Improved fixed-budget results via drift analysis

Kötzing, Timo; Witt, Carsten

Published in:
Parallel Problem Solving from Nature

Link to article, DOI:
10.1007/978-3-030-58115-2_45

Publication date:
2020

Document Version
Peer reviewed version

Citation (APA):
Kötzing, T., & Witt, C. (2020). Improved fixed-budget results via drift analysis. In T. Bäck, M. Preuss, A. Deutz, M. Emmerich, H. Wang, C. Doerr, & H. Trautmann (Eds.), Parallel Problem Solving from Nature (pp. 648-660). Springer. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) Vol. 12270 LNCS https://doi.org/10.1007/978-3-030-58115-2_45
Improved Fixed-Budget Results via Drift Analysis

Timo Kötzing\textsuperscript{1} and Carsten Witt\textsuperscript{2}

\textsuperscript{1} Hasso Plattner Institute, Potsdam, Germany, timo.koetzing@hpi.de
\textsuperscript{2} Technical University of Denmark, Kgs. Lyngby, Denmark, cawi@dtu.dk

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 \textsc{Leading-Ones} 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 \textit{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 \textit{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 \textsc{OneMax} and \textsc{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 8. 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 7). 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 12). 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 mathemat-
ical 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. Due to space limitations, all proofs have
been removed from this article. A full technical report is available at [? ].

2 Preliminaries

The concrete objective functions we are concerned with in this paper are One-
Max and LeadingOnes, studied in a large number of papers. These two func-
tions 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.

\begin{verbatim}
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         with probability \( 1/n \): \( y_i \leftarrow 1 - y_i \);
6     if \( f(y) \geq f(x) \) then \( x \leftarrow y \);
\end{verbatim}

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 2.** 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

1. $E(X_t - X_{t+1} \mid X_t) = (2 - 2^{1-X_t})(1 - 1/n)^{X_t}/n$
2. $\Pr(X_{t+1} \neq X_t \mid X_t; T > t) = (1 - 1/n)^{X_t} n^{1 - X_t}/n$
3. For $j \geq 1$, $\Pr(X_{t+1} = X_t - j) \leq \frac{1}{n} (\frac{1}{2})^{j-1}$
4. $G_t := X_t - X_{t+1}$ is a random variable with support $0, \ldots, X_t$ 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

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

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

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 3.** We say that a drift function $h: S \to \mathbb{R}^{\geq 0}$ is **greed-admitting** if $id - h$ (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 \(\delta \times\)
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 4 (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 \(F_t.\) Let \(h: S \rightarrow \mathbb{R}^\geq 0\) be
a convex and greed-admitting function such that we have the drift condition
\((D\text{-}ut)}\n\[ E(X_t - X_{t+1} \mid F_t) \geq h(X_t). \]
Define \(\tilde{h}(x) = x - h(x).\) Thus, the drift condition is equivalent to
\((D\text{-}ut')}\n\[ E(X_{t+1} \mid F_t) \leq \tilde{h}(X_t). \]
We have that, for all \(t \geq 0,\)
\[ E(X_t \mid F_0) \leq \tilde{h}^t(X_0) \]
and, in particular,
\[ E(X_t) \leq \tilde{h}^t(E(X_0)). \]

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 5 (Direct Fixed-Budget Drift, limited time).** Let \(X_t, t \geq 0,\)
be a stochastic process on \(S \subseteq \mathbb{R},\) adapted to a filtration \(F_t.\) 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 \(h'(0) \in [0, 1]\) and we have the drift condition
\((D\text{-}lt)}\n\[ E(X_t - X_{t+1} \mid F_t; t < T) \geq h(X_t). \]
Define \(\bar{h}(x) = x - h(x).\) Thus, the drift condition is equivalent to
\[ \tilde{h}(x) = x - h(x). \] Recall from the preliminaries that \(f^i\) is the \(i\)-times self-composition of a function \(f.\)
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,
\[
\mathbb{E}(X_t) \leq \tilde{h}^t(\mathbb{E}(X_0)) - \frac{\tilde{h}(0)}{\tilde{h}'(0)} \cdot \Pr(t \geq T \mid \mathcal{F}_0).
\]

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 9 and the discussion about drift theory in general in [14].

**Theorem 6.** 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 \( t \),
\[
\text{if } t \geq \sum_{i=m}^{n-1} \frac{1}{h(i)} \text{ then } \tilde{h}^t(n) \leq m.
\]

### 3.1 Application to OneMax

In this section we show how we can apply Theorem 4 by using the optimization of the (1+1) EA on OneMax as an example (where we have multiplicative drift).

**Theorem 7.** 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
\[
\mathbb{E}(V_t) \geq n(1 - \exp(-t/(en))/2).
\]

### 3.2 Application to LeadingOnes

In this section we want to use Theorem 5 to the progress of the (1+1) EA on LeadingOnes. The result is summarized in the following theorem.

**Theorem 8.** 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 \( t \),
\[
\mathbb{E}(V_t) \geq \begin{cases} 
\frac{2t}{n} - O(1), & \text{if } t = O(n^{3/2}); \\
\frac{2t}{n}(1 - o(1)), & \text{if } t = o(n^2); \\
n \ln(1 + \frac{2t}{n^2}) - O(1), & \text{if } t \leq \frac{e-1}{2}n^2 - n^{3/2}.
\end{cases}
\]
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 9.** Let $X_t, t \geq 0$, be a stochastic process, adapted to a filtration $\mathcal{F}_t$, on $S := \{0\} \cup \mathbb{R}^x_{x_{\text{min}}}$ for some $x_{\text{min}} > 0$. Let $T := \min\{t \geq 0 \mid X_t = 0\}$ and $h: S \to \mathbb{R}^{>0}$ be a non-decreasing function such that $E(X_t - X_{t+1} \mid \mathcal{F}_t; t < T) \geq h(X_t)$. Define $g: S \to \mathbb{R}$ by

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

Then it holds that

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

4.1 Additive Drift as Special Case

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

**Theorem 10.** 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 $E(X_t - X_{t+1} \mid \mathcal{F}_t; t < T) \geq \delta$. Then we have

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

The theorem is a corollary to Theorem 9 by using $x_{\text{min}} = \delta$, the smallest value for which the condition of a drift of at least $\delta$ 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 11.** 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 $t \leq \frac{e-1}{3} n^2 - \frac{n^3}{2} \log(n)$, we have

$$E(V_t) \geq \frac{2t}{e n} - O(1).$$
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 8.

5 Variable Drift and Concentration Inequalities

The expected $g(X_t)$-value derived in Theorem 9 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 9, 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 12.** 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 $1 - 1/n^3$,

\[-n \ln \left( 1 - 2t/n^2 + O(\sqrt{t \log n}/n^{3/2}) \right) \leq V_t\]

\[-n \ln \left( 1 - 2t/n^2 - O(\sqrt{t \log n}/n^{3/2}) \right) \geq V_t.\]

(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 turns out as $O(\sqrt{t \log n}/n^{3/2})$. This is asymptotically smaller than the additive error term of order $\Omega(n^{3/2+\epsilon})$ 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 12 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 \( \sum_{t=0}^{t-1} D_t = Y_t - Y_0 \) 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 \( E(e^{\lambda D_t} \mid F_t) \leq e^{\lambda^2 \nu_t^2 / 2} \) for a certain parameter \( \nu_t \) and all \( \lambda < 1/b_t \), where \( b_t \) is another parameter. In particular, the bound has to remain true when \( \lambda \) becomes arbitrarily small.

In one of our concrete applications of the martingale difference technique, the inequality \( E(e^{\lambda D_t} \mid F_t) \leq e^{\lambda^2 \nu_t^2 / 2} \) is true for certain values of \( \lambda \) below a threshold \( 1/b^* \), but does not hold if \( \lambda \) 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 13.** Let \( Y_t, t \geq 0 \), be a supermartingale, adapted to a filtration \( F_t \), 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 \( \nu_t, t \geq 0 \), such that for \( \lambda \in [1/b_1, 1/b_2] \) it holds that \( E(e^{\lambda D_t} \mid F_t) \leq e^{\lambda^2 \nu_t^2 / 2} \). Then for all \( t \geq 0 \) it holds that

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

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

5.2 Preparing an Upper Tail Bound via the Martingale Difference Method

We now return to Theorem 9 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 13, and inverting \( g \), we will obtain a bound on the probability of the event \( g(X_t) \geq \E(g(X_t)) \).
As we want to prove Theorem 12, 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 $h(X_t) := E(X_t - X_{t+1} \mid X_t)$ according to Lemma 2 and $g(X_t) = 1/h(1) + \int_1^{X_t} 1/h(z) \, dz$ according to Lemma 9, we will establish the following bound on the moment-generating function (mgf.) of the drift of our concrete $g$.

**Lemma 14.** Let $T$ denote the optimization time of the (1+1) EA on LeadingOnes. If $\lambda \leq 1/(2en)$ then $E(e^{\lambda (g(X_t) - g(X_{t+1}) + \Pr(T > t))} \mid X_t) = e^{O(\lambda^2 n)}$.

Looking into Theorem 13 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 13, 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., $E(Y_{t+1} \mid Y_t) \geq Y_t$. 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 15.** 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 $X_t > \log n$.

Recall that the aim is to apply Theorem 13 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 16.** The mgf. of $D_t = Y_t - Y_{t-1} = g(X_{t+1}) - g(X_t) + \Pr(T > t) + r(t, n)$ satisfies $E(e^{\lambda D_t} \mid X_t) = e^{O(\lambda^2 n)}$ for all $\lambda \in [1/n^2, 1/(2en)]$.

Hence, we can satisfy the assumptions of Theorem 13 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 $\Delta_t^{(1)} = g(X_t) - g(X_{t+1}) + \sum_{s=0}^{t-1} \Pr(T > s)$ is a supermartingale and the sequence $\Delta_t^{(b)} = g(X_t) - g(X_{t+1}) + \sum_{s=0}^{t-1} \Pr(T > s) + r(t, n)$, where $r(t, n) = O(1/n)$, is a submartingale. We also know from Theorem 9 that $E(g(X_t) \mid F_0) \leq \ldots$
$g(X_n) - \sum_{i=0}^{T-1} \Pr(T > s)$. Hence, using Theorem 13 with respect to the $\Delta_i^{(t)}$-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 13 with respect to the $\Delta_i^{(h)}$-sequence, choosing $b_2 = 2en$ and $b_1 = n^2$.

**Theorem 17.**

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

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

$$\Pr(g(X_t) \leq E(g(X_t)) - d - tr(t, n)) \leq \begin{cases} e^{-d/(4\nu n)}, & \text{if } d \geq Ct; \\ e^{-\Omega(d^2/(\nu n))}, & \text{if } \frac{Ct^2}{n} \leq d < Ct; \end{cases}$$

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

As mentioned above, Theorem 9 gives us an upper bound on $E(g(X_t))$ but we would like to know an upper bound on $E(X_t)$. Unfortunately, since $g$ is not concave, it does not hold that $E(X_t) \leq g^{-1}(E(g(X_t)))$. However, using the concentration inequalities above, we can show that $E(X_t)$ 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 $\Pr(g(X_t) > E(g(X_t)) + d^*) \leq 1/n^3$. If $g(X_t) \leq E(g(X_t)) + d^*$, the concavity of $g$ implies that the $E(X_t)$-value is maximized if $g(X_t)$ takes the value $E(g(X_t)) + d^*$ with probability $\frac{E(g(X_t))}{E(g(X_t)) + d^*}$ and is 0 otherwise. Since $g(X_t) = O(n^2)$, we altogether have

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

The omitted proof of Theorem 12 makes this idea concrete.

## 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.
Bibliography

[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] 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.

[8] 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.

[9] Thomas Jansen and Christine Zarges. Reevaluating immune-inspired hypermutations using the fixed budget perspective. IEEE Transactions on Evolutionary Computation, 18(5):674–688, 2014.

[10] Daniel Johannsen. Random Combinatorial Structures and Randomized Search Heuristics. PhD thesis, Universit¨at des Saarlandes, Saarbr¨ucken, Germany and the Max-Planck-Institut f¨ur Informatik, 2010.

[11] Timo K¨otzing. Concentration of first hitting times under additive drift. Algorithmica, 75:490–506, 2016.

[12] Timo K¨otzing and Carsten Witt. Improved Fixed-Budget Results via Drift Analysis. 2020. URL http://arxiv.org/abs/2006.07019.

[13] 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.

[14] 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.

[15] Colin McDiarmid. Concentration. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors, Probabilistic Methods for Algorithmic Discrete Mathematics, page 195–247. 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/~mjwain/stat210b/Chap2_TailBounds_Jan22_2015.pdf.