n}\log n)$ and $T = \Theta(\min\{\mu, \sqrt{n}\})$), whereby we easily satisfy the assumption $1/\alpha = \mathrm{o}(\mu)$. As $T = \mathrm{O}(\sqrt{n})$ and s constant, we also satisfy the assumption $\alpha = \mathrm{O}(\sqrt{n}/\mu)$. Moreover, $\alpha \leq \gamma < 1$ by definition. Now, Lemma 19 states that the probability of the random walk on p_j reaching a total displacement of $-5/6$ (or hitting the lower border before) within the phase of length T is at least

$$\left(\frac{(|s|\,\alpha)^{\frac{1}{2}}}{\kappa} - \frac{(|s|\,\alpha)^{\frac{3}{2}}}{\kappa^3}\right)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{\kappa^2}{2\cdot|s|\alpha}} - \mathrm{O}\left(\frac{1}{\sqrt{\alpha\mu}}\right)\,. \tag{9}$$

In order to bound the last expression from below, we distinguish between two cases. If $\mu \leq \sqrt{n}$, then $\alpha = \Omega(1)$ and (9) is at least

$$\Omega(1) - \mathrm{O}\left(\frac{1}{\sqrt{\mu}}\right) = \Omega(1)\,,$$

since $T = \Omega(\mu) = \Omega(\log n) = \omega(1)$. If $\mu \geq \sqrt{n}$, then we have $T = \Omega(\sqrt{n})$. Since $1/\alpha = \mathrm{o}(\log n)$, we estimate (9) from below by

$$\Omega\left(\frac{1}{\mathrm{o}(\sqrt{\log n})} \cdot \mathrm{e}^{-\mathrm{o}(\ln n)}\right) - \mathrm{O}\left(\frac{\log n}{n^{1/4}}\right) \geq n^{-\eta}\,,$$

for some $\eta = \eta(n) = o(1)$. Combining this with the probability of not exceeding $5/6$, the probability of p_j hitting the lower border within T iterations is, in any case, $\Omega(n^{-\eta})$. Note that this argumentation applies to every of the last $n/2$ bits, and, as explained in Section 2.2, the bounds derived hold independently for all these bits. Hence by Chernoff bounds, with probability $1 - 2^{-\Omega(n^{1-\eta})}$, the number of frequencies from the second half that hit the lower border within T iterations is $\Omega(n^{1-\eta})$.

A frequency that has hit the lower border $1/n$ somewhere in the phase may recover (i.e., reach a larger value) by the end of the phase. However, for each bit the probability of not recovering is at least

$$\left(1 - \frac{1}{n}\right)^{T\lambda} \geq e^{-o(\log n)} = n^{-\eta'}$$

for some $\eta' = o(1)$, since we consider $T = O(\sqrt{n})$ iterations and $\lambda = o(\sqrt{n}\log n)$ samples per iteration. Again applying Chernoff bounds leaves $\Omega(n^{1-\eta-\eta'})$ bits at the lower border at iteration T with probability $1 - 2^{-\Omega(n^{1-\eta-\eta'})}$.

Now, making use of Lemma 7 gives us the desired run time bound. □

4. Relaxing the condition on the population size

Theorem 6, which most of the paper was concerned with, assumed that $\lambda = (1+\beta)\mu$ for some constant $\beta > 0$. We think that the lower bound of $\Omega(n \log n)$ holds for all combinations of μ and λ. As a step toward a proof of this conjecture, we extend our lower bound toward all $\mu \leq c \log n$ for a sufficiently small constant $c > 0$. This includes the extreme case of $\mu = 1$, for which no matching upper bound has been proved up to date.

Theorem 20. *Let $\mu \leq c \log n$ for a sufficiently small constant $c > 0$, and let $\lambda = n^{O(1)}$. Then the optimization time of UMDA on* OneMax *is $\Omega(\lambda + n \log n)$ with high probability and in expectation.*

Proof. The lower bound λ follows since UMDA will sample the optimum in the first iteration only with a probability of $2^{-\Theta(n)}$. Thus, with high probability, all λ offspring from the first generation need to be evaluated. In the following, we assume $\lambda = O(n \log n)$ since otherwise nothing is left to show.

We now follow the ideas underlying the proof of Theorem 8 by showing that the best μ individuals from the initial generation are still close to uniform, resulting in many frequencies being set to their minimum $1/n$. Note that the mentioned theorem considered all λ individuals from the initial generation, which are uniform on the search space. Here we focus on the best μ from the initial population, which violates the independence.

By Chernoff bounds, the probability that at least one of the λ initial individuals has $3n/4$ or more 1s is at most $\lambda e^{-\Omega(n)} = e^{-\Omega(n)}$. In the following, we condition on this not happening. Let us consider an arbitrary individual of the λ initial individuals. Clearly, given that it has k 1s, the actual distribution of 1s is uniform over all permutations of k 1s. This still applies to the selected best μ individuals since OneMax is unbiased with respect to permutations, i.e., only depends on the number of 1s. Hence, we get the following property $(*)$: if we consider an arbitrary individual from the μ best, then every bit in it takes the value 1 with the same probability p^* (not necessarily independently of the other bits). Since the expected number of 1s is bounded by $3n/4$, we have $p^* \leq 3/4$; otherwise, the expected value would be larger, which we excluded. Pessimistically assuming that all λ individuals have $3n/4$ 1s, we obtain $p^* = 3/4$ and have established the property $(*)$ independently for all individuals (also when arguing only about the best μ ones) but still not independently for all bits.

We now consider an arbitrary bit position i from one of the best μ individuals. If bit i takes the value 0, then the $3n/4$ 1s have to be taken at positions other than i and are uniformly distributed among these positions. Hence, any bit $j \neq i$ takes the value 1 with probability at most $(3n/4)/(n-1)$ and 0 with probability at least $1 - (3n/4)/(n-1) = (n/4-1)/(n-1)$. Altogether, independently of the outcome of i, bit j takes the value 0 with probability at least $\min\{1/4, (n/4-1)/(n-1)\} = (n/4-1)/(n-1)$. We iterate this argument over an arbitrary set S^* consisting of at most $n/8$ bits (e.g., the first $n/8$ positions). Hence, every of these bits takes the value 0 with probability at least

$$\frac{\frac{n}{4} - \frac{n}{8}}{n - \frac{n}{8}} = \frac{1}{7},$$

independently of the other bits in S^*. As this applies independently to all μ best individuals, each bit in S^* is set to 0 in all μ best individuals with probability at least

$$\left(\frac{1}{7}\right)^{\mu},$$

independently of the other bits in S^*.

Thanks to the independence achieved by the estimations, we can now apply Chernoff bounds w.r.t. to the sum of the indicator random variables associated with the events "bit i is set to 0 in all μ best individuals" over all $i \in S^*$. The

expected number of such bits is at least $\ell := (n/8)(1/7)^{\mu}$. If we choose $\mu \leq c \log n$ for a sufficiently small constant $c > 0$, we obtain, for a constant $c' > 0$, that $\ell \geq n^{c'}/8$. Moreover, the probability that fewer than $n^{c'}/9$ bits take the value 0 in all μ best individuals is $2^{-\Omega(n)}$ then. We assume this to happen and note that the failure probability altogether is $2^{-\Omega(n)}$. Now Lemma 7 yields the theorem. □

5. Conclusions

We have analyzed UMDA on OneMax and obtained the general bound $\Omega(\lambda + \mu\sqrt{n} + n \log n)$ on its expected run time for combinations of μ and λ where $\lambda = O(\mu)$ or $\mu \leq c \log n$ (for a sufficiently small constant c). This lower bound analysis is the first of its kind and contributes advanced techniques, including potential functions.

We note that our lower bound for UMDA is tight in many cases, as has been shown recently [8,10]. We also note that our main result assumes $\lambda = O(\mu)$. However, we do not think that larger λ can be beneficial; if $\lambda = \alpha\mu$, for $\alpha = \omega(1)$, the progress due to 2nd-class individuals can be by a factor of at most α bigger; however, also the computational effort per generation would grow by this factor. Still, we have not presented a formal proof for all such cases.

Further run time analyses of UMDA or other EDAs for other classes of functions are an obvious subject for future research. In this respect, we hope that our technical contributions are useful and can be extended toward a more general lower bound technique at some point.

Acknowledgments

Financial support by the Danish Council for Independent Research (DFF-FNU 4002-00542) is gratefully acknowledged.

The authors would like to thank the anonymous reviewers of the conference and journal version for their comments, which greatly improved the quality of this paper.

References

[1] P. Larrañaga, J.A. Lozano (Eds.), Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation, Genet. Algorithms Evol. Comput., vol. 2, Springer, 2002.
[2] M. Hauschild, M. Pelikan, An introduction and survey of estimation of distribution algorithms, Swarm Evol. Comput. 1 (3) (2011) 111–128.
[3] S. Droste, A rigorous analysis of the compact genetic algorithm for linear functions, Nat. Comput. 5 (3) (2006) 257–283.
[4] F. Neumann, D. Sudholt, C. Witt, A few ants are enough: ACO with iteration-best update, in: Proc. of GECCO '10, 2010, pp. 63–70.
[5] D. Dang, P.K. Lehre, Simplified runtime analysis of estimation of distribution algorithms, in: Proc. of GECCO '15, 2015, pp. 513–518.
[6] T. Friedrich, T. Kötzing, M.S. Krejca, EDAs cannot be balanced and stable, in: Proc. of GECCO '16, 2016, pp. 1139–1146, http://doi.acm.org/10.1145/2908812.2908895.
[7] T. Friedrich, T. Kötzing, M.S. Krejca, A.M. Sutton, The benefit of recombination in noisy evolutionary search, in: Proc. of ISSAC '15, 2015, pp. 140–150.
[8] 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.
[9] D. Sudholt, C. Witt, Update strength in EDAs and ACO: how to avoid genetic drift, in: Proc. of GECCO '16, 2016, pp. 61–68.
[10] C. Witt, Upper bounds on the runtime of the univariate marginal distribution algorithm on OneMax, in: Proc. of GECCO '17, 2017, pp. 1415–1422.
[11] D. Sudholt, A new method for lower bounds on the running time of evolutionary algorithms, IEEE Trans. Evol. Comput. 17 (3) (2013) 418–435.
[12] C. Witt, Tight bounds on the optimization time of a randomized search heuristic on linear functions, Combin. Probab. Comput. 22 (2) (2013) 294–318.
[13] H. Mühlenbein, G. Paass, From recombination of genes to the estimation of distributions I. Binary parameters, in: Proc. of PPSN IV, 1996, pp. 178–187.
[14] T. Chen, P.K. Lehre, K. Tang, X. Yao, When is an estimation of distribution algorithm better than an evolutionary algorithm?, in: Proc. of CEC '09, 2009, pp. 1470–1477.
[15] T. Chen, K. Tang, G. Chen, X. Yao, On the analysis of average time complexity of estimation of distribution algorithms, in: Proc. of CEC '07, 2007, pp. 453–460.
[16] T. Chen, K. Tang, G. Chen, X. Yao, Rigorous time complexity analysis of univariate marginal distribution algorithm with margins, in: Proc. of CEC '09, 2009, pp. 2157–2164.
[17] T. Chen, K. Tang, G. Chen, X. Yao, Analysis of computational time of simple estimation of distribution algorithms, IEEE Trans. Evol. Comput. 14 (1) (2010) 1–22.
[18] S. Droste, T. Jansen, I. Wegener, Upper and lower bounds for randomized search heuristics in black-box optimization, Theory Comput. Syst. 39 (2006) 525–544.
[19] J. Jägersküpper, T. Storch, When the plus strategy outperforms the comma strategy—and when not, in: Proc. of FOCI '07, 2007, pp. 25–32.
[20] M.S. Krejca, C. Witt, Lower bounds on the run time of the univariate marginal distribution algorithm on OneMax, in: Proc. of FOGA '17, 2017, pp. 65–79.
[21] D. Sudholt, C. Witt, Update strength in EDAs and ACO: how to avoid genetic drift, arXiv:1607.04063.
[22] W. Feller, An Introduction to Probability Theory and Its Applications, vol. 2, Wiley, 1971.
[23] P.S. Oliveto, C. Witt, Improved time complexity analysis of the simple genetic algorithm, Theoret. Comput. Sci. 605 (2015) 21–41.
[24] W. Feller, An Introduction to Probability Theory and Its Applications, vol. 1, Wiley, 1968.

Please cite this article in press as: M.S. Krejca, C. Witt, Lower bounds on the run time of the Univariate Marginal Distribution Algorithm on OneMax, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2018.06.004