

Improved Fixed-Budget Results via Drift Analysis

Timo Ktzing
Hasso Plattner Institute
Potsdam
Germany

Carsten Witt
Technical University of Denmark
Kgs. Lyngby
Denmark

June 15, 2020

Abstract

Fixed-budget theory is concerned with computing or bounding the fitness value achievable by randomized search heuristics within a given budget of fitness function evaluations. Despite recent progress in fixed-budget theory, there is a lack of general tools to derive such results. We transfer drift theory, the key tool to derive expected optimization times, to the fixed-budget perspective. A first and easy-to-use statement concerned with iterating drift in so-called greed-admitting scenarios immediately translates into bounds on the expected function value. Afterwards, we consider a more general tool based on the well-known variable drift theorem. Applications of this technique to the LEADINGONES benchmark function yield statements that are more precise than the previous state of the art.

1 Introduction

Randomized search heuristics are a class of optimization algorithms which use probabilistic choices with the aim of maximizing or minimizing a given objective function. Typical examples of such algorithms use inspiration from nature in order to determine the method of search, most prominently evolutionary algorithms, which use the concepts of *mutation* (slightly altering a solution) and *selection* (giving preference to solutions with better objective value).

The theory of randomized search heuristics aims at understanding such heuristics by explaining their optimization behavior. Recent results are typically phrased as run time results, for example by giving upper (and lower) bounds on the expected time until a solution of a certain quality (typically the best possible quality) is found. This is called the (expected) *optimization time*. A different approach, called *fixed-budget analysis*, bounds the quality of the current solution of the heuristic after a given amount of time. In order to ease the analysis and by convention, in this theoretical framework *time* is approximated as the number of evaluations of the objective function (called fitness evaluations).

In this paper we are concerned with the approach of giving a fixed-budget analysis. This approach was introduced to the analysis of randomized search heuristics by

Jansen and Zarges [9], who derived fixed-budget results for the classical example functions ONEMAX and LEADINGONES by bounding the expected progress in each iteration. A different perspective was proposed by Doerr, Jansen, Witt and Zarges [2], who showed that fixed-budget statements can be derived from bounds on optimization times if these exhibit strong concentration. Lengler and Spooner [15] proposed a variant of multiplicative drift for fixed-budget results and the use of differential equations in the context of ONEMAX and general linear functions. Nallaperuma, Neumann and Sudholt [17] applied fixed-budget theory to the analysis of evolutionary algorithms on the traveling salesman problem and Jansen and Zarges [10] to artificial immune systems. The quality gains of optimal black-box algorithms on ONEMAX in a fixed-budget perspective were analyzed by Doerr, Doerr and Yang [3]. In a recent technical report, He, Jansen and Zarges [6] consider so-called unlimited budgets to estimate fitness values in particular for points of time larger than the expected optimization time. A recent survey by Jansen [8] summarizes the state of the art in the area of fixed-budget analysis.

There are general methods easing the analysis of randomized search heuristics. Most importantly, in order to derive bounds on the optimization time, we can make use of drift theory. Drift theory is a general term for a collection of theorems that consider random processes and bound the expected time it takes the process to reach a certain value - the first-hitting time. The beauty and appeal of these theorems lie in them usually having few restrictions but yielding strong results. Intuitively speaking, in order to use a drift theorem, one only needs to estimate the expected change of a random process - the drift - at any given point in time. Hence, a drift theorem turns expected local changes of a process into expected first-hitting times. In other words, local information of the process is transformed into global information. See [14] for an extensive discussion of drift theory.

In contrast to the numerous drift theorems available for bounding the optimization time, there is no corresponding theorem for making a fixed-budget analysis apart from one for the multiplicative case given in [15]. With this paper we aim to provide several such drift theorems, applicable in different settings and with a different angle of conclusions. In each our main goal is to provide an *upper bound* on the distance to the optimum after t iterations, for t less than the expected optimization time. Upper bounds alone do not allow for a fair comparison of algorithms, since a bad upper bound does not exclude the possibility of a good performance of an algorithm; for this, we require lower bounds. However, one of our techniques also allows us to derive lower bounds. Furthermore, when upper and lower bounds are close together we can conclude that the derived bounds are correspondingly tight, highlighting the quality of our methods.

We start, in Section 3, by giving a theorem which iteratively applies local drift estimates to derive a global drift estimate after t iterations. Crucial for this theorem is that the drift condition is *unlimited time*, by which we mean that the drift condition has to hold for all times t , not just (which is the typical case in the literature for drift theorems) those before the optimum is hit. This theorem is applicable in the case where there is no optimum (and optimization progresses indefinitely) and in the case that, in the optimum, the drift is 0. In order to bypass these limitations we also give a variant

in Section 3 which allows for *limited time* drift, where the drift condition only needs to hold before the optimum is hit; however, in this case we pick up an additional error term in the result, derived from the possibility of hitting the optimum within the allowed time budget of t . Thus, in order to apply this theorem, one will typically need concentrations bounds for the time to hit the optimum.

For both these theorems, the drift function (bounding the drift) has to be convex and *greed-admitting*, which intuitively says that being closer to the goal is always better in terms of the expected state after an additional iteration, while search points closer to the goal are required to have weaker drift. These conditions are fulfilled in many sample applications; as examples we give analyses of the (1+1) EA on LEADINGONES and ONEMAX. Note that these analyses seem to be rather tight, but we do not offer any lower bounds, since our techniques crucially only apply in one direction (owing to an application of Jensen’s Inequality to convex drift functions).

In Section 4 we use a potential-based approach and give a variable drift theorem for fixed-budget analysis. As a special case, where the drift function is constant, we give an additive drift theorem for fixed-budget analysis and derive a result for (1+1) EA on LEADINGONES. In general, the approach bounds the expected value of the potential but not of the fitness. Therefore, we also study how to derive a bound on the fitness itself, both from above and from below, by inverting the potential function and using tail bounds on its value. The approach uses a generalized theorem showing tail bounds for martingale differences, which overcomes a weakness of existing martingale difference theorems in our specific application. This generalization may be of independent interest.

Our results allow for giving strong fixed-budget results which were not obtainable before. For the (1+1) EA on LEADINGONES with a budget of $t = o(n^2)$ iterations, the original paper [9] gives a lower bound of $2t/n - o(t/n)$ for the expected fitness after t iterations, which we recover with a simple proof in Theorem 6. Our theorem also allows budgets closer to the expected optimization time, where we get a lower bound of $n \ln(1 + 2t/n^2) - O(1)$.

For the (1+1) EA on ONEMAX, no concrete formula for a bound on the fitness value after t iterations was known: The original work [9] could only handle RLS on ONEMAX, not the (1+1) EA. The multiplicative drift theorem of [15] allows for deriving a lower bound of $n/2 + t/(2e)$ for $t = o(n)$ using a multiplicative drift constant of $(1 - 1/n)^n/n$. Since our drift theorem allows for variable drift, we can give a bound of $n/2 + t/(2\sqrt{e}) - o(t)$ for the (1+1) EA on ONEMAX with $t = o(n)$ (see Theorem 5). Note that [15] also gives bounds for values of t closer to the expected optimization time.

Furthermore, we are not only concerned with expected values but also give strong concentration bounds. We consider the (1+1) EA on LEADINGONES and show that the fitness after t steps is strongly concentrated around its expectation (see Theorem 10). The error term obtained is asymptotically smaller than in the previous work [2] and the statement is also less complex.

Fixed-budget results that hold with high probability are crucial for the analysis of algorithm configurators [5]. These configurators test different algorithms for fixed budgets in order to make statements about their appropriateness in a given setting.

Thus, we believe that this work also contributes to the better understanding of the strengths and weaknesses of algorithm configurators.

The remainder of the paper is structured as follows. Next we give mathematical preliminaries, covering problem and algorithm definitions as well as some well-known results from the literature which we require later. In Section 3 we give our direct fixed-budget drift theorems, as well as its applications to the (1+1) EA on ONEMAX and LEADINGONES. In Section 4 we give a variable fixed-budget drift theorem and its corollary for additive drift. We show how to apply this variable fixed-budget drift theorem to obtain very strong bounds in Section 5. We conclude in Section 6.

2 Preliminaries

The concrete objective functions we are concerned with in this paper are ONEMAX and LEADINGONES, studied in a large number of papers. These two functions are defined as follows. For a fixed natural number n , the functions map bit strings $x \in \{0, 1\}^n$ of length n to natural numbers such that

$$\text{ONEMAX}(x) = \sum_{i=1}^n x_i$$

is the number of 1s in the bit string x and

$$\text{LEADINGONES}(x) = \sum_{i=1}^n \prod_{j=1}^i x_j$$

is the number of leading 1s in x before the first 0 (if any, n otherwise).

We consider for application only one algorithm, the well-known (1+1) EA given in Algorithm 1 below.

Algorithm 1: The (1+1) EA for maximizing function f

```

1 choose  $x$  from  $\{0, 1\}^n$  uniformly at random;
2 while optimum not reached do
3    $y \leftarrow x$ ;
4   for  $i = 1$  to  $n$  do
5      $\lfloor$  with probability  $1/n$ :  $y_i \leftarrow 1 - y_i$ ;
6   if  $f(y) \geq f(x)$  then  $x \leftarrow y$ ;
```

For any function f and $i \geq 0$, we let f^i denote the i -times self-composition of f (with f^0 being the identity).

2.1 Known Results for the (1+1) EA on LeadingOnes

We will use the following concentration result from [2], bounding the optimization time of the (1+1) EA on LEADINGONES.

Theorem 1 ([2, Theorem 7]). *For all $d \leq 2n^2$, the probability that the optimization time of the (1+1) EA on LEADINGONES deviates from its expectation of $(1/2)(n^2 - n)((1 + 1/(n - 1))^n - 1)$ by at least d , is at most $4 \exp(-d^2/(20e^2n^3))$.*

The following lemma collects some important and well-known results for the optimization process of the (1+1) EA on LEADINGONES.

Lemma 1. *Consider the (1+1) EA on LEADINGONES, let x_t denote its search point at time t and $X_t = n - \text{LEADINGONES}(x_t)$ the fitness distance. Then*

- (a) $E(X_t - X_{t+1} \mid X_t) = (2 - 2^{1-X_t})(1 - 1/n)^{n-X_t}/n$
- (b) $\Pr(X_{t+1} \neq X_t \mid X_t; T > t) = (1 - 1/n)^{n-X_t} \frac{1}{n}$
- (c) For $j \geq 1$, $\Pr(X_{t+1} = X_t - j) \leq \frac{1}{n} \left(\frac{1}{2}\right)^{j-1}$
- (d) $G_t := X_t - X_{t+1}$ is a random variable with support $0, \dots, X_t$ and the following conditional distribution on $G_t \geq 1$:
 - $\Pr(G_t = i) = (1/2)^i$ for $i < X_t$
 - $\Pr(G_t = X_t) = (1/2)^{X_t-1}$

For the moment-generating function of this G_t (conditional on $G_t \geq 1$) it holds that

$$E(e^{\eta G_t} \mid X_t) = \frac{(e^\eta/2)^{X_t}(1 - e^\eta) + (e^\eta/2)}{1 - e^\eta/2}.$$

- (e) The expected optimization time equals $\frac{n^2-n}{2} \left(\left(1 + \frac{1}{n-1}\right)^n - 1 \right)$, which is $\frac{e-1}{2}n^2 \pm O(n)$.

Proof. The proofs of the first three statements can be found in in [2] and [13, Lemma 12 of technical report]. For the first part of the fourth statement, we recall from these papers that the $X_t - 1$ bits after the first 0 are uniform and independent. Hence, the probability of observing $i - 1 < X_t$ of these so-called free-riders is $(1/2)^i$ since $i - 1$ bits have to be set to 1 and the i -th bit to 0. If $i = X_t - 1$ then all i bits have to be set to 1, which has probability $(1/2)^i$.

For the moment-generating function, we write (using the first part of the fourth statement)

$$E(e^{\eta G_t} \mid X_t) = \sum_{j=1}^{X_t-1} \left(\frac{1}{2}\right)^j e^{\eta j} + \left(\frac{1}{2}\right)^{X_t-1} e^{\eta X_t}$$

Since, by the geometric series,

$$\sum_{j=1}^{X_t-1} \left(\frac{1}{2}\right)^j e^{\eta j} = \frac{e^\eta/2 - (e^\eta/2)^{X_t}}{1 - e^\eta/2},$$

we have

$$\begin{aligned} \mathbb{E}(e^{\eta G_t} \mid X_t) &= \frac{e^\eta/2 - (e^\eta/2)^{X_t}}{1 - e^\eta/2} + 2 \left(\frac{e^\eta}{2} \right)^{X_t} \\ &= \frac{e^\eta/2 - (e^\eta/2)^{X_t} + 2(e^\eta/2)^{X_t}(1 - e^\eta/2)}{1 - e^\eta/2} \\ &= \frac{(e^\eta/2)^{X_t}(1 - e^\eta) + (e^\eta/2)}{1 - e^\eta/2}. \end{aligned}$$

The fifth statement is due to [1]. □

3 Direct Fixed-Budget Drift Theorems

In this section we give a drift theorem which gives a fixed-budget result without the detour via first hitting times. The idea is to focus on drift which gets monotonically weaker as we approach the optimum, but where being closer to the optimum is still better in terms of drift. To this end, we make the following definition.

Definition 1. *We say that a drift function $h: S \rightarrow \mathbb{R}^{>0}$ is greed-admitting if $\text{id} - h$ (the function $x \mapsto x - h(x)$) is monotone non-decreasing.*

Intuitively, this formalizes the idea that being closer to the goal is always better (i. e. greed is good). Greed could be bad, if from one part of the search space, the drift is much higher than when being a bit closer, so that being a bit closer does not balance out the loss in drift. Note that any given differentiable h is greed-admitting if and only if $h' \leq 1$.

Typical drift functions are greed-admitting. For example, if we drift on integers, in many situations drift is less than 1, while being closer means being at least one step closer, so being closer is always better in this sense. An example monotone process on $\{0, 1, 2\}$ which has a drift which is not greed-admitting is the following: X_0 is 2 and the process moves to any of the states 0, 1, 2 uniformly. State 0 is the target state, from state 1 there is only a very small probability to progress to 0 (say 0.1). Then it is better to stay in state 2 than be trapped in state 1, if the goal is to progress to state 0.

We now give two different versions of the direct fixed-budget drift theorem. The first considers *unlimited time*, that is, the situation where drift carries on for an arbitrary time (and does not stop once a certain threshold value is reached). This is applicable in situations where there is no end to the process (for example for random walks on the line) or when the drift eventually goes all the way down to 0 so that the drift condition holds vacuously even when no progress is possibly any more (this is for example the case for multiplicative drift, where the drift is δ times the current value, which is naturally 0 once 0 has been reached). Note that this is a very strong requirement of the theorem, leading to a strong conclusion.

A special case of the following theorem is given in [15], where drift is necessarily multiplicative.

Theorem 2 (Direct Fixed-Budget Drift, unlimited time). *Let $X_t, t \geq 0$, be a stochastic process on $S \subseteq \mathbb{R}$, adapted to a filtration \mathcal{F}_t . Let $h: S \rightarrow \mathbb{R}^{\geq 0}$ be a **convex and greed-admitting** function such that we have the drift condition*

$$\text{(D-ut)} \quad \mathbb{E}(X_t - X_{t+1} \mid \mathcal{F}_t) \geq h(X_t).$$

Define $\tilde{h}(x) = x - h(x)$. Thus, the drift condition is equivalent to

$$\text{(D-ut')} \quad \mathbb{E}(X_{t+1} \mid \mathcal{F}_t) \leq \tilde{h}(X_t).$$

We have that, for all $t \geq 0$,¹

$$\mathbb{E}(X_t \mid \mathcal{F}_0) \leq \tilde{h}^t(X_0)$$

and, in particular,

$$\mathbb{E}(X_t) \leq \tilde{h}^t(\mathbb{E}(X_0)).$$

Proof. Note that \tilde{h} is concave, since the second derivative of $\text{id} - h$ is $-h''$. We have, using this concavity of for Jensen's Inequality, for all t ,

$$\begin{aligned} \mathbb{E}(X_{t+1} \mid \mathcal{F}_0) &= \mathbb{E}(\mathbb{E}(X_{t+1} \mid \mathcal{F}_t) \mid \mathcal{F}_0) \\ &\leq \mathbb{E}(\mathbb{E}(\tilde{h}(X_t) \mid \mathcal{F}_t) \mid \mathcal{F}_0) \\ &= \mathbb{E}(\tilde{h}(X_t) \mid \mathcal{F}_0) \\ &\leq \tilde{h}(\mathbb{E}(X_t \mid \mathcal{F}_0)). \end{aligned}$$

Thus, the claim follows by induction with \tilde{h} being non-decreasing (since h is greed-admitting). The second statement of the theorem follows with Jensen's Inequality. \square

Now we get to the second version of the theorem, considering the more frequent case where no guarantee on the drift can be given once the optimum has been found. This weaker requirement leads to a weaker conclusion.

Theorem 3 (Direct Fixed-Budget Drift, limited time). *Let $X_t, t \geq 0$, be a stochastic process on $S \subseteq \mathbb{R}$, $0 = \min S$, adapted to a filtration \mathcal{F}_t . Let $T := \min\{t \geq 0 \mid X_t = 0\}$ and $h: S \rightarrow \mathbb{R}^{\geq 0}$ be a **differentiable, convex and greed-admitting** function such that $\tilde{h}'(0) \in]0, 1]$ and we have the drift condition*

$$\text{(D-lt)} \quad \mathbb{E}(X_t - X_{t+1} \mid \mathcal{F}_t; t < T) \geq h(X_t).$$

Define $\tilde{h}(x) = x - h(x)$. Thus, the drift condition is equivalent to

$$\text{(D-lt')} \quad \mathbb{E}(X_{t+1} \mid \mathcal{F}_t; t < T) \leq \tilde{h}(X_t).$$

We have that, for all $t \geq 0$,

$$\mathbb{E}(X_t \mid \mathcal{F}_0) \leq \tilde{h}^t(X_0) + \frac{\tilde{h}(0)}{\tilde{h}'(0)}$$

and, in particular,

$$\mathbb{E}(X_t) \leq \tilde{h}^t(\mathbb{E}(X_0)) - \frac{\tilde{h}(0)}{\tilde{h}'(0)} \cdot \Pr(t \geq T \mid \mathcal{F}_0).$$

¹Recall from the preliminaries that f^i is the i -times self-composition of a function f .

Proof. We let $m = -\frac{\tilde{h}(0)}{\tilde{h}'(0)}$. Recall that we assume that $0 < \tilde{h}'(0) \leq 1$. We now define a new process which mimics $(X_t)_t$, but which has to make one additional step down after reaching 0. In order to have 0 be the target of this new process, we will shift the old process accordingly. We let $Y_0 = X_0 + m$ and, for all $t \geq 0$,

$$Y_{t+1} = \begin{cases} X_{t+1} + m, & \text{if } t + 1 < T; \\ 0, & \text{else if } Y_t = 0; \\ 0, & \text{else, with probability } \tilde{h}'(0); \\ m, & \text{otherwise.} \end{cases}$$

Intuitively, Y_t behaves like $X_t + m$, but once X_t hits the optimum, it will stay at m until, with probability $\tilde{h}'(0)$, it jumps to 0. We now give a drift function for this process in order to apply Theorem 2.

Define h_0 such that, for all $x \geq 0$,

$$h_0(x) = \begin{cases} x \cdot \tilde{h}'(0), & \text{if } x < m; \\ h(x - m), & \text{otherwise.} \end{cases}$$

To see that h_0 is convex, note that it is convex on both the parts less than m and above m ; furthermore, the left- and right-derivative in m coincide. Furthermore, h_0 is greed-admitting since it is differentiable with derivative $\tilde{h}'(0) \leq 1$ for all $x \leq m$ and with derivative at most 1 for $x > m$ from h being greed-admitting. We see that $\mathbb{E}(Y_t - Y_{t+1} | \mathcal{F}_t) \geq h_0(X_t)$, either by the corresponding statement about $(X_t)_t$ and h or by the drift in case of $Y_t = m$ being $m \cdot \tilde{h}'(0) = -\tilde{h}(0) = h(0) = h_0(m)$. Thus, we can apply Theorem 2 and get $\mathbb{E}(Y_t | \mathcal{F}_0) \leq \tilde{h}_0^t(Y_0)$. By induction we get $\tilde{h}_0^t(Y_0) = \tilde{h}_0^t(X_0 + m) = \tilde{h}^t(X_0) + m$. From $X_t \leq Y_t - m \cdot \mathbb{1}[t < T]$ we thus get

$$\begin{aligned} \mathbb{E}(X_t | \mathcal{F}_0) &\leq \mathbb{E}(Y_t | \mathcal{F}_0) - m \cdot \Pr(t < T | \mathcal{F}_0) \\ &\leq \tilde{h}_0^t(Y_0) - m \cdot \Pr(t < T | \mathcal{F}_0) \\ &= \tilde{h}^t(X_0) + m - m \cdot \Pr(t < T | \mathcal{F}_0) \\ &= \tilde{h}^t(X_0) + m \cdot \Pr(t \geq T | \mathcal{F}_0). \end{aligned}$$

This concludes the proof. □

With the following theorem we give a general way of iterating a greed-admitting function, as necessary for the application of the previous two theorems. From this we can see the similarity of this approach to the method of variable drift theory where the inverse of h is integrated over, see Theorem 7 and the discussion about drift theory in general in [14].

Theorem 4. *Let h be greed-admitting and let $\tilde{h} = \text{id} - h$. Then we have, for all starting points n and all target points $m < n$ and all time budgets t ,*

$$\text{if } t \geq \sum_{i=m}^{n-1} \frac{1}{h(i)} \text{ then } \tilde{h}^t(n) \leq m.$$

Proof. The idea of this proof is that each application of \tilde{h} on some value $\geq i$ gains at least $h(i)$, so gaining this amount at least $1/h(i)$ times decreases a value of at most $i+1$ to a value of at most i . A simple induction then gives the claimed result. More formally, for all i and all $t \geq 1/h(i)$, we have that $\tilde{h}^t(i+1) \leq i$. Thus we inductively get, for all k , if $t \geq \sum_{i=n-k}^{n-1} 1/h(i)$, then $\tilde{h}^t(n) \leq n-k$. Using the induction statement for $k = n-m$ gives the result. \square

3.1 Application to OneMax

In this section we show how we can apply Theorem 2 by using the optimization of the (1+1) EA on ONEMAX as an example (where we have multiplicative drift).

Theorem 5. *Let V_t be the number of 1s which the (1+1) EA on ONEMAX has found after t iterations of the algorithm. Then we have, for all t ,*

$$\mathbb{E}(V_t) \geq \begin{cases} \frac{n}{2} + \frac{t}{2\sqrt{e}} - O(1), & \text{if } t = O(\sqrt{n}); \\ \frac{n}{2} + \frac{t}{2\sqrt{e}}(1 - o(1)), & \text{if } t = o(n). \end{cases}$$

Furthermore, for all t , we have $\mathbb{E}(V_t) \geq n(1 - \exp(-t/(en)))/2$.

Proof. We can apply the unlimited time theorem (Theorem 2) to the (1+1) EA on ONEMAX by using the drift function $h(x) = (1 - 1/n)^{n-x} \frac{x}{n}$. This function is convex and greed-admitting, and it also applies in case the process already reached the optimum of 0 (since $h(0) = 0$). We now need to estimate \tilde{h}^t .

In order to apply Theorem 4, we estimate as follows (using estimates for the harmonic sum which use $c = o(n)$).

$$\begin{aligned} \sum_{i=n/2-c}^{n/2} \frac{1}{h(i)} &= \sum_{i=n/2-c}^{n/2} (1 - 1/n)^{i-n} \frac{n}{i} \\ &\leq n \sum_{i=n/2-c}^{n/2} (1 - 1/n)^{-c-n/2} \frac{1}{i} \\ &\leq n(1 - 1/n)^{-c-n/2} \sum_{i=n/2-c}^{n/2} \frac{1}{i} \\ &\leq n \exp((c + n/2)/n) (\ln(n/2) - \ln(n/2 - c) + O(1/n)) \\ &= n\sqrt{e}e^{c/n} (-\ln((n/2 - c)/(n/2)) + O(1/n)) \\ &= n\sqrt{e}e^{c/n} (-\ln(1 - 2c/n) + O(1/n)) \\ &\leq n\sqrt{e}(1 + c/n + c^2/n^2) (2c/n + O(1/n)) \\ &= 2\sqrt{e}(1 + c/n + c^2/n^2) (c + O(1)). \end{aligned}$$

For $c = \sqrt{n}$ the last term is at most $2\sqrt{e}c + O(1)$ and for $c = o(n)$ it is $2\sqrt{e}c(1 + o(1))$. Thus, we get the claimed bounds with Theorem 4. Regarding the ‘‘furthermore’’ clause,

we argue more directly about \tilde{h}^t by observing that, for all x , $\tilde{h}(x) \leq (1 - 1/(en))x$ and thus, by a straightforward induction (similar to the proof of [15, Theorem 1]) we get, for all x, t , $\tilde{h}^t(x) \leq (1 - 1/(en))^t x \leq e^{-t/(en)}x$. This gives the desired result with initial state $n/2$. \square

3.2 Application to LeadingOnes

In this section we want to use Theorem 3 to the progress of the (1+1) EA on LEADING-ONES. The result is summarized in the following theorem.

Theorem 6. *Let V_t be the number of leading 1s which the (1+1) EA on LEADINGONES has found after t iterations of the algorithm. We have, for all t ,*

$$E(V_t) \geq \begin{cases} \frac{2t}{n} - O(1), & \text{if } t = O(n^{3/2}); \\ \frac{2t}{n} \cdot (1 - o(1)), & \text{if } t = o(n^2); \\ n \ln(1 + \frac{2t}{n^2}) - O(1), & \text{if } t \leq \frac{e-1}{2}n^2 - n^{3/2}. \end{cases}$$

Proof. For the derivation of fitness drift of the (1+1) EA on LEADINGONES, see the first item of Lemma 1. We want to use Theorem 3 to get our fixed-budget result.

However, in order to make our analysis, we artificially change the fitness value of the all-1s string to $n + 1$ (rather than n). The result of this change is in the expected fitness gain: if any fitness is gained at all, the total gain is usually (for plain LEADINGONES) 1 plus the number of “free rider” bits, additional bits after the first that happen to be set to 1. There cannot be an arbitrary number of them (since the bit string is finite – of size n), so the total expected number of bits gained is slightly less than 2: it is $2 - (1/2)^{n-1-x}$. By artificially changing the fitness value of the perfect string to $n + 1$ we now have an expected value increase of at least 2, as long as the best bit string has not been found (conditional on making an improvement at all).

Note that this change of the fitness value of the all-1s string changes the final result only by at most 1, which is consumed by the O -notation.

Thus, we can use the drift function

$$h: [0, n] \rightarrow \mathbb{R}^{>0}, x \mapsto \left(1 - \frac{1}{n}\right)^{n-x} \frac{2}{n}.$$

We have that h is greed-admitting (since the drift changes only very little, it would have to change by more than 1 between two distance 1 states) and convex (since the exponential function is convex). Note that without the artificial change mentioned above the actual drift would not have been convex.

We let $\tilde{h}(x) = x - h(x)$. We want to apply Theorems 3 so we note that

$$\tilde{h}(0) = - \left(1 - \frac{1}{n}\right)^n \frac{2}{n} = -\Theta(1/n)$$

and

$$\tilde{h}'(0) = 1 + \left(1 - \frac{1}{n}\right)^n \frac{2}{n} \ln(1 - 1/n) = 1 - o(1).$$

In order to estimate the t -fold application of \tilde{h} on 0 we use Theorem 4. Let $c = n - m$. We have

$$\begin{aligned}
\sum_{i=m}^{n-1} \frac{1}{h(i)} &= \sum_{i=m}^{n-1} \left(1 - \frac{1}{n}\right)^{i-n} \frac{n}{2} \\
&= \frac{n}{2} \sum_{j=1}^c \left(1 - \frac{1}{n}\right)^{-j} \\
&= \frac{n}{2} \left(\frac{\left(1 - \frac{1}{n}\right)^{-c-1} - 1}{\frac{1}{1 - 1/n} - 1} - 1 \right) \\
&= \frac{n(n-1)}{2} \left(\left(1 - \frac{1}{n}\right)^{-c-1} - 1 - \frac{1}{n-1} \right) \\
&\leq \frac{n(n-1)}{2} \left(\exp\left(\frac{c+1}{n}\right) - 1 - \frac{1}{n-1} \right).
\end{aligned}$$

From this we already get the third and most general claimed bound using the concentration bound given in Theorem 1 with an appropriate $d = \Theta(n^{3/2})$, where the probability of having reached the optimum is some constant.

We can continue the estimates as

$$\begin{aligned}
&\leq \frac{n(n-1)}{2} \left(\frac{c+1}{n} + \left(\frac{c+1}{n}\right)^2 - \frac{1}{n-1} \right) \\
&\leq \frac{n-1}{2} \left(c+1 + \frac{(c+1)^2}{n} - \frac{1}{n-1} \right).
\end{aligned}$$

This term is at most $(c+1)n/2 + o(n)$ for $c = o(\sqrt{n})$; and $cn/2 + O(n)$ for $c = \Theta(\sqrt{n})$ and $(1+o(1))cn/2$ for $c = o(n)$.

We use Theorem 1 again to see that the (1+1) EA on LEADINGONES is done in $o(n^2)$ steps with probability at most $\exp(-n)$, which suffices to get the first two desired bounds with the help of Theorems 3 and 4. \square

4 Variable Drift Theorem for Fixed Budget

We now turn to an alternative approach to derive fixed-budget results via drift analysis. Our method is based on variable drift analysis that was introduced to the analysis of randomized search heuristics by Johannsen [11]. Crucially, variable drift analysis applies a specific transformation, the so-called potential function g , to the state space. Along with bounds on the hitting times, we obtain the following theorem estimating the expected value of the potential function after t steps. Subsequently, we will discuss how this information can be used to analyze the untransformed state.

Theorem 7. *Let X_t , $t \geq 0$, be a stochastic process, adapted to a filtration \mathcal{F}_t , on $S := \{0\} \cup \mathbb{R}^{\geq x_{\min}}$ for some $x_{\min} > 0$. Let $T := \min\{t \geq 0 \mid X_t = 0\}$ and $h: S \rightarrow \mathbb{R}^{>0}$ be*

a non-decreasing function such that $\mathbb{E}(X_t - X_{t+1} \mid \mathcal{F}_t; t < T) \geq h(X_t)$. Define $g: S \rightarrow \mathbb{R}$ by

$$g(x) := \begin{cases} \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^x \frac{1}{h(z)} dz & \text{if } x \geq x_{\min} \\ 0 & \text{otherwise} \end{cases}.$$

Then it holds that

$$\mathbb{E}(g(X_t) \mid \mathcal{F}_0) \leq g(X_0) - \sum_{s=0}^{t-1} \Pr(s < T).$$

Proof. Since h is non-decreasing, g is concave. We claim that the drift of the g -value is bounded from below by 1, formally

$$\mathbb{E}(g(X_t) - g(X_{t+1}) \mid \mathcal{F}_t; t < T) \geq 1 \quad (1)$$

To prove the claim, we use standard arguments from the proof of the variable drift theorem for expected hitting times. Expanding the definition of g , we obtain

$$\begin{aligned} \mathbb{E}(g(X_t) - g(X_{t+1}) \mid \mathcal{F}_t; t < T) &= \int_{x_{\min}}^{X_t} \frac{1}{h(z)} dz \\ &\quad - \mathbb{E}\left(\int_{x_{\min}}^{X_{t+1}} \frac{1}{h(z)} dz \mid \mathcal{F}_t\right). \end{aligned}$$

By Jensen's inequality and the concavity of g , we have

$$\mathbb{E}(g(X_t) - g(X_{t+1}) \mid \mathcal{F}_t; t < T) \geq \int_{x_{\min}}^{X_t} \frac{1}{h(z)} dz - \int_{x_{\min}}^{\mathbb{E}(X_{t+1} \mid \mathcal{F}_t)} \frac{1}{h(z)} dz,$$

which, since $\mathbb{E}(X_{t+1} \mid \mathcal{F}_t; t < T) \leq X_t - h(X_t)$, is at least

$$\int_{X_t - h(X_t)}^{X_t} \frac{1}{h(z)} dz \geq \int_{X_t - h(X_t)}^{X_t} \frac{1}{h(X_t)} dz = 1,$$

where the inequality used that $h(z)$ is non-decreasing.

We proceed by estimating $\mathbb{E}(g(X_t))$ in an inductive fashion. By the law of total probability,

$$\mathbb{E}(g(X_1) \mid \mathcal{F}_0) = g(X_0) - \Pr(0 < T) (g(X_0) - \mathbb{E}(g(X_1) \mid \mathcal{F}_0; 0 < T))$$

so with (1),

$$\mathbb{E}(g(X_1) \mid \mathcal{F}_0) \leq g(X_0) - \Pr(T > 0).$$

Noting that

$$\begin{aligned} &\mathbb{E}(g(X_t) \mid \mathcal{F}_0) \\ &= \mathbb{E}(g(X_{t-1}) - \mathbb{E}(g(X_{t-1}) - g(X_t) \mid \mathcal{F}_{t-1}; t-1 < T) \mid \mathcal{F}_0), \end{aligned}$$

we get by the induction hypothesis and (1) that

$$\begin{aligned} \mathbb{E}(g(X_t) \mid \mathcal{F}_0) &\leq \mathbb{E}(g(X_{t-1}) \mid \mathcal{F}_0) - \Pr(t-1 < T) \\ &\leq g(X_0) - \sum_{s=0}^{t-2} \Pr(s < T) - \Pr(t-1 < T) \end{aligned}$$

altogether

$$\mathbb{E}(g(X_t) \mid \mathcal{F}_0) \leq g(X_0) - \sum_{s=0}^{t-1} \Pr(s < T).$$

as suggested. \square

4.1 Additive Drift as Special Case

A special case of variable drift is additive drift, when the drift function h is constant.

Theorem 8. *Let X_t , $t \geq 0$, be a stochastic process, adapted to a filtration \mathcal{F}_t , on $S := \mathbb{R}^{\geq 0}$. Let $T := \min\{t \geq 0 \mid X_t = 0\}$ and $\delta \in \mathbb{R}^{>0}$ be such that $\mathbb{E}(X_t - X_{t+1} \mid \mathcal{F}_t; t < T) \geq \delta$. Then we have*

$$\mathbb{E}(X_t \mid \mathcal{F}_0) \leq X_0 - \delta \sum_{s=0}^{t-1} \Pr(s < T).$$

The theorem is a corollary to Theorem 7 by using $x_{\min} = \delta$, the smallest value for which the condition of a drift of at least δ can still be obtained, and thus the smallest value (other than 0) that the process can attain.

As a sample application, we can now derive an estimate of the best value found by the (1+1) EA on LEADINGONES within t steps, using the concentration result from [2] given in Theorem 1.

Theorem 9. *Let V_t be the number of leading 1s which the (1+1) EA on LEADINGONES has found after t iterations of the algorithm. Then, for all $t \leq \frac{e-1}{2}n^2 - n^{3/2} \log(n)$, we have*

$$\mathbb{E}(V_t) \geq \frac{2t}{en} - O(1).$$

Proof. We drift on the potential which assigns each bit string its number of leading ones, except for the all-1 string which has a potential of $n+1$. A quick computation shows that this leads to an expected increase in potential of 2, conditional on the potential increasing at all (without the “+1” for the all-1-string, it would have been slightly less than 2). We use drift on this potential, which is, for all current potential values $x < n$, now lower bounded by

$$\left(1 - \frac{1}{n}\right)^x \frac{2}{n} \geq \frac{2}{en}.$$

Thus, the result follows with Theorem 8 and the concentration bound given in Theorem 1. \square

Note that the result was proven very easily with a direct application of the additive version of the fixed-budget drift theorem in combination with a strong result on concentration. The price paid for this simplicity is that the lead constant in this time bound is not tight, as can be seen by comparing with the results given in Theorem 6.

5 Variable Drift and Concentration Inequalities

The expected $g(X_t)$ -value derived in Theorem 7 is not very useful unless it allows us to make conclusions on the underlying X_t -value. The previous application in Section 4.1 only gives tight bounds in case that the drift is more or less constant throughout the search space. This is not the case for ONEMAX and LEADINGONES where the drift increases with the distance to the optimum (e.g., for ONEMAX the drift is $\Theta(1/n)$ at distance 1 and $\Theta(1)$ as distance $n/2$; for LEADINGONES the drift can vary by a term of roughly e). Hence, looking back into Theorem 7, we now are interested in characterizing $g(X_t)$ more precisely than just in terms of expected value. If we manage to establish concentration of $g(X_t)$ then we can (after inverting g) derive a maximum of the X_t -value that holds with sufficient probability. Our main result achieved along this path is the following one.

Theorem 10. *Let V_t be the number of leading 1s which the $(1+1)$ EA on LEADINGONES has found after t iterations. Then for $t = \omega(n \log n)$ and $t \leq (e-1)n^2/2 - cn^{3/2}\sqrt{\log n}$, where c is a sufficiently large constant the following statements hold. (a) With probability at least $1 - 1/n^3$,*

$$\begin{aligned} -n \ln\left(1 - 2t/n^2 + O(\sqrt{t \log n}/n^{3/2})\right) &\leq V_t \\ -n \ln\left(1 - 2t/n^2 - O(\sqrt{t \log n}/n^{3/2})\right) &\geq V_t. \end{aligned}$$

(b) $E(V_t) = -n \ln(1 - 2t/n^2 + O(\sqrt{t \log n}/n^{3/2}))$.

To compare with previous work, we note that the additive error is $O(\sqrt{t \log n}/n^{1/2})$. This is asymptotically smaller than the additive error term of order $\Omega(n^{3/2+\varepsilon})$ that appears in the fixed-budget statements of [2] and moreover, it depends on t . Also, we think that the formulation of our statement is less complex than in that paper.

The proof of Theorem 10 overcomes several technical challenges. The first idea is to apply established concentration inequalities for stochastic processes. Since (after a reformulation discussed below) the process of g -values describes a (super)martingale, it is natural to take the method of bounded martingale differences. However, since there is no ready-to-use theorem for all our specific martingales, we present a generalization of martingale concentration inequalities in the following subsection Section 5.1. The concrete application is then given in Sections 5.2 onwards.

5.1 Tail Bounds for Martingale Differences

The classical method of bounded martingale differences [16] considers a (super)martingale Y_t , $t \geq 0$, and its corresponding martingale differences $D_t = Y_{t+1} - Y_t$. Given certain

boundedness conditions for D_t (e. g., that $|D_t| \leq c$ for a constant c almost surely), it is shown that the sum of martingale differences $\sum_{i=0}^{t-1} D_i = Y_t - Y_0$ does not deviate much from its expectation Y_0 (resp. is not much bigger in the case of supermartingales). This statement remains essentially true if D_t is allowed to have unbounded support but exhibits a strong concentration around its expected value. Usually, this concentration is formulated in terms of a so-called *subgaussian* (or, similarly, *subexponential*) property [4, 12]. Roughly speaking, this property requires that the moment-generating function (mgf.) of the differences can be bounded as $\mathbb{E}(e^{\lambda D_t} | \mathcal{F}_t) \leq e^{\lambda^2 \nu_t^2 / 2}$ for a certain parameter ν_t and all $\lambda < 1/b_t$, where b_t is another parameter. In particular, the bound has to remain true when λ becomes arbitrarily small.

In one of our concrete applications of the martingale difference technique, the inequality $\mathbb{E}(e^{\lambda D_t} | \mathcal{F}_t) \leq e^{\lambda^2 \nu_t^2 / 2}$ is true for certain values of λ below a threshold $1/b^*$, but does not hold if λ is much smaller than $1/b^*$. We therefore show that the concentration of the sums of martingale differences to some extent remains true if the inequality only holds for $\lambda \in [1/a^*, 1/b^*]$ where $a^* > b^*$ is another parameter. The approach uses well-known arguments for the proof of concentration inequalities. Here, we were inspired by the notes [18], which require the classical subexponential property, though.

Theorem 11. *Let Y_t , $t \geq 0$, be a supermartingale, adapted to a filtration \mathcal{F}_t , and let $D_t = Y_{t+1} - Y_t$ be the corresponding martingale differences. Assume that there are $0 < b_2 < b_1 \leq \infty$ and a sequence ν_t , $t \geq 0$, such that for $\lambda \in [1/b_1, 1/b_2]$ it holds that $\mathbb{E}(e^{\lambda D_t} | \mathcal{F}_t) \leq e^{\lambda^2 \nu_t^2 / 2}$. Then for all $t \geq 0$ it holds that*

$$\Pr(Y_t - Y_0 \geq d) \leq \begin{cases} e^{-d/(2b_2)} & \text{if } d \geq \frac{\sum_{i=0}^{t-1} \nu_i^2}{b_2} \\ e^{-d^2/(2\sum_{i=0}^{t-1} \nu_i^2)} & \text{if } \frac{\sum_{i=0}^{t-1} \nu_i^2}{b_1} \leq d < \frac{\sum_{i=0}^{t-1} \nu_i^2}{b_2} \end{cases}$$

The theorem holds analogously for submartingales with respect to the tail bound $\Pr(Y_t - Y_0 \leq -d)$.

Proof. We consider the mgf. of the sum $S_t := \sum_{i=0}^{t-1} D_i$. Using the usual Chernoff-type approach, we have for all $\lambda \geq 0$ that

$$\Pr(S_t \geq d) = \Pr(e^{\lambda S_t} \geq e^{\lambda d}) \leq e^{-\lambda d} \mathbb{E}(e^{\lambda S_t} | \mathcal{F}_0)$$

To bound the last mgf., we note that by the law of total expectation,

$$\begin{aligned} \mathbb{E}(e^{\lambda S_t} | \mathcal{F}_0) &= \mathbb{E}\left(e^{\lambda \sum_{i=0}^{t-2} D_i} \cdot \mathbb{E}(e^{\lambda D_{t-1}} | \mathcal{F}_{t-1}) | \mathcal{F}_0\right) \\ &\leq \mathbb{E}(e^{\lambda S_{t-1}} | \mathcal{F}_0) e^{\lambda^2 \nu_{t-1}^2 / 2}, \end{aligned}$$

where the last inequality used the assumption from the theorem, which is valid after we assume $1/b_1 \leq \lambda \leq 1/b_2$. Iterating this argument, we obtain

$$\mathbb{E}(e^{\lambda S_t} | \mathcal{F}_0) \leq e^{\lambda^2 \sum_{i=0}^{t-1} \nu_i^2 / 2},$$

hence

$$\Pr(S_t \geq d) \leq e^{-\lambda d} e^{\lambda^2 \sum_{i=0}^{t-1} \nu_i^2 / 2}.$$

In the following, we write $V_t := \sum_{i=0}^{t-1} \nu_i^2$. We now distinguish between the two cases for d displayed in the lemma. If $d \geq V_t/b_2$, this leads to

$$\Pr(S_t \geq d) \leq e^{-\lambda d} e^{\lambda^2 db_2/2},$$

which, choosing $\lambda := 1/b_2$, yields

$$\Pr(S_t \geq d) \leq e^{-d/b_2 + (1/b_2)^2 db_2/2} = e^{-d/(2b_2)}.$$

If $V_t/b_1 \leq d < V_t/b_2$ we choose $\lambda := d/V_t \in [1/b_1, 1/b_2]$. Then

$$\Pr(S_t \geq d) \leq e^{-\lambda d + \lambda^2 V_t/2} = e^{-\frac{d}{V_t} d + \frac{d^2}{V_t^2} \frac{V_t}{2}} = e^{-d^2/(2V_t)},$$

which, after substituting V_t , proves the theorem. \square

5.2 Preparing an Upper Tail Bound via the Martingale Difference Method

We now return to Theorem 7 and would like to show concentration of $g(X_t)$ in order to show a bound for X_t that holds with sufficiently high probability. Note that by the statement of the theorem, we immediately have that $Y_t := g(X_t) + \sum_{s=0}^{t-1} \Pr(T > s)$ is a supermartingale. By bounding the probability of $Y_t \geq d$ for arbitrary $t \geq 0$ and $d \geq 0$, i. e., establishing concentration of the supermartingale Y_t via Theorem 11, and inverting g , we will obtain a bound on the probability of the event $g(X_t) \geq \mathbb{E}(g(X_t))$.

As we want to prove Theorem 10, the application is again the (1+1) EA on the LEADINGONES function, so $X_t = n - \text{LEADINGONES}(x_t)$ is the fitness distance of the LEADINGONES-value at time t from the target.

Defining $h(X_t) := \mathbb{E}(X_t - X_{t+1} \mid X_t)$ according to Lemma 1 and $g(X_t) = 1/h(1) + \int_1^{X_t} 1/h(z) dz$ according to Lemma 7, we will establish the following bound on the moment-generating function (mgf.) of the drift of our concrete g .

Lemma 2. *Let T denote the optimization time of the (1+1) EA on LEADINGONES. If $\lambda \leq 1/(2en)$ then $\mathbb{E}(e^{\lambda(g(X_{t+1}) - g(X_t) + \Pr(T > t))} \mid X_t) = e^{O(\lambda^2 n)}$.*

Proof. We write $Y_t = g(X_t) + \sum_{s=0}^{t-1} \Pr(T > s)$ and $D_t = Y_{t+1} - Y_t = g(X_{t+1}) - g(X_t) + \Pr(T > t)$. Without loss of generality the process stops from time T on so that $g(X_{t+1}) - g(X_t) = 0$ for all $t \geq T$. Hence, conditional on $T \leq t$ we have $\mathbb{E}(e^{\lambda D_t}) = e^{\lambda 0 + 0} = 1 \leq e^{\lambda^2 n}$. We now consider the interesting case that $T > t$. Then

$$\mathbb{E}(e^{\lambda D_t} \mid X_t) = \mathbb{E}(e^{\lambda(g(X_{t+1}) - g(X_t) + 1)} \mid X_t)$$

since $\Pr(T > t) = 1$ on our condition. To bound this mgf., we shall exploit that

$$g(X_{t+1}) - g(X_t) \leq \frac{X_{t+1} - X_t}{h(X_t)}$$

(by the concavity of g , noting that the difference is negative). Hence, $g(X_{t+1}) - g(X_t)$ is stochastically dominated by $\frac{X_{t+1} - X_t}{h(X_t)}$.

Applying the law of total probability with respect to an improving step (using the second item from Lemma 1) and writing $G_t = X_t - X_{t+1}$, we obtain

$$\begin{aligned} \mathbb{E}(e^{\lambda(g(X_{t+1}) - g(X_t) + 1)} \mid X_t) &\leq \left(1 - \frac{1}{n}\right)^{n - X_t} \frac{1}{n} \mathbb{E}(e^{\frac{\lambda}{h(X_t)}(-G_t) + \lambda} \mid X_t) \\ &\quad + \left(1 - \left(1 - \frac{1}{n}\right)^{n - X_t} \frac{1}{n}\right) e^\lambda \end{aligned} \quad (2)$$

We will write $\eta = \frac{\lambda}{h(X_t)}$ in the following and assume $0 \leq \eta = o(1)$, which implies the same for λ . Using the well-known inequalities $1 + x + x^2 \geq e^x \geq 1 + x$ and $1 - x + x^2 \geq e^{-x} \geq 1 - x$ for $0 \leq x \leq 1$, we obtain the following bound on (2):

$$\begin{aligned} &\mathbb{E}(e^{\lambda(g(X_{t+1}) - g(X_t) + 1)} \mid X_t) \\ &\leq \left(1 - \frac{1}{n}\right)^{n - X_t} \frac{1}{n} (1 + \lambda + \lambda^2) \mathbb{E}(e^{-\eta G_t} \mid X_t) \\ &\quad + \left(1 - \left(1 - \frac{1}{n}\right)^{n - X_t} \frac{1}{n}\right) (1 + \lambda + \lambda^2). \end{aligned} \quad (3)$$

The most challenging part is now to bound $\mathbb{E}(e^{-\eta G_t} \mid X_t)$ (still conditional on $G_t \geq 1$) in such a way that we obtain an estimate depending on λ only. We use the shorthand $X_t = i$ from now on. From Lemma 1 and again using the inequalities with e^x we have

$$\begin{aligned} \mathbb{E}(e^{-\eta G_t} \mid X_t; G_t \geq 1) &= \frac{(e^{-\eta/2})^i (1 - e^{-\eta}) + (e^{-\eta/2})}{1 - e^{-\eta/2}} \\ &\leq \frac{(\frac{1-\eta+\eta^2}{2})^i (1 - (1 - \eta)) + \frac{1-\eta+\eta^2}{2}}{1 - (1/2)(1 - \eta + \eta^2)} \\ &= \frac{(\frac{1-\eta+\eta^2}{2})^i \eta + \frac{1-\eta+\eta^2}{2}}{(1/2) + \eta/2 - \eta^2/2} \\ &\leq \frac{(\frac{1}{2})^{i-1} \eta + 1 - \eta + \eta^2}{1 + \eta - \eta^2} \\ &\leq \left(\frac{1}{2}\right)^{i-1} \eta + \frac{1 - \eta + \eta^2}{1 + \eta - \eta^2} \\ &= \left(\frac{1}{2}\right)^{i-1} \eta + 1 - \frac{2\eta - 2\eta^2}{1 + \eta - \eta^2} \\ &= 1 + \left(\frac{1}{2}\right)^{i-1} \eta - 2\eta + O(\eta^2) \end{aligned}$$

We are now ready to substitute $h(i) = (1 - 1/n)^{n-i}(1/n)(2 - (\frac{1}{2})^{i-1})$ (Lemma 1) in $\eta = \lambda/h(i)$ and obtain from the preceding estimate that

$$\begin{aligned} \mathbb{E}(e^{-\eta G_t} \mid X_t; G_t \geq 1) &\leq 1 + \eta \left(\left(\frac{1}{2} \right)^{i-1} - 2 \right) + O(\eta^2) \\ &= 1 + \frac{\lambda n \left(1 - \frac{1}{n}\right)^{i-n}}{2 - \left(\frac{1}{2}\right)^{i-1}} \left(\left(\frac{1}{2} \right)^{i-1} - 2 \right) + O(\eta^2) \\ &= 1 - \lambda n \left(1 - \frac{1}{n}\right)^{i-n} + O(\lambda^2 n^2), \end{aligned}$$

where the last inequality used $1/h(i) = O(n)$. Plugging this back into (3), we have

$$\begin{aligned} &\mathbb{E}(e^{\lambda(g(X_{t+1}) - g(X_t) + 1)} \mid X_t = i) \\ &= \left(1 - \frac{1}{n}\right)^{n-i} \frac{1}{n} (1 + \lambda + \lambda^2) \left(1 - \lambda n \left(1 - \frac{1}{n}\right)^{i-n} + O(\lambda^2 n^2)\right) \\ &\quad + \left(1 - \left(1 - \frac{1}{n}\right)^{n-i} \frac{1}{n}\right) (1 + \lambda + \lambda^2) \\ &= 1 + \lambda + \lambda^2 - \lambda + O(\lambda^2 n) \leq e^{\lambda^2 + O(\lambda^2 n)} = e^{O(\lambda^2 n)} \end{aligned} \tag{4}$$

for $\lambda \leq 1/(2en)$. The restriction on λ follows from the fact that $\eta \leq 1$ has to be ensured, which follows for $\lambda \leq 1/(2en)$ from the bound $h(x) \leq 2en$. \square

Looking into Theorem 11 the required subexponential property of the martingale difference D_t has been proven with $\nu = O(\sqrt{n})$ and $\lambda \leq 1/(2en) = 1/b^*$. Before we formally apply this lemma, we also establish concentration in the other direction.

5.3 Preparing a Lower Tail Bound

We will now complement the upper tail bound for g that we prepared in the previous subsection with a lower tail bound. The aim is again to apply Theorem 11, this time with respect to the sequence $Y_t = g(X_t) + \sum_{s=0}^{t-1} \Pr(T > s) + r(t, n)$, where $X_t = n - \text{LEADINGONES}(x_t)$ is still the fitness distance of the LEADINGONES-value at time t from the target and $r(t, n)$ is an “error term” that we will prove to be $O(1/n)$ if $g(X_t) > \log n$. Moreover, $r(t, n) = 0$ if $g(X_t) = 0$. The first step is to prove that Y_t is a submartingale, i. e., $\mathbb{E}(Y_{t+1} \mid Y_t) \geq Y_t$. Afterwards, we bound the mgf. of $D_t = Y_t - Y_{t-1} = g(X_{t+1}) - g(X_t) + \Pr(T > t) + r(t, n)$.

Lemma 3. *The sequence $Y_t = g(X_t) + \sum_{s=0}^{t-1} \Pr(T > s) + r(t, n)$ is a submartingale with $r(t, n) = O(1/n)$ for $X_t > \log n$.*

Proof. We first note that nothing is to show if $g(X_t) = 0$. Hence, we assume $g(X_t) > 0$ in the following. From Section 5 we know that $\mathbb{E}(g(X_t) - g(X_{t+1}) \mid X_t) \geq 1$. Basically,

this holds since the function g is concave and the slope at point X_t equals $1/h(X_t)$, so that g scales the drift of the X_t -process at this point, which is $h(X_t)$, by a factor of $1/h(X_t)$. We will next show that the error incurred by estimated the drift of g via the slope $1/h(X_t)$ is small. The approach is similar to [7], who analyzed this kind of error with respect to the drift function of the (1+1) EA on ONEMAX.

We claim that if $X_t \geq \log n$ then

$$\mathbb{E}(g(X_t) - g(X_{t+1}) \mid X_t) \leq 1 + O(1/n).$$

To prove the claim, we show

$$\mathbb{E}(g(X_t) - g(X_{t+1}) \mid X_t) - \frac{\mathbb{E}(X_t - X_{t+1} \mid X_t)}{h(X_t)} = O(1/n)$$

instead. Substituting the definition of g and noting that $X_{t+1} \leq X_t$, this is equivalent to

$$\mathbb{E}\left(\int_{X_{t+1}}^{X_t} 1/h(i) \, di \mid X_t\right) - \frac{\mathbb{E}(X_t - X_{t+1} \mid X_t)}{h(X_t)}$$

and, due to the discrete state space, it is also equivalent to

$$\mathbb{E}\left(\sum_{i=X_{t+1}}^{X_t} \frac{1}{h(i)} \mid X_t\right) - \frac{\mathbb{E}(X_t - X_{t+1} \mid X_t)}{h(X_t)}.$$

By the definition of expectation, this difference equals

$$\begin{aligned} & \sum_{j=1}^{X_t} \Pr(X_{t+1} = X_t - j) \left(\sum_{k=X_t-j+1}^{X_t} \frac{1}{h(k)} - \frac{j}{h(X_t)} \right) \\ & \leq \sum_{j=1}^{X_t} \Pr(X_{t+1} = X_t - j) \left(\frac{j}{h(X_t - j + 1)} - \frac{j}{h(X_t)} \right), \end{aligned} \quad (5)$$

where we used that h is non-decreasing. After an index manipulation (formally, writing $j' = j - 1$), we are left with the task of bounding

$$\left(\frac{1}{h(X_t - j')} - \frac{1}{h(X_t)} \right)$$

for $j' \geq 1$ (since for $j' = 0$ the difference is 0). Writing the difference as

$$\frac{h(X_t) - h(X_t - j')}{h(X_t)h(X_t - j')}$$

and using the bounds

$$\frac{(1 - 1/n)^{n-i}}{2n} \leq h(i) = \frac{(1 - 1/n)^{n-i}}{(2 - (1/2)^{i-1})n} \leq \frac{n(1 + 1/n)(1 - 1/n)^{n-i}}{2n}$$

according to Lemma 1 for $i \geq \log n$, we obtain for $X_t \geq 2 \log n$ and $j' \leq \log n$ that

$$\begin{aligned} \frac{h(X_t) - h(X_t - j')}{h(X_t)h(X_t - j')} &\leq \frac{4n^2 \left(1 + \frac{1}{n}\right) \left(\left(1 - \frac{1}{n}\right)^{n-X_t} - \left(1 - \frac{1}{n}\right)^{n-X_t+j'} \right)}{2n \left(1 - \frac{1}{n}\right)^{n-X_t+n-X_t+j'}} \\ &= \frac{2n \left(1 + \frac{1}{n}\right) \left(1 - \frac{1}{n}\right)^{n-X_t} \left(1 - \left(1 - \frac{1}{n}\right)^{j'}\right)}{\left(1 - \frac{1}{n}\right)^{n-X_t+n-X_t+j'}} \\ &\leq 2e \left(1 + \frac{1}{n}\right)^2 j', \end{aligned}$$

where the last inequality used Bernoulli's inequality and the estimate $(1 - 1/n)^n \geq e^{-1}(1 - 1/n)$.

Together with the estimate $\Pr(X_{t+1} = X_t - j) \leq \frac{1}{n} \left(\frac{1}{2}\right)^{j-1}$ from Lemma 1 and recalling the index transformation, we bound (5) from above by

$$\sum_{j=1}^{X_t} \frac{1}{n} \left(\frac{1}{2}\right)^{j-1} j 2e(1 + 1/n)^2(j-1) = O(1/n)$$

as suggested. \square

Recall that the aim is to apply Theorem 11 with respect to the submartingale sequence $Y_t = g(X_t) + \sum_{s=0}^{t-1} \Pr(T > s) + r(t, n)$. To this end, we shall bound the mgf. of $D_t = Y_t - Y_{t-1} = g(X_{t+1}) - g(X_t) + \Pr(T > t) + r(t, n)$ in the following way.

Lemma 4. *The mgf. of $D_t = Y_t - Y_{t-1} = g(X_{t+1}) - g(X_t) + \Pr(T > t) + r(t, n)$ satisfies $\mathbb{E}(e^{\lambda D_t} | X_t) = e^{O(\lambda^2 n)}$ for all $\lambda \in [1/n^2, 1/(2en)]$.*

Proof. Without loss of generality the process stops from time T on so that $g(X_{t+1}) - g(X_t) = 0$ for all $t \geq T$. Hence, conditional on $T \leq t$ we have $\mathbb{E}(e^{\lambda D_t}) = e^{\lambda 0 + 0} = 1 \leq e^{\lambda^2 \nu^2 / 2}$. We now consider the interesting case that $T > t$. Then

$$\mathbb{E}(e^{\lambda D_t} | X_t) = \mathbb{E}(e^{\lambda(g(X_{t+1}) - g(X_t) + 1 + r(t, n))} | X_t)$$

since $\Pr(T > t) = 1$ on our condition. This mgf. differs from the one investigated in Lemma 2 and its proof only by the factor $e^{\lambda r(t, n)}$. Adjusting (4) accordingly and plugging in $r(n) = r(t, n) = O(1/n)$, we obtain

$$\begin{aligned} &\mathbb{E}(e^{\lambda D_t} | X_t = i) \tag{6} \\ &= (1 - 1/n)^{n-i} \frac{1}{n} (1 + \lambda(1 + r(n))) \\ &\quad + \lambda^2 (1 + r(n))^2 (1 - \lambda n (1 - 1/n)^{i-n} + O(\lambda^2 n^2)) \\ &\quad + \left(1 - (1 - 1/n)^{n-i} \frac{1}{n}\right) (1 + \lambda(1 + r(n)) + \lambda^2 (1 + r(n))^2) \\ &= 1 + \lambda + \lambda^2 + \lambda r(n) + \lambda^2 (2r(n) + r^2(n)) - \lambda + O(\lambda^2 n) \\ &\leq e^{\lambda^2 + O(\lambda/n) + O(\lambda^2/n) + O(\lambda^2 n)} = e^{O(\lambda/n + \lambda^2 n)}, \end{aligned}$$

for $\lambda \leq 1/(2en)$. Since $\lambda/n \leq \lambda^2 n$ for $\lambda \geq 1/n^2$, we have $\mathbb{E}(e^{\lambda D_t} \mid X_t = i) = e^{O(\lambda^2 n)}$ for all $\lambda \in [1/n^2, 1/(2en)]$. \square

Hence, we can satisfy the assumptions of Theorem 11 with $b_2 = 2en$ and $b_1 = n^2$. We will apply this theorem in the following subsection, where we put everything together.

5.4 Main Concentration Result – Putting Everything Together

In the previous subsections we have derived (w. r. t. LEADINGONES) that the sequence $\Delta_t^{(\ell)} = g(X_t) - g(X_{t+1}) + \sum_{s=0}^{t-1} \Pr(T > s)$ is a supermartingale and the sequence $\Delta_t^{(h)} = g(X_t) - g(X_{t+1}) + \sum_{s=0}^{t-1} \Pr(T > s) + r(t, n)$, where $r(t, n) = O(1/n)$, is a submartingale. We also know from Theorem 7 that $\mathbb{E}(g(X_t) \mid \mathcal{F}_0) \leq g(X_0) - \sum_{s=0}^{T-1} \Pr(T > s)$. Hence, using Theorem 11 with respect to the $\Delta_t^{(\ell)}$ -sequence, choosing $b_1 = \infty$ and $b_2 = 2en$ according to our analysis of the mgf., we obtain (since $\nu^2 = O(n)$) the first statement of the following theorem. Its second statement follows by applying Theorem 11 with respect to the $\Delta_t^{(h)}$ -sequence, choosing $b_2 = 2en$ and $b_1 = n^2$.

Theorem 12.

$$\Pr(g(X_t) \geq \mathbb{E}(g(X_t)) + d) \leq \begin{cases} e^{-d/(4en)}, & \text{if } d \geq Ct; \\ e^{-\Omega(d^2/(tn))}, & \text{otherwise,} \end{cases}$$

where $C = \nu^2/(4en) = O(1)$. Moreover,

$$\Pr(g(X_t) \leq \mathbb{E}(g(X_t)) - d - tr(t, n)) \leq \begin{cases} e^{-d/(4en)}, & \text{if } d \geq Ct; \\ e^{-\Omega(d^2/(tn))}, & \text{if } \frac{C't}{n} \leq d < Ct; \end{cases}$$

where $C = \nu^2/(4en) = \Theta(1)$ and $C' = \nu^2/n = \Theta(1)$.

As mentioned above, Theorem 7 gives us an upper bound on $\mathbb{E}(g(X_t))$ but we would like to know an upper bound on $\mathbb{E}(X_t)$. Unfortunately, since g is concave, it does **not** hold that $\mathbb{E}(X_t) \leq g^{-1}(\mathbb{E}(g(X_t)))$. However, using the concentration inequalities above, we can show that $\mathbb{E}(X_t)$ is not much bigger than the right-hand side of this wrong estimate. Given $t > 0$, we choose a $d^* > 0$ for the tail bound such that $\Pr(g(X_t) > \mathbb{E}(g(X_t)) + d^*) \leq 1/n^3$. If $g(X_t) \leq \mathbb{E}(g(X_t)) + d^*$, the concavity of g implies that the $\mathbb{E}(X_t)$ -value is maximized if $g(X_t)$ takes the value $\mathbb{E}(g(X_t)) + d^*$ with probability $\frac{\mathbb{E}(g(X_t))}{\mathbb{E}(g(X_t)) + d^*}$ and is 0 otherwise. Since $g(X_t) = O(n^2)$, we altogether have

$$\begin{aligned} \mathbb{E}(X_t) &\leq \frac{1}{n^3} O(n^2) + g^{-1}(\mathbb{E}(g(X_t)) + d^*) \frac{\mathbb{E}(g(X_t))}{\mathbb{E}(g(X_t)) + d^*} \\ &= g^{-1}(\mathbb{E}(g(X_t)) + d^*) \frac{\mathbb{E}(g(X_t))}{\mathbb{E}(g(X_t)) + d^*} + o(1). \end{aligned}$$

We will now make this concrete and conclude by presenting the proof of our main theorem from this section.

Proof of Theorem 10. First of all, we need some handy estimates for $g(a) = \frac{1}{h(1)} + \sum_1^a \frac{1}{h(i)} di$. Since i is an integer, we can also integrate over $1/h(\lceil i \rceil)$ instead and obtain

$$g(a) = \sum_{i=1}^a \frac{1}{h(i)} = \sum_{i=1}^a (2 - (1/2)^{i-1})^{-1} (1 - 1/n)^{i-n} n$$

using the expression from Lemma 1. So,

$$g(a) \leq \frac{n}{2} \sum_{i=1}^a (1 - 1/n)^{i-n} = \frac{n}{2} (n-1) (1 - 1/n)^{-n} \left(1 - \left(1 - \frac{1}{n} \right)^a \right).$$

Also, since $2 - (1/2)^{i-1} \geq 2 - 2/n$ for $i \geq \log n$, we have

$$\begin{aligned} g(a) &\geq \left(1 - \frac{1}{n} \right) \sum_{i=\log n+1}^a \frac{1}{2} (1 - 1/n)^{i-n} \\ &\geq \frac{(n-1)^2}{2} (1 - 1/n)^{-n} \left(1 - \left(1 - \frac{1}{n} \right)^a \right) - 2n \log n \\ &\geq \frac{en^2}{2} \left(1 - \left(1 - \frac{1}{n} \right)^a \right) - 3n \log n, \end{aligned} \tag{7}$$

where the last bound holds for sufficiently large n .

Let $X_t := n - V_t$ be the fitness distance at time t . Using Theorem 12 with $d = c^* \sqrt{tn \ln n}$, where c^* is a sufficiently large constant, and noting that $d < Ct$ by our assumption on t (for n large enough), we have

$$\Pr(g(X_t) \geq \mathbb{E}(g(X_t)) + d) \leq e^{-\Omega(d^2/(tn))} \leq 1/n^3.$$

So, if we have a bound on $\mathbb{E}(g(X_t))$, we will obtain a bound on $\mathbb{E}(X_t)$ via the inverse of g computed above.

We define $T^* = \frac{e-1}{2} n^2$ and note that this reflects the expected optimization time $\mathbb{E}(T)$ of the (1+1) EA on LEADINGONES up to a relative error of $1 + O(1/n)$ according to Lemma 1. Since $t \leq T^* - cn^{3/2} \sqrt{\log n}$, we obtain from Theorem 1 by choosing c sufficiently large that $\Pr(T \leq t) \leq 1/n^3$. Hence, from Theorem 7, $\mathbb{E}(g(X_t)) \leq \frac{e-1}{2} n^2 - t + O(1/n)$.

Altogether,

$$\Pr\left(g(X_t) \geq \frac{e-1}{2} n^2 - t + O(\sqrt{tn \log n})\right) \leq \frac{1}{n^3}.$$

We now invert g , more precisely the lower bound (7) since g is increasing and we want to bound the pre-image from above. Hence, we obtain that

$$z = g(a) \geq \frac{en^2}{2} \left(1 - \left(1 - \frac{1}{n} \right)^a \right) - 3n \log n$$

implies

$$a(z) \leq \frac{\ln(1 - \frac{2z+6n \log n}{en^2})}{\ln(1 - 1/n)}.$$

Using $z = \frac{e-1}{2}n^2 - t + O(\sqrt{tn \log n})$, we finally have

$$\begin{aligned} X_t &\leq \frac{\ln(1 - \frac{(e-1)n^2 - 2t + O(\sqrt{tn \log n}) + 6n \log n}{en^2})}{\ln(1 - 1/n)} \\ &= \frac{\ln(1/e - 2t/(en^2) + O(\sqrt{t \log n}/n^{3/2}))}{\ln(1 - 1/n)} \\ &= \frac{-1 + \ln(1 - 2t/n^2 + O(\sqrt{t \log n}/n^{3/2}))}{\ln(1 - 1/n)} \end{aligned}$$

with probability at least $1 - n^{-3}$. Estimating $\ln(1 - 1/n) = -1/n \pm O(1/n^2)$ and moving the error term into the logarithm, the lower bound on $V_t = n - X_t$ from the first claim follows of the theorem after straightforward manipulations.

The upper bound on V_t is proved almost analogously. We use the $\Delta^{(h)}$ sequence instead of the $\Delta^{(\ell)}$ sequence and note that the $r(t, n)$ terms vanish in the $O(\sqrt{t \log n}/n^{3/2})$ errors. The only requirement that has to be met additionally is that Lemma 3 only holds for $X_t > \log n$. However, by a straightforward change of Theorem 1 by choosing c as a sufficiently large constant we have not only $\Pr(T \leq t) \leq 1/n^3$ but also $\Pr(\min\{s \mid X_s \leq \log n\} \leq t) \leq 1/n^3$.

For the expected value, we note that $X_t \leq n$, so by the law of total probability, and again estimating $\ln(1 - 1/n) = -1/n \pm O(1/n^2)$, the claim follows. \square

6 Conclusions

We have described two general approaches that derive fixed-budget results via drift analysis. The first approach is concerned with iterating drifts either in an unbounded time scenario, or, using bounds on hitting times, in the scenario that the underlying process stops at some target state. Applying this approach to the ONEMAX or LEADINGONES functions, we obtain strong lower bounds on the expected fitness value after a given number of iterations. The second approach is based on variable drift analysis and tail bounds for martingale differences. Exemplified for the LEADINGONES function, this technique allows us to derive statements that are more precise than the previous state of the art. We think that our drift theorems can be useful for future fixed-budget analyses.

References

- [1] Süntje Böttcher, Benjamin Doerr, and Frank Neumann. Optimal fixed and adaptive mutation rates for the LeadingOnes problem. In *Proc. of PPSN 2010*, volume 6238 of *Lecture Notes in Computer Science*, pages 1–10. Springer, 2010.

- [2] Benjamin Doerr, Thomas Jansen, Carsten Witt, and Christine Zarges. A method to derive fixed budget results from expected optimisation times. In *Proc. of GECCO 2013*, pages 1581–1588. ACM Press, 2013.
- [3] Benjamin Doerr, Carola Doerr, and Jing Yang. Optimal parameter choices via precise black-box analysis. *Theoretical Computer Science*, 801:1–34, 2020.
- [4] Xiequan Fan, Ion Grama, and Quansheng Liu. Exponential inequalities for martingales with applications. *Electronic Journal of Probability*, 20:22 pp., 2015. URL <https://doi.org/10.1214/EJP.v20-3496>.
- [5] George T. Hall, Pietro Simone Oliveto, and Dirk Sudholt. On the impact of the cutoff time on the performance of algorithm configurators. In *Proc. of GECCO '19*, pages 907–915. ACM Press, 2019.
- [6] Jun He, Thomas Jansen, and Christine Zarges. Unlimited budget analysis of randomised search heuristics. *CoRR*, abs/1909.03342, 2019. URL <http://arxiv.org/abs/1909.03342>.
- [7] Hsien-Kuei Hwang and Carsten Witt. Sharp bounds on the runtime of the (1+1) EA via drift analysis and analytic combinatorial tools. In *Proc. of FOGA 2015*, pages 1–12. ACM Press, 2019. ISBN 978-1-4503-6254-2.
- [8] Thomas Jansen. Analysing stochastic search heuristics operating a fixed budget. In Benjamin Doerr and Frank Neumann, editors, *Theory of Evolutionary Computation: Recent Developments in Discrete Search Spaces*, pages 249–270. Springer, 2020.
- [9] Thomas Jansen and Christine Zarges. Fixed budget computations: a different perspective on run time analysis. In *Proc. of GECCO 2012*, pages 1325–1332. ACM Press, 2012.
- [10] Thomas Jansen and Christine Zarges. Reevaluating immune-inspired hypermutations using the fixed budget perspective. *IEEE Transactions on Evolutionary Computation*, 18(5):674–688, 2014.
- [11] Daniel Johannsen. *Random Combinatorial Structures and Randomized Search Heuristics*. PhD thesis, Universität des Saarlandes, Saarbrücken, Germany and the Max-Planck-Institut für Informatik, 2010.
- [12] Timo Kötzing. Concentration of first hitting times under additive drift. *Algorithmica*, 75:490506, 2016.
- [13] Per Kristian Lehre and Carsten Witt. Concentrated hitting times of randomized search heuristics with variable drift. In *Proc. of ISAAC 2014*, volume 8889 of *Lecture Notes in Computer Science*, pages 686–697. Springer, 2014. Extended version at <http://arxiv.org/abs/1307.2559>.

- [14] Johannes Lengler. Drift analysis. In Benjamin Doerr and Frank Neumann, editors, *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*, pages 89–131. Springer, 2020.
- [15] Johannes Lengler and Nicholas Spooner. Fixed budget performance of the (1+1) EA on linear functions. In *Proceedings of FOGA 2015*, pages 52–61. ACM Press, 2015.
- [16] Colin McDiarmid. Concentration. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, page 195247. Springer, 1998.
- [17] Samadhi Nallaperuma, Frank Neumann, and Dirk Sudholt. Expected fitness gains of randomized search heuristics for the traveling salesperson problem. *Evolutionary Computation*, 25(4), 2017.
- [18] M. Wainwright. Basic tail and concentration bounds. Technical report, 2015. Lecture Notes, Univ. of Berkeley, https://www.stat.berkeley.edu/~mhwain/stat210b/Chap2_TailBounds_Jan22_2015.pdf.