

ARTICLE

Tail bounds on hitting times of randomized search heuristics using variable drift analysis[†]

P. K. Lehre¹ and C. Witt^{2*}

¹School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK and ²DTU Compute, Technical University of Denmark, Kongens Lyngby, Denmark

*Corresponding author. Email: cawi@dtu.dk

(Received 3 May 2018; revised 23 September 2020; accepted 10 September 2020; first published online 5 November 2020)

Abstract

Drift analysis is one of the state-of-the-art techniques for the runtime analysis of randomized search heuristics (RSHs) such as evolutionary algorithms (EAs), simulated annealing, *etc.* The vast majority of existing drift theorems yield bounds on the expected value of the hitting time for a target state, for example the set of optimal solutions, without making additional statements on the distribution of this time. We address this lack by providing a general drift theorem that includes bounds on the upper and lower tail of the hitting time distribution. The new tail bounds are applied to prove very precise sharp-concentration results on the running time of a simple EA on standard benchmark problems, including the class of general linear functions. On all these problems, the probability of deviating by an r -factor in lower-order terms of the expected time decreases exponentially with r . The usefulness of the theorem outside the theory of RSHs is demonstrated by deriving tail bounds on the number of cycles in random permutations. All these results handle a position-dependent (variable) drift that was not covered by previous drift theorems with tail bounds. Finally, user-friendly specializations of the general drift theorem are given.

2020 MSC Codes: Primary: 68W50; Secondary: 68W20, 68Q87

1. Introduction

Randomized search heuristics (RSHs) such as simulated annealing, evolutionary algorithms (EAs), ant colony optimization, *etc.*, are highly popular techniques in black-box optimization, *i.e.* the problem of optimizing a function with only oracle access to the function. These heuristics often imitate some natural process, and are rarely designed with analysis in mind. Their extensive use of randomness, such as in the mutation operator, render the underlying stochastic processes non-trivial. While the theory of RSHs is less developed than the theory of classical, randomized algorithms, significant progress has been made in the past decade [2, 10, 19, 33]. This theory has mainly focused on the optimization time, which is the random variable $T_{A,f}$ defined as the number of oracle accesses the heuristic A makes before the maximal argument of f is found. Most classical studies considered the expectation of $T_{A,f}$, but more information about the distribution of the optimization time is often needed. For example, the expectation can be deceptive when the runtime distribution has a high variance. Also, tail bounds can be helpful for other performance

[†] A preliminary version of this paper appeared in the proceedings of ISAAC 2014 [28].

measures such as fixed-budget computation, which seeks to estimate the approximation quality as a function of time [8, 20, 25].

Results on the runtime of RSHs were obtained after relevant analytical techniques were developed, some adopted from other fields, others developed specifically for RSHs. Drift analysis, which is a central method for analysing the hitting time of stochastic processes, was introduced to the analysis of simulated annealing as early as 1988 [37]. Informally, it allows long-term properties of a discrete-time stochastic process $(X_t)_{t \in \mathbb{N}_0}$ to be inferred from properties of the one-step change $\Delta_t := X_t - X_{t+1}$. In the context of EAs, there has been particular interest in the random variable T_a defined as the smallest t such that $X_t \leq a$. For example, if X_t represents the ‘distance’ of the current solution in iteration t to an optimum, then T_0 is the optimization time.

Since its introduction to evolutionary computation by He and Yao in 2001 [16], drift analysis has been widely used to analyse the optimization time of EAs. Many drift theorems have been introduced, such as additive [16], multiplicative [7, 9], variable [21, 32, 36], and population [26] drift theorems. Drift analysis is also used outside the theory of RSHs, for example in queuing theory [3, 12]. The widespread use of these techniques in separated research fields has made it difficult to get an overview and a unified presentation of the drift theorems; see the recent survey by Lengler [30]. However, at least for the case of expected first hitting times under additive drift, theorems that are as general as possible have been obtained in the meantime [24]. Drift analysis is also related to other areas, such as stochastic differential equations and stochastic difference relations.

Most drift theorems used in the theory of RSHs relate to the expectation of the hitting time T_a , and there are fewer results about the tails $\mathbb{P}[T_a > t]$ and $\mathbb{P}[T_a < t]$. From the simple observation that

$$\mathbb{P}[T_a > t] \leq \mathbb{P}\left[\sum_{i=0}^t \Delta_i < a - X_0\right],$$

the problem is reduced to bounding the deviation of a sum of random variables. If the Δ_t were independent and identically distributed, then one would be in the familiar scenario of Chernoff/Hoeffding-like bounds. The stochastic processes originating from RSHs are rarely so simple; in particular, the Δ_t are often dependent variables, and their distributions are not explicitly given. However, bounds of the form $\mathbb{E}[\Delta_t | X_t] \geq h(X_t)$ for some function h often hold. The drift is called *variable* when h is a non-constant function. The variable drift theorem provides bounds on the expectation of T_a given some conditions on h . However, there have been no general tail bounds from a variable drift condition. The only results in this direction seem to be the tail bounds for probabilistic recurrence relations from [22]; however, this scenario is restricted to monotonically decreasing stochastic processes.

Our contribution is a new, general drift theorem that provides sharp-concentration results for the hitting time of stochastic processes with variable drift, along with concrete advice and examples how to apply it. The theorem is used to bound the tails of the optimization time of the well-known $(1 + 1)$ EA [11] to the benchmark problems ONEMAX and LEADINGONES, as well as the class of linear functions, which is an intensively studied problem in the area [40]. The results show that the distribution is highly concentrated around the expectation. The probability of deviating by an r -factor in lower-order terms decreases exponentially with r . In an application outside the theory of RSHs, we analyse the drift in probabilistic recurrence relations, showing that the number of cycles in a random permutation is sharply concentrated around the expectation $H_n \approx \ln n$.

This paper is structured as follows. Section 2 introduces notation and basics of drift analysis. Section 3 presents the general drift theorem with tail bounds and suggestions for user-friendly corollaries. Section 4 applies the tail bounds from our theorem. Sharp-concentration results on the running time of the $(1 + 1)$ EA on ONEMAX, LEADINGONES and general linear functions are

obtained. An application outside the theory of RSHs with respect to random recurrence relations is described at the end of this section (Section 4.2). In all these applications, the probability of deviating by an r -factor in lower-order terms of the expected time decreases exponentially with r .

2. Preliminaries

We analyse time-discrete stochastic processes represented by a sequence of non-negative random variables $(X_t)_{t \in \mathbb{N}_0}$. For example, X_t could represent a certain distance value of an RSH from an optimum. We adopt the convention that the process should pass below some threshold $a \geq 0$ ('minimizes' its state) and define the first hitting time $T_a := \min\{t \mid X_t \leq a\}$. If the actual process seeks to maximize its state, typically a straightforward mapping allows us to stick to the convention of minimization. In an important special case, we are interested in the hitting time T_0 of target state 0; for example when a $(1 + 1)$ EA, a very simple RSH, is run on the well-known ONEMAX problem and we are interested in the first point of time where the number of zero-bits becomes zero. Note that T_a is a stopping time and that we assume that the stochastic process is adapted to some filtration $(\mathcal{F}_t)_{t \in \mathbb{N}_0}$, such as its natural filtration $\sigma(X_0, \dots, X_t)$.

Our main goal is to describe properties of the distribution of the first hitting time T_a , hence some information about the stochastic process before that time is required. In particular, we consider the expected one-step change of the process

$$\delta_t := \mathbb{E}[X_t - X_{t+1} ; X_t > a \mid \mathcal{F}_t],$$

the so-called *drift*. For any event A and random variable X , we use the well-established notation $\mathbb{E}[X ; A \mid \mathcal{F}_t] := \mathbb{E}[X \cdot \mathbb{1}[A] \mid \mathcal{F}_t]$, where $\mathbb{1}[\]$ is the indicator function. Note that δ_t in general is a random variable since the outcomes of X_0, \dots, X_t are random. Suppose we manage to bound the random variable δ_t from below by some real number $\delta^* > 0$, conditioning on $X_t \geq a$. This is the same as bounding

$$\mathbb{E}[X_t - X_{t+1} - \delta^* ; X_t > a \mid \mathcal{F}_t] \geq 0,$$

except for the case that $\mathbb{P}[X_t > a]$, where the conditioning does not work; however, this difference is unimportant for our analysis of first hitting time. Then, informally speaking, we know that the process, conditioned on not having reached the target, decreases its state in expectation by at least δ^* in every step, and the additive drift theorem (see Theorem 2.1 below) will provide a bound on T_0 that only depends on X_0 and δ^* . In fact, the very natural-looking result $\mathbb{E}[T_0 \mid \mathcal{F}_0] \leq X_0/\delta^*$ will be obtained. However, bounds on the drift might be more complicated. For example, a bound on δ_t might depend on X_t or states at even earlier points of time, e.g. if the progress decreases as the current state decreases. This is often the case in applications to EAs.

As pointed out, the drift δ_t is in general a random variable and should not be confused with the 'expected drift' $\mathbb{E}[\delta_t] = \mathbb{E}[\mathbb{E}[X_t - X_{t+1} ; X_t > a \mid \mathcal{F}_t]]$, which is rarely available since it averages over the whole history of the stochastic process. Drift as used in this paper is based on the inspection of the progress from one step to another, taking into account every possible history. This one-step inspection often makes it easy to come up with bounds on δ_t . Drift theorems could also be formulated based on expected drift, possibly allowing a stronger statement on the first hitting time. However, in many applications it is infeasible to bound the expected value of the drift in a precise enough way for a stronger statement to be obtained. See [18] for one of the rare analyses of 'expected drift', which we will not get into in this paper.

We now cite the first drift theorem for additive drift. It goes back to [16] and has subsequently been generalized in various ways, for example by removing unnecessary assumptions such as a discrete search space and the Markov property. The formulation closely follows [24]. For convenience, we demand a bounded state space for the lower bound; variants for two-sided unbounded spaces are discussed in [24].

Theorem 2.1 (additive drift, following [24]). *Let $(X_t)_{t \in \mathbb{N}_0}$, be a stochastic process, adapted to a filtration $(\mathcal{F}_t)_{t \in \mathbb{N}_0}$, over some state space $S \subseteq \mathbb{R}$, and let $b, \delta_u, \delta_\ell > 0$. Letting $\Delta_t := X_t - X_{t+1}$ and $T_0 := \min\{t \mid X_t \leq 0\}$, we have the following.*

- (i) *If $\mathbb{E}[\Delta_t - \delta_u; X_t > 0 \mid \mathcal{F}_t] \geq 0$ and $X_t \geq 0$ for all $t \in \mathbb{N}_0$, then $\mathbb{E}[T_0 \mid \mathcal{F}_0] \leq X_0/\delta_u$.*
- (ii) *If $\mathbb{E}[\Delta_t - \delta_\ell; X_t > 0 \mid \mathcal{F}_t] \leq 0$ and $X_t \leq b$ for all $t \in \mathbb{N}_0$, then $\mathbb{E}[T_0 \mid \mathcal{F}_0] \geq X_0/\delta_\ell$.*

Additive drift concerns the simple scenario where there is a progress of at least δ_u from all non-optimal states towards the target in (i) and a progress of at most δ_ℓ in (ii). Since the δ -values are independent of X_t , one has to use the worst-case drift over all non-optimal X_t . This might lead to very bad bounds on the first hitting time, which is why more general theorems were developed (as mentioned in the Introduction). Interestingly, these more general theorems are often proved based on Theorem 2.1 using an appropriate mapping (sometimes called *Lyapunov function, potential function, distance function* or *drift function*) from the original state space to a new one. Informally, the mapping ‘smooths out’ position-dependent drift into an (almost) position-independent drift. We will use the same approach when deriving concrete tail bounds in Section 4.

3. General drift theorem

In this section we present our general drift theorem. As pointed out in the Introduction, we strive for a general statement, partly at the expense of simplicity. More user-friendly specializations will be given later. Nevertheless, the underlying idea of the complicated-looking general theorem is the same as in all drift theorems. We look into the one-step drift $\delta_t = \mathbb{E}[X_t - X_{t+1} \mid \mathcal{F}_t]$, which is a random variable that may depend on the complete history of the process up to time t . Then we assume we have an (upper or lower) bound $h(X_t)$ on the drift, formally $\delta_t \geq h(X_t)$ or $\delta_t \leq h(X_t)$, where the bound depends on X_t only, *i.e.* a possibly smaller σ -algebra than \mathcal{F}_t . Based on h , we define a new function g (see Remark 3.1), with the aim of ‘smoothing out’ the dependency, and the drift with respect to g (formally, $\mathbb{E}[g(X_t) - g(X_{t+1}) \mid \mathcal{F}_t]$) is analysed. Statements (i) and (ii) of the following Theorem 3.2 provide bounds on $\mathbb{E}[T_0]$ based on the drift with respect to g . In fact, g can be defined in a very similar way to existing variable-drift theorems [21, 32, 36], such that statements (i) and (ii) can be understood as generalized variable drift theorems for upper and lower bounds on the expected hitting time, respectively.

Statements (iii) and (iv) consider general $a \geq 0$ and concern tail bounds on the hitting time T_a , the main focus of this paper. Here moment-generating functions (MGFs) of the drift with respect to g come into play, formally $\mathbb{E}[e^{-\lambda(g(X_t) - g(X_{t+1}))} \mid \mathcal{F}_t]$ is bounded. Bounds on the MGF may depend on the point of time t , as captured by the bounds $\beta_u(t)$ and $\beta_\ell(t)$. Section 4 gives an example where the mapping g smooths out the position-dependent drift into a (nearly) position-independent and time-independent drift, while the MGF of the drift with respect to g still depends on the current point (and indirectly on the expected position) of time t .

Our drift theorem generalizes virtually all existing drift theorems concerned with a drift towards the target, including variable drift theorems for upper [21, 32, 36] and lower bounds [5, 6, 14], a non-monotone variable drift theorem [13] and multiplicative drift theorems [4, 7, 40]. Our theorem also generalizes fitness-level theorems [38, 39], another well-known technique in the analysis of randomized search heuristics. Some examples of such generalizations are shown in a supplementary technical report [29]; however, often the proof in the original publication already makes it explicit that the additive drift theorem is applied with respect to an appropriately defined potential function. Note that we do not consider the case of negative drift (drift away from the target) as studied in [31], [34] and [35], since this scenario is handled with structurally different techniques.

Remark 3.1. If, for some function $h: \mathbb{R}_{\geq x_{\min}} \rightarrow \mathbb{R}^+$ where $x_{\min} > 0$ and $1/h(x)$ is integrable on $\mathbb{R}_{\geq x_{\min}}$, we have either

$$\mathbb{E}[X_t - X_{t+1} - h(X_t); X_t \geq x_{\min} | \mathcal{F}_t] \geq 0$$

or

$$\mathbb{E}[X_t - X_{t+1} - h(X_t); X_t \geq x_{\min} | \mathcal{F}_t] \leq 0,$$

it is recommended to define the function g in Theorem 3.2 as $g(x) := x/h(x_{\min})$ for all $x < x_{\min}$, and otherwise for all $x \geq x_{\min}$

$$g(x) := \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^x \frac{1}{h(y)} dy.$$

Theorem 3.2 (general drift theorem). Let $(X_t)_{t \in \mathbb{N}_0}$, be a stochastic process, adapted to a filtration $(\mathcal{F}_t)_{t \in \mathbb{N}_0}$, over some state space $S \subseteq \mathbb{R}$. For some $a \geq 0$ (assumed 0 in (i) and (ii)), let $T_a = \min\{t | X_t \leq a\}$. Moreover, let $g: S \cup \{0, a\} \rightarrow \mathbb{R}_{\geq 0}$ be a function such that $g(0) = 0$ and $g(x) > g(a)$ for all $x \in S \cap (a, \infty)$.

(i) If

$$\mathbb{E}[g(X_t) - g(X_{t+1}) - \alpha_u; X_t > 0 | \mathcal{F}_t] \geq 0 \quad \text{for all } t \in \mathbb{N}_0 \text{ and some } \alpha_u > 0,$$

then $\mathbb{E}[T_0 | \mathcal{F}_0] \leq g(X_0)/\alpha_u$.

(ii) If there is $x_{\max} > 0$ such that $g(X_t) \leq x_{\max}$ and

$$\mathbb{E}[g(X_t) - g(X_{t+1}) - \alpha_\ell; X_t > 0 | \mathcal{F}_t] \leq 0 \quad \text{for all } t \in \mathbb{N}_0 \text{ and some } \alpha_\ell > 0,$$

then $\mathbb{E}[T_0 | \mathcal{F}_0] \geq g(X_0)/\alpha_\ell$.

(iii) If there exists $\lambda > 0$ and a function $\beta_u: \mathbb{N}_0 \rightarrow \mathbb{R}^+$ such that

$$\mathbb{E}[e^{-\lambda(g(X_t) - g(X_{t+1}))} - \beta_u(t); X_t > a | \mathcal{F}_t] \leq 0 \quad \text{for all } t \in \mathbb{N}_0,$$

then

$$\mathbb{P}[T_a > t^* | \mathcal{F}_0] < \left(\prod_{r=0}^{t^*-1} \beta_u(r) \right) \cdot e^{\lambda(g(X_0) - g(a))} \quad \text{for } t^* > 0.$$

(iv) If there exists $\lambda > 0$ and a function $\beta_\ell: \mathbb{N}_0 \rightarrow \mathbb{R}^+$ such that

$$\mathbb{E}[e^{\lambda(g(X_t) - g(X_{t+1}))} - \beta_\ell(t); X_t > a | \mathcal{F}_t] \leq 0 \quad \text{for all } t \in \mathbb{N}_0,$$

then

$$\mathbb{P}[T_a < t^* | \mathcal{F}_0] \leq \left(\sum_{s=1}^{t^*-1} \prod_{r=0}^{s-1} \beta_\ell(r) \right) \cdot e^{-\lambda(g(X_0) - g(a))} \quad \text{for } t^* > 0 \text{ and } X_0 > a.$$

If additionally the set of states $S \cap \{x | x \leq a\}$ is absorbing, then

$$\mathbb{P}[T_a < t^* | \mathcal{F}_0] \leq \left(\prod_{r=0}^{t^*-1} \beta_\ell(r) \right) \cdot e^{-\lambda(g(X_0) - g(a))}.$$

Statement (ii) is also valid (but useless) if the expected hitting time is infinite.

Special cases of (iii) and (iv). If

$$\mathbb{E}[e^{-\lambda(g(X_t) - g(X_{t+1}))} - \beta_u; X_t > a | \mathcal{F}_t] \leq 0$$

for some time-independent β_u , then statement (iii) simplifies to the bound

$$\mathbb{P}[T_a > t^* \mid \mathcal{F}_0] < \beta_u^{t^*} \cdot e^{\lambda(g(X_0)-g(a))},$$

and similarly for statement (iv).

The tail bounds in (iii) and (iv) are obtained easily by the exponential method (a generalized Chernoff bound), which idea is also implicit in [15].

Proof of Theorem 3.2. Since $g(X_t) = 0$ if and only if $X_t = 0$, and since the image of g is bounded from below by 0 and additionally by x_{\max} in item (ii), the first two items follow from the classical additive drift theorem (Theorem 2.1). To prove item (iii) we consider the stopped process that does not move after time T_a . We now use ideas implicit in [15] and argue that

$$\begin{aligned} \mathbb{P}[T_a > t^* \mid \mathcal{F}_0] &\leq \mathbb{P}[X_{t^*} > a \mid \mathcal{F}_0] \\ &\leq \mathbb{P}[g(X_{t^*}) > g(a) \mid \mathcal{F}_0] \\ &= \mathbb{P}[e^{\lambda g(X_{t^*})} > e^{\lambda g(a)} \mid \mathcal{F}_0] \\ &< \mathbb{E}[e^{\lambda g(X_{t^*})-\lambda g(a)} \mid \mathcal{F}_0], \end{aligned}$$

where the second inequality uses that $X_{t^*} > a$ implies $g(X_{t^*}) > g(a)$, the equality that $x \mapsto e^x$ is a bijection, and the last inequality is Markov's inequality. Now

$$\begin{aligned} \mathbb{E}[e^{\lambda g(X_{t^*})} \mid \mathcal{F}_0] &= \mathbb{E}[e^{\lambda g(X_{t^*-1})} \cdot \mathbb{E}[e^{-\lambda(g(X_{t^*-1})-g(X_{t^*}))} \mid \mathcal{F}_{t^*-1}] \mid \mathcal{F}_0] \\ &\leq \mathbb{E}[e^{\lambda g(X_{t^*-1})} \mid \mathcal{F}_0] \cdot \beta_u(t^* - 1) \end{aligned}$$

using the prerequisite from item (iii). Unfolding the remaining expectation inductively (note that this does not assume independence of $g(X_{r-1}) - g(X_r)$), we get

$$\mathbb{E}[e^{\lambda g(X_{t^*})} \mid \mathcal{F}_0] \leq e^{\lambda g(X_0)} \prod_{r=0}^{t^*-1} \beta_u(r),$$

altogether

$$\mathbb{P}[T_a > t^* \mid \mathcal{F}_0] < e^{\lambda(g(X_0)-g(a))} \prod_{r=0}^{t^*-1} \beta_u(r),$$

which proves item (iii).

Item (iv) is proved similarly to (iii). Using a union bound and that $X_{t^*} \leq a$ follows from $g(X_{t^*}) \leq g(a)$,

$$\mathbb{P}[T_a < t^* \mid \mathcal{F}_0] \leq \sum_{s=1}^{t^*-1} \mathbb{P}[g(X_s) \leq g(a) \mid \mathcal{F}_0] \quad \text{for } t^* > 0,$$

assuming $X_0 > a$. Moreover,

$$\mathbb{P}[g(X_s) \leq g(a) \mid \mathcal{F}_0] = \mathbb{P}[e^{-\lambda g(X_s)} \geq e^{-\lambda g(a)} \mid \mathcal{F}_0] \leq \mathbb{E}[e^{-\lambda g(X_s)+\lambda g(a)} \mid \mathcal{F}_0]$$

again using Markov's inequality. By the prerequisites, we get

$$\mathbb{E}[e^{-\lambda g(X_s)} \mid \mathcal{F}_0] \leq e^{-\lambda g(X_0)} \prod_{r=0}^{s-1} \beta_\ell(r).$$

Altogether,

$$\mathbb{P}[T_a < t^* \mid \mathcal{F}_0] \leq \sum_{s=1}^{t^*-1} e^{-\lambda(g(X_0)-g(a))} \prod_{r=0}^{s-1} \beta_\ell(r).$$

If furthermore $S \cap \{x \mid x \leq a\}$ is absorbing, then the event $X_{t^*} \leq a$ is necessary for $T_a < t^*$. In this case

$$\mathbb{P}[T_a < t^* \mid \mathcal{F}_0] \leq \mathbb{P}[g(X_{t^*}) \leq g(a) \mid \mathcal{F}_0] \leq e^{-\lambda(g(X_0) - g(a))} \prod_{r=0}^{t^*-1} \beta_\ell(r). \quad \square$$

Condition (iii) and (iv) of Theorem 3.2 involve an MGF, which may be tedious to compute. Inspired by [15] and [27], we show that bounds on the MGFs follow from more user-friendly conditions based on stochastic dominance, here denoted by \leq .

Theorem 3.3. *Let $(X_t)_{t \in \mathbb{N}_0}$ be a stochastic process, adapted to a filtration $(\mathcal{F}_t)_{t \in \mathbb{N}_0}$, over some state space $S \subseteq \{0\} \cup \mathbb{R}_{\geq x_{\min}}$, where $x_{\min} \geq 0$ and $0 \in S$. Let $h: \mathbb{R}_{\geq x_{\min}} \rightarrow \mathbb{R}^+$ be a function such that $1/h(x)$ is integrable on $\mathbb{R}_{\geq x_{\min}}$. Suppose there exist a random variable Z and some $\lambda > 0$ such that $\int_{X_{t+1}}^{X_t} 1/h(\max\{x_{\min}, x\}) dx \leq Z$ for $X_t \geq x_{\min}$ and all $t \in \mathbb{N}_0$, and $\mathbb{E}[e^{\lambda Z}] = D$ for some $D > 0$. Then the following two statements hold for the first hitting time $T := \min\{t \mid X_t = 0\}$.*

(i) If

$$\mathbb{E}[X_t - X_{t+1} - h(X_t); X_t \geq x_{\min} \mid \mathcal{F}_t] \geq 0 \quad \text{for all } t \in \mathbb{N}_0$$

and h is non-decreasing, then for any $\delta > 0$, $\eta := \min\{\lambda, \delta\lambda^2/(D - 1 - \lambda)\}$ and $t^* > 0$ it holds that

$$\mathbb{P}[T > t^* \mid \mathcal{F}_0] \leq \exp\left(\eta \left(\int_{x_{\min}}^{X_0} \frac{1}{h(x)} dx - (1 - \delta)t^*\right)\right).$$

(ii) If

$$\mathbb{E}[X_t - X_{t+1} - h(X_t); X_t \geq x_{\min} \mid \mathcal{F}_t] \leq 0 \quad \text{for all } t \in \mathbb{N}_0$$

and h is non-increasing, then for any $\delta > 0$, $\eta := \min\{\lambda, \delta\lambda^2/(D - 1 - \lambda)\}$ and $t^* > 0$ it holds on $X_0 > 0$ that

$$\mathbb{P}[T < t^* \mid \mathcal{F}_0] \leq \exp\left(\eta \left((1 + \delta)t^* - \int_{x_{\min}}^{X_0} \frac{1}{h(x)} dx\right)\right) \frac{1}{\eta(1 + \delta)}.$$

If state 0 is absorbing, then

$$\mathbb{P}[T < t^* \mid \mathcal{F}_0] \leq \exp\left(\eta \left((1 + \delta)t^* - \int_{x_{\min}}^{X_0} \frac{1}{h(x)} dx\right)\right).$$

Remark 3.4. Theorem 3.3 assumes

$$\left| \int_{X_{t+1}}^{X_t} \frac{1}{h}(\max\{x_{\min}, x\}) dx \right| \leq Z.$$

This is implied by $|X_{t+1} - X_t|(1/\inf_{x \geq x_{\min}} h(x)) \leq Z$.

Proof. As in Remark 3.1, let

$$g(x) := \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^x \frac{1}{h(y)} dy \quad \text{for } x \geq x_{\min}$$

and

$$g(x) := \frac{x}{h(x_{\min})} \quad \text{for } x < x_{\min}.$$

Let

$$\Delta_t := g(X_t) - g(X_{t+1}) = \int_{X_{t+1}}^{X_t} \frac{1}{h(\max\{x_{\min}, x\})} dx$$

and assume $X_t \geq x_{\min}$. We now prove statement (i). Since h is non-increasing, g is concave, and using Jensen’s inequality, we have (suppressing \mathcal{F}_t)

$$\mathbb{E}[\Delta_t] \geq g(X_t) - g(\mathbb{E}[X_{t+1}]) \geq \frac{\mathbb{E}[\Delta_t]}{h(X_t)} \geq 1$$

by our assumption $\mathbb{E}[\Delta_t] \geq h(X_t)$. To satisfy condition (iii) of Theorem 3.2, we note, still on $X_t \geq x_{\min}$, that

$$\begin{aligned} \mathbb{E}[e^{-\eta\Delta_t}] &= 1 - \eta\mathbb{E}[\Delta_t] + \sum_{k=2}^{\infty} \frac{\eta^k \mathbb{E}[\Delta_t^k]}{k!} \\ &\leq 1 - \eta\mathbb{E}[\Delta_t] + \eta^2 \sum_{k=2}^{\infty} \frac{\eta^{k-2} \mathbb{E}[|\Delta_t|^k]}{k!} \\ &\leq 1 - \eta\mathbb{E}[\Delta_t] + \eta^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2} \mathbb{E}[|\Delta_t|^k]}{k!} \\ &= 1 - \eta + \frac{\eta^2}{\lambda^2} (e^{\lambda Z} - \lambda\mathbb{E}[Z] - 1), \end{aligned}$$

where we have used $\mathbb{E}[\Delta_t] \geq 1$ and $\lambda \geq \eta$. Since $|\Delta_t| \leq Z$, also $\mathbb{E}[Z] \geq 1$. Using $e^{\lambda Z} = D$ and $\eta \leq \delta\lambda^2/(D - 1 - \lambda)$, we obtain

$$\mathbb{E}[e^{-\eta\Delta_t}] \leq 1 - \eta + \delta\eta = 1 - (1 - \delta)\eta \leq e^{-\eta(1-\delta)}.$$

Setting $\beta_u := e^{-\eta(1-\delta)}$ and using η as the λ of Theorem 3.2 proves statement (i).

For statement (ii), analogous calculations prove

$$\mathbb{E}[e^{\eta\Delta_t}] \leq 1 + (1 + \delta)\eta \leq e^{\eta(1+\delta)}.$$

We set $\beta_\ell := e^{\eta(1+\delta)}$, use η as the λ of Theorem 3.2(iv) and note that

$$\frac{e^{\lambda(1+\delta)t^*} - e^{\lambda(1+\delta)}}{e^{\lambda(1+\delta)} - 1} \leq \frac{e^{\lambda(1+\delta)t^*}}{\lambda(1 + \delta)},$$

which was to be proved. If additionally an absorbing state 0 is assumed, the stronger upper bound follows from the corresponding statement in Theorem 3.2(iv). □

4. Applications of the tail bounds

We now show that Theorem 3.2 together with the function g defined explicitly in Remark 3.1 constitute a general and precise tool for analysis of stochastic processes. It provides sharp tail bounds on the running time of randomized search heuristics, which were not previously obtained by drift analysis, as well as tail bounds on random recursions, such as those in analysis of random permutations (see Section 4.2). Most existing drift theorems, including an existing result proving tail bounds with multiplicative drift, can be phrased as special cases of the general drift theorem; see [29] for examples. Relatively recently, in [23], different tail bounds were proved for the scenario of additive drift using classical concentration inequalities such as Azuma–Hoeffding bounds. These bounds are not directly comparable to the ones from our general drift theorem; they are more specific but yield even stronger exponential bounds.

Algorithm 1 (1+1) Evolutionary Algorithm (EA)

Choose uniformly at random an initial bit string $x_0 \in \{0, 1\}^n$.
for $t := 0$ **to** ∞ **do**
 Create x' by flipping each bit in x_t i.i.d. with probability $1/n$ (*mutation*).
 $x_{t+1} := x'$ if $f(x') \geq f(x_t)$, and $x_{t+1} := x_t$ otherwise (*selection*).
end for

We first give sharp tail bounds on the optimization time of the (1 + 1) EA (see Algorithm 1) which maximizes pseudo-Boolean functions $f: \{0, 1\}^n \rightarrow \mathbb{R}$. The optimization time is defined in the canonical way at the smallest t such that x_t is an optimum. We consider classical benchmark problems from the theory of RSHs. Despite their simplicity, their analysis has turned out to be surprisingly difficult and research is still ongoing.

4.1 OneMax, linear functions and LeadingOnes

A simple pseudo-Boolean function is given by $\text{ONEMAX}(x_1, \dots, x_n) = x_1 + \dots + x_n$. It is included in the class of so-called linear functions $f(x_1, \dots, x_n) = w_1x_1 + \dots + w_nx_n$, where $w_i \in \mathbb{R}$ for $1 \leq i \leq n$. We start by citing very precise bounds on the expected optimization time of the (1 + 1) EA on ONEMAX and then prove the new tail bounds. The lower bounds obtained will imply results for all linear functions.

Theorem ([17]). *The expected optimization time of the (1 + 1) EA on ONEMAX is*

$$en \ln n - c_1n + (e/2) \ln n + c_2 + O((\log n)/n),$$

where $c_1 = 1.892541 \dots$ and $c_2 = 0.597899 \dots$ are explicitly computable constants.

We now derive the sharp tail bounds. The following upper concentration inequality in Theorem 4.1 is not new but is already implicit in the classical work on multiplicative drift analysis [9, 7]. A similar upper bound is even available for all linear functions [40]. By contrast, the lower concentration inequality is a novel and non-trivial result.

Theorem 4.1. *The optimization time of the (1 + 1) EA on ONEMAX is at least $en \ln n - cn - ren$, where c is a constant, with probability at least $1 - e^{-r/2}$ for any $r \geq 0$. It is at most $en \ln n + ren$ with probability at least $1 - e^{-r}$.*

Proof of Theorem 4, upper tail. The upper tail is well known and can be easily derived from the multiplicative drift theorem [7]. Let X_t denote the number of zeros at time t . Since

$$\mathbb{E}[X_t - X_{t+1} \mid X_t] \geq \left(\frac{X_t}{n}\right) \left(1 - \frac{1}{n}\right)^{n-1} \geq \frac{X_t}{(en)},$$

one can choose $\delta := 1/(en)$ as the parameter of the multiplicative drift theorem. Then the upper bound follows since $X_0 \leq n$ and $x_{\min} = 1$. □

We now consider the lower tail. The aim is to prove it using Theorem 3.2(iv), which includes a bound on the moment-generating function of the drift of g . We first set up the h (and thereby the g) used for our purposes. The following lemma bounds the drift and prepares the definition of h , which is given in the subsequent Lemma 4.2.

Lemma 4.1. *Let X_t denote the number of zeros of the current search point of the (1 + 1) EA on ONEMAX. Then*

$$\left(1 - \frac{1}{n}\right)^{n-i} \frac{i}{n} \leq \mathbb{E}[X_t - X_{t+1} \mid X_t = i] \leq \left(\left(1 - \frac{1}{n}\right) \left(1 + \frac{i}{(n-1)^2}\right)\right)^{n-i} \frac{i}{n}.$$

Proof. The lower bound considers the expected number of flipping zero-bits, assuming that no one-bit flips. The upper bound is obtained in the proof of Lemma 6 in [6] and is denoted by $S_1 \cdot S_2$, but is not made explicit in the statement of the lemma. \square

Lemma 4.2. Consider the $(1 + 1)$ EA on ONEMAX and let the random variable X_t denote the current number of zeros at time $t \geq 0$. Then

$$h(i) := \exp\left(-1 + 2\frac{\lceil i \rceil}{n}\right) \cdot \left(\frac{\lceil i \rceil}{n}\right) \cdot \left(1 + \frac{c^*}{n}\right),$$

where $c^* > 0$ is a sufficiently large constant, satisfies the condition

$$\mathbb{E}[X_t - X_{t+1} \mid X_t = i] \leq h(i) \quad \text{for } i \in [n] := \{1, \dots, n\}.$$

Moreover, with $x_{\min} := 1$, define

$$g(i) := \frac{\min(i, x_{\min})}{h(x_{\min})} + \int_{x_{\min}}^{\max(i, x_{\min})} \frac{1}{h(y)} dy$$

and $\Delta_t := g(X_t) - g(X_{t+1})$. Then, for $i \in [n]$,

$$g(i) = \sum_{j=1}^i \frac{1}{h(j)} \quad \text{and} \quad \Delta_t \leq \sum_{j=X_{t+1}+1}^{X_t} e^{1-\frac{2X_{t+1}}{n}} \cdot \left(\frac{n}{j}\right).$$

Proof. According to Lemma 4.1,

$$h^*(i) := \left(\left(1 - \frac{1}{n}\right)\left(1 + \frac{i}{(n-1)^2}\right)\right)^{n-i} \frac{i}{n}$$

is an upper bound on the drift. For some sufficiently large constant $c^* > 0$ we have

$$h^*(i) \leq e^{-1+i/n+(i(n-i))/n^2} \cdot \frac{i}{n} \cdot \left(\frac{1+i/(n-1)^2}{1+i/n^2}\right)^{n-i} \leq e^{-1+2i/n} \cdot \frac{i}{n} \cdot \left(1 + \frac{c^*}{n}\right) = h(i),$$

where we used $1 + x \leq e^x$ twice. Therefore $\mathbb{E}[X_t - X_{t+1} \mid X_t = i] \leq h(i)$.

The representation of $g(i)$ as a sum follows immediately from h due to the ceilings. The bound on Δ_t follows from h by estimating

$$e^{-1+2\lceil i \rceil/n} \cdot \left(1 + \frac{c^*}{n}\right) \geq e^{-1+2i/n}. \quad \square$$

The next lemma provides a bound on the MGF of the drift of g , which will depend on the current state. Later the state will be estimated based on the current point of time, leading to a time-dependent bound on the MGF. Note that we do not need the whole natural filtration based on X_0, \dots, X_t but only X_t , since we have a Markov chain.

Lemma 4.3. Let $\lambda := 1/(en)$ and $i \in [n]$. Then

$$\mathbb{E}[e^{\lambda \Delta_t} \mid X_t = i] \leq 1 + \lambda + \frac{2\lambda}{i} + o\left(\frac{\lambda}{\log n}\right).$$

Proof. We distinguish between three major cases.

Case 1. $i = 1$. Then $X_{t+1} = 0$, implying $\Delta_t \leq en$, with probability

$$(1/n)(1 - 1/n)^{n-1} = (1/(en))(1 + 1/(n-1))$$

and $X_{t+1} = i$ otherwise. We get

$$\begin{aligned} \mathbb{E}[e^{\lambda \Delta_t} \mid X_t = i] &\leq \frac{1}{en} \cdot e^1 + \left(1 - \frac{1}{en}\right) + O\left(\frac{1}{n^2}\right) \\ &\leq 1 + \frac{e-1}{en} + O\left(\frac{1}{n^2}\right) \\ &\leq 1 + \lambda + \frac{(e-2)\lambda}{i} + o\left(\frac{\lambda}{\ln n}\right). \end{aligned}$$

Case 2. $2 \leq i \leq \ln^3 n$. Let $Y := i - X_{t+1}$ and note that $\mathbb{P}[Y \geq 2] \leq (\ln^6 n)/n^2$ since a zero-bit flips with probability at most $(\ln^3 n)/n$. We consider two subcases with respect to Y .

Case 2a. $2 \leq i \leq \ln^3 n$ and $Y \geq 2$. The largest value of Δ_t is taken when $Y = i$. Using Lemma 4.2 and estimating the i th harmonic number, we have $\lambda \Delta_t \leq (\ln i) + 1 \leq 3(\ln \ln n) + 1$. The contribution to the MGF is bounded by

$$\mathbb{E}[e^{\lambda \Delta_t} \cdot \mathbb{1}[X_{t+1} \leq i - 2] \mid X_t = i] \leq e^{3 \ln \ln n + 1} \cdot \left(\frac{\ln^6 n}{n^2}\right) = o\left(\frac{\lambda}{\ln n}\right).$$

Case 2b. $2 \leq i \leq \ln^3 n$ and $Y < 2$. Then $X_{t+1} \geq X_t - 1$, which implies

$$\Delta_t \leq en(\ln(X_t) - \ln(X_{t+1})).$$

We obtain

$$\mathbb{E}[e^{\lambda \Delta_t} \cdot \mathbb{1}[X_{t+1} \geq i - 1] \mid X_t = i] \leq \mathbb{E}[e^{\ln(i/X_{t+1})}] \leq \mathbb{E}[e^{\ln(1+(i-X_{t+1})/(i-1))}] = \mathbb{E}\left[1 + \frac{Y}{i-1}\right],$$

where the first inequality estimated

$$\sum_{i=j+1}^k \frac{1}{i} \leq \ln(k/j)$$

and the second one used $X_{t+1} \geq i - 1$. From Lemma 4.1, we get

$$\mathbb{E}[Y] \leq \frac{i}{en} \left(1 + O\left(\frac{\ln^3 n}{n}\right)\right) \quad \text{for } i \leq \ln^3 n.$$

This implies

$$\begin{aligned} \mathbb{E}\left[1 + \frac{i - X_{t+1}}{i - 1}\right] &\leq 1 + \frac{i}{en(i-1)} \left(1 + O\left(\frac{\ln^3 n}{n}\right)\right) \\ &= 1 + \frac{1}{en} \cdot \left(1 + \frac{1}{i-1}\right) \left(1 + O\left(\frac{\ln^3 n}{n}\right)\right) \\ &= 1 + \lambda + \frac{2\lambda}{i} + o\left(\frac{\lambda}{\ln n}\right), \end{aligned}$$

using $i/(i-1) \leq 2$ in the last step. Adding the bounds from the two subcases proves the lemma in Case 2.

Case 3. $i > \ln^3 n$. Note that

$$\mathbb{P}[Y \geq \ln n] \leq \binom{n}{\ln n} \left(\frac{1}{n}\right)^{\ln n} \leq \frac{1}{(\ln n)!}.$$

We further subdivide the case according to whether or not $Y \geq \ln n$.

Case 3a. $i > \ln^3 n$ and $Y \geq \ln n$. Since $\Delta_t \leq en(\ln n + 1)$, we get

$$\mathbb{E}[e^{\lambda \Delta_t} \cdot \mathbb{1}[X_{t+1} \leq i - \ln^3 n] \mid X_t = i] \leq \frac{1}{(\ln n)!} \cdot e^{\ln n + 1} = o\left(\frac{\lambda}{\ln n}\right).$$

Case 3b. $i > \ln^3 n$ and $Y < \ln n$. Then, using Lemma 4.2 and proceeding as in Case 2b,

$$\begin{aligned} \mathbb{E}[e^{\lambda \Delta_t} \cdot \mathbb{1}[X_{t+1} > i - \ln n] \mid X_t = i] &\leq \mathbb{E}[e^{\lambda \exp(1-2(i-\ln n)/n) \cdot n \ln(i/X_{t+1})} \mid X_t = i] \\ &= \mathbb{E}\left[\left(1 + \frac{i - X_{t+1}}{X_{t+1}}\right)^{\exp((-2i + \ln n)/n)}\right]. \end{aligned}$$

Using $i > \ln^3 n$ and Jensen’s inequality, the last expectation is at most

$$\begin{aligned} \left(1 + \mathbb{E}\left[\frac{i - X_{t+1}}{X_{t+1}}\right]\right)^{\exp((-2i + \ln n)/n)} &\leq \left(1 + \mathbb{E}\left[\frac{Y}{i - \ln n}\right]\right)^{\exp((-2i + \ln n)/n)} \\ &\leq \left(1 + \mathbb{E}\left[\frac{Y}{i(1 - 1/\ln^2 n)}\right]\right)^{\exp((-2i + \ln n)/n)}, \end{aligned}$$

where the last inequality used $i > \ln^3 n$. Since

$$\mathbb{E}[Y] \leq e^{-1+2i/n} \frac{i}{n} \left(1 + \frac{c^*}{n}\right),$$

we conclude

$$\begin{aligned} \mathbb{E}[e^{\lambda \Delta_t} \cdot \mathbb{1}[X_{t+1} > i - \ln n] \mid X_t = i] &\leq \left(1 + \frac{e^{2i/n}}{en(1 - 1/\ln^2 n)}\right)^{\exp((-2i + \ln n)/n)} \\ &\leq \left(1 + \frac{1}{en(1 - 1/\ln^2 n)}\right) \left(1 + O\left(\frac{\ln n}{n^2}\right)\right) \\ &\leq 1 + \lambda + o\left(\frac{\lambda}{\ln n}\right), \end{aligned}$$

where we used $(1 + ax)^{1/a} \leq 1 + x$ for $x \geq 0$ and $a \geq 1$. Adding up the bounds from the two subcases, we have proved the lemma in Case 3.

Altogether, for all $i \in [n]$,

$$\mathbb{E}[e^{\lambda \Delta_t} \mid X_t = i] \leq 1 + \lambda + \frac{2\lambda}{i} + o\left(\frac{\lambda}{\ln n}\right). \quad \square$$

The bound on the MGF of Δ_t derived in Lemma 4.3 is particularly large for $i = O(1)$, i.e. if the current state X_t is small. If $X_t = O(1)$ held during the whole optimization process, we could not prove the lower tail in Theorem 4.1 from the lemma. However, it is easy to see that $X_t = i$ only holds for an expected number of at most en/i steps. Hence most of the time the term $2\lambda/i$ is negligible, and the time-dependent $\beta_\ell(t)$ -term from Theorem 3.2(iv) comes into play. We make this precise in the following proof, where we iteratively bound the probability of the process being at ‘small’ states.

Proof of Theorem 4.1, lower tail. With overwhelming probability $1 - 2^{-\Omega(n)}$ due to Chernoff bounds, $X_0 \geq (1 - \varepsilon)n/2$ for an arbitrarily small constant $\varepsilon > 0$, which we assume to happen. We consider phases in the optimization process. Phase 1 starts with initialization and ends before the first step where $X_t < e^{(\ln n - 1)/2} = \sqrt{n} \cdot e^{-1/2}$. Phase i , where $i > 1$, follows phase $i - 1$ and ends before the first step where $X_t < \sqrt{n} \cdot e^{-i/2}$. Obviously, the optimum is not found before the end of phase $\ln(n)$; however, this does not yet tell us anything about the optimization time.

Phase i is called *typical* if it does not end before time $eni - 1$. We will prove inductively that the probability of one of the first i phases not being typical is at most $c'e^{i/2}/\sqrt{n} = c'e^{(i-\ln n)/2}$ for some constant $c' > 0$. This implies the theorem since an optimization time of at least $en \ln n - cn - ren$ is implied by the event that phase $\ln n - \lceil r - c/e \rceil$ is typical, which has probability at least

$$1 - c'e^{(-r+c/e+1)/2} = 1 - e^{-r/2} \quad \text{for } c = e(2 \ln c' + 1).$$

Fix some $k > 1$ and assume for the moment that all the first $k - 1$ phases are typical. Then for $1 \leq i \leq k - 1$, we have $X_t \geq \sqrt{ne}^{-i/2}$ in phase i , i.e. when $en(i - 1) \leq t \leq eni - 1$. We analyse the event that additionally phase k is typical, which subsumes the event that $X_t \geq \sqrt{ne}^{-k/2}$ throughout phase k . According to Lemma 4.3, we get in phase $i \in [k]$

$$\mathbb{E}[e^{\lambda \Delta_t} | X_t] \leq 1 + \lambda + 2\lambda e^{i/2}/\sqrt{n} + o\left(\frac{\lambda}{\ln n}\right) \leq \exp\left(\lambda + \frac{2\lambda e^{i/2}}{\sqrt{n}} + o\left(\frac{\lambda}{\ln n}\right)\right).$$

The expression now depends on the time only, so for $\lambda := 1/(en)$

$$\begin{aligned} \prod_{t=0}^{enk-1} \mathbb{E}[e^{\lambda \Delta_t} | X_0] &\leq \exp\left(\lambda enk + \frac{2\lambda en}{\sqrt{n}} \sum_{i=1}^k e^{i/2} + enk \cdot o\left(\frac{\lambda}{\ln n}\right)\right) \\ &\leq \exp\left(k + \frac{6e^{k/2}}{n\sqrt{n}} + o(1)\right) \\ &\leq e^{k+o(1)}, \end{aligned}$$

using that $k \leq \ln n$. By Theorem 3.2(iv) for $a = \sqrt{ne}^{-k/2}$ and $t = enk - 1$, we obtain

$$\mathbb{P}[T_a < t] \leq e^{k+o(1)-\lambda(g(X_0)-g(\sqrt{ne}^{-k/2}))}.$$

It is easy to see that $g(X_0) \geq en \ln n - c''n$ for some constant $c'' > 0$ (which is assumed large enough to subsume the $-O(\log n)$ term). Moreover, $g(x) \leq en(\ln x + 1)$ according to Lemma 4.2. We get

$$\mathbb{P}[T_a < t] \leq e^{k+o(1)-\ln n+O(1)-k/2+(\ln n)/2} = e^{(k-\ln n+O(1))/2} = c''' e^{k/2}/\sqrt{n}$$

for some sufficiently large constant $c''' > 0$, which proves the bound on the probability of phase k not being typical (without making statements about the earlier phases). The probability that all phases up to and including phase k are typical is at least

$$1 - \frac{\left(\sum_{i=1}^k c''' e^{i/2}\right)}{\sqrt{n} \geq 1} = c' \frac{e^{k/2}}{\sqrt{n}} \quad \text{for a constant } c' > 0. \quad \square$$

We now deduce a concentration inequality with respect to linear functions, essentially depending on all variables, i.e. functions of the kind $f(x_1, \dots, x_n) = w_1x_1 + \dots + w_nx_n$, where $w_i \neq 0$. This intensively studied function class contains ONEMAX [40].

Theorem 4.2. *The optimization time of the $(1 + 1)$ EA on any linear function with non-zero weights is at least $en \ln n - cn - ren$, where c is a constant, with probability at least $1 - e^{-r/2}$ for any $r \geq 0$. It is at most $en \ln n + (1 + r)en + O(1)$ with probability at least $1 - e^{-r}$.*

Proof. The upper tail is proved in Theorem 5.1 in [40]. The lower bound follows from the lower tail in Theorem 4.1 and the fact that the optimization time within the class of linear functions is stochastically smallest for ONEMAX (Theorem 6.2 in [40]). \square

Finally we consider the $(1 + 1)$ EA on $\text{LEADINGONES}(x_1, \dots, x_n) := \sum_{i=1}^n \prod_{j=1}^i x_j$, another intensively studied standard benchmark problem from the analysis of RSHs. Tail bounds on the optimization time of the $(1 + 1)$ EA on LEADINGONES were derived in [8]. This result represents a fundamentally new contribution, but suffers from the fact that it depends on a very specific structure and closed formula for the optimization time. Using a simplified version of Theorem 3.2 (see Theorem 3.3), it is possible to prove similarly strong tail bounds without needing this exact formula. As in [8], we are interested in a more general statement. Let $T(a)$ be the number of steps until the $(1 + 1)$ EA has reached a LEADINGONES -value of at least a , where $0 \leq a \leq n$. Let $X_t := \max\{0, a - \text{LEADINGONES}(x_t)\}$ be the distance from the target a at time t . Lemma 4.4 states the drift of $(X_t)_{t \in \mathbb{N}_0}$ exactly; see also [8].

Lemma 4.4. For all $i > 0$,

$$\mathbb{E}[X_t - X_{t+1} \mid X_t = i] = (2 - 2^{-n+a-i+1}) \left(\frac{1 - 1}{n}\right)^{a-i} \left(\frac{1}{n}\right).$$

Proof. The leftmost zero-bit is at position $a - i + 1$. To increase the LEADINGONES -value (it cannot decrease), it is necessary to flip this bit and not to flip the first $a - i$ bits, which is reflected by the last two terms in the lemma. The first term is due to the expected number of free-rider bits (a sequence of previously random bits after the leftmost zero that happen to be all 1 at the time of improvement). Note that there can be between 0 and $n - a + i - 1$ such bits. By the usual argumentation using a geometric distribution [11], the expected number of free-riders in an improving step equals

$$\sum_{k=0}^{n-a+i-1} k \cdot \left(\frac{1}{2}\right)^{\min\{n-a+i-1, k+1\}} = 1 - 2^{-n+a-i+1},$$

hence the expected progress in an improving step is $2 - 2^{-n+a-i+1}$. □

Statements (ii) and (iii) of the following theorem provide the announced tail bounds. For completeness, statement (i) states the expected hitting time derived in [8].

Theorem 4.3. Let $T(a)$ the time for the $(1 + 1)$ EA to reach a LEADINGONES -value of at least a . Moreover, let $r \geq 0$. Then we have the following.

(i) $\mathbb{E}[T(a)] = \frac{n^2 - n}{2} \left(\left(1 + \frac{1}{n - 1}\right)^a - 1 \right).$

(ii) For $0 < a \leq n - \log n$, with probability at least $1 - e^{-\Omega(rn^{-3/2})}$,

$$T(a) \leq \frac{n^2}{2} \left(\left(1 + \frac{1}{n - 1}\right)^a - 1 \right) + r.$$

(iii) For $\log^2 n - 1 \leq a \leq n$, with probability at least $1 - e^{-\Omega(rn^{-3/2})} - e^{-\Omega(\log^2 n)}$,

$$T(a) \geq \frac{n^2 - n}{2} \left(\left(1 + \frac{1}{n - 1}\right)^a - 1 - \frac{2 \log^2 n}{n} \right) - r.$$

Proof. Statement (i) is already contained in [8], so we turn to (ii). From Lemma 4.4, $h(x) = (2 - 2/n)(1 - 1/n)^{a-x}/n$ is a non-decreasing lower bound on the drift $\mathbb{E}[X_t - X_{t+1} \mid X_t = x]$ if $x \geq \log n$. To bound the change of the g -function, we observe that $h(x) \geq 1/(en)$ for all $x \geq 1$. This means that $X_t - X_{t+1} = k$ implies $g(X_t) - g(X_{t+1}) \leq enk$. Moreover, to change the LEADINGONES -value by k , it is necessary that the first zero-bit flips (which has probability $1/n$), and $k - 1$ free-riders occur. The change only gets stochastically larger if we assume an infinite supply of

free-riders. Hence $g(X_t) - g(X_{t+1})$ is stochastically dominated by $Z = enY$, where Y is 0 with probability $1 - 1/n$ and, follows the geometric distribution with parameter $1/2$ otherwise. Thus the MGF of Y equals

$$\mathbb{E}[e^{\lambda Y}] = \left(1 - \frac{1}{n}\right)e^0 + \frac{1}{n} \cdot \frac{1/2}{e^{-\lambda} - (1 - 1/2)} \leq 1 + \frac{1}{n(1 - 2\lambda)},$$

where we have used $e^{-\lambda} \geq 1 - \lambda$. For the MGF of Z it follows that

$$\mathbb{E}[e^{\lambda Z}] = \mathbb{E}[e^{\lambda enY}] \leq 1 + \frac{1}{n(1 - 2en\lambda)}.$$

For $\lambda := 1/(4en)$ we get

$$D := \mathbb{E}[e^{\lambda Z}] = 1 + \frac{2}{n} = 1 + 8e\lambda,$$

i.e. $D - 1 - \lambda = (8e - 1)\lambda$. We get

$$\eta := \frac{\delta\lambda^2}{D - 1 - \lambda} = \frac{\delta\lambda}{8e - 1} = \frac{\delta}{4en(8e - 1)}$$

(which is less than λ if $\delta \leq 8e - 1$). Choosing $\delta := n^{-1/2}$, we obtain $\eta = Cn^{-3/2}$ for $C := 1/((8e - 1)(4e))$.

We set

$$t := \left(\int_{x_{\min}}^{X_0} \frac{1}{h(x)} dx + r\right)/(1 - \delta)$$

in statement (i) of Theorem 3.3. The integral within t can be bounded according to

$$\begin{aligned} U &:= \int_{x_{\min}}^{X_0} \frac{1}{h(x)} dx \\ &\leq \sum_{i=1}^a \frac{1}{(2 - 2/n)(1 - 1/n)^{a-i}/n} \\ &= \left(\frac{1}{2} + \frac{1}{2n - 2}\right) \cdot n \cdot \frac{(1 + 1/(n - 1))^a - 1}{1/(n - 1)} \\ &= \frac{n^2}{2} \left(\left(1 + \frac{1}{n - 1}\right)^a - 1\right). \end{aligned}$$

Hence, using the theorem, we get

$$\mathbb{P}[T > t] = \mathbb{P}[T > (U + r)/(1 - \delta)] \leq e^{-\eta r} \leq e^{-Cn^{-3/2}}.$$

Since $U \leq en^2$ and $1/(1 - \delta) \leq 1 + 2\delta = 1 + 2n^{-1/2}$, we get

$$\mathbb{P}[T \geq U + 2en^{3/2} + 2r] \leq e^{-Cn^{-3/2}}.$$

Using the upper bound on U derived above, we obtain as suggested

$$\mathbb{P}\left[T \geq \frac{n^2}{2} \left(\left(1 + \frac{1}{n - 1}\right)^a - 1\right) + r\right] = e^{-\Omega(n^{-3/2})}.$$

Finally, we prove statement (iii) of this theorem in a rather symmetrical way to statement (ii). We can choose $h(x) := 2(1 - 1/n)^{a-x}/n$ as an upper bound on the drift $\mathbb{E}[X_t - X_{t+1} | X_t = x]$. The estimation of the $\mathbb{E}[e^{\lambda Z}]$ still applies. We set

$$t := \left(\int_{x_{\min}}^{X_0} \frac{1}{h(x)} dx - r\right)/(1 - \delta).$$

Moreover, we assume $X_0 \geq n - \log^2 n - 1$, which happens with probability at least $1 - e^{-\Omega(\log^2 n)}$. Note that

$$\begin{aligned} L &:= \int_{x_{\min}}^{X_0} \frac{1}{h(x)} dx \\ &\geq \sum_{i=1}^{a-\log^2 n} \frac{1}{2(1-1/n)^{a-i}/n} \\ &= \frac{n^2-n}{2} \left(\left(1 + \frac{1}{n-1}\right)^a - \left(1 + \frac{1}{n-1}\right)^{\log^2 n} \right) \\ &\geq \frac{n^2-n}{2} \left(\left(1 + \frac{1}{n-1}\right)^a - 1 - \frac{\log^2 n}{n} \right), \end{aligned}$$

where the last inequality used $e^x \leq 1 + 2x$ for $x \leq 1$ and $e^x \geq 1 + x$ for $x \in \mathbb{R}$. Statement (ii) of Theorem 3.3 yields (since state 0 is absorbing)

$$\mathbb{P}[T < t] = \mathbb{P}[T < (L - r)/(1 + \delta)] \leq e^{-\eta r} \leq e^{-Crn^{-3/2}}.$$

Now, since

$$\frac{L - r}{1 + \delta} \geq (L - r) - \delta(L - r) \geq L - r - en^{3/2}$$

(using $L \leq en^2$), we get statement (iii) by calculations analogous to those above. □

4.2 An application to probabilistic recurrence relations

Drift analysis is not only useful in the theory of RSHs, but also in classical computer science. Here we study the probabilistic recurrence relation $T(n) = a(n) + T(h(n))$, where n is the problem size, $a(n)$ is the amount of work at the current level of recursion, and $h(n)$ is a random variable denoting the size of the problem at the next recursion level. The asymptotic distribution (letting $n \rightarrow \infty$) of the number of cycles is well studied [1], but there are few results for finite n . Karp [22] studied this scenario using probabilistic techniques different from ours. Assuming knowledge of $\mathbb{E}[h(n)]$, he proved upper tail bounds for $T(n)$; more precisely, he analysed the probability of $T(n)$ exceeding the solution of the ‘deterministic’ process $T(n) = a(n) + T(\mathbb{E}[h(n)])$.

We pick up the example from [22, Section 2.4] on the number of cycles in a permutation $\pi \in S_n$ drawn uniformly at random, where S_n denotes the set of all permutations of the n elements $[n]$. A cycle is a subsequence of indices i_1, \dots, i_ℓ such that $\pi(i_j) = i_{(j \bmod \ell) + 1}$ for $1 \leq j \leq \ell$. Each permutation partitions the elements into disjoint cycles. The expected number of cycles in a random permutation is $H_n = \ln n + \Theta(1)$. Moreover, the length of the cycle containing any fixed element is uniform on $[n]$. This leads to the probabilistic recurrence $T(n) = 1 + T(h(n))$ for the random number of cycles, where $h(n)$ is uniform on $\{0, \dots, n - 1\}$. As a result, [22] shows that the number of cycles is larger than $\log_2(n + 1) + a$ with probability at most 2^{-a+1} . Note that $\log_2(n)$, which is the solution of the deterministic recurrence, is by a constant factor away from the expected value. Lower tail bounds are not obtained in [22]. However, our drift theorem implies that the number of cycles is sharply concentrated around its expectation.

Theorem 4.4. *Let N be the number of cycles in a random permutation of $[n]$. Then*

$$\mathbb{P}[N < (1 - \varepsilon)(\ln n)] \leq \exp\left(-\frac{\varepsilon^2}{4}(1 - o(1)) \ln n\right)$$

for any constant $0 < \varepsilon < 1$. And for any constant $\varepsilon > 0$,

$$\mathbb{P}[N \geq (1 + \varepsilon)((\ln n) + 1)] \leq \exp\left(-\frac{\min\{\varepsilon, \varepsilon^2\}}{6} \ln n\right).$$

Proof. We regard the recurrence as a stochastic process, where $X_t, t \geq 0$, denotes the number of elements not yet included in a cycle, and $X_0 = n$. If $X_t = i$ then X_{t+1} is uniform on $\{0, \dots, i - 1\}$ [22]. Note that N equals the first hitting time for $X_t = 0$, which is denoted by T_0 in our notation. Clearly N is stochastically larger than T_a for any $a > 0$.

We now prove the lower tail using Theorem 3.2(iv). We compute $\mathbb{E}[X_{t+1} | X_t] = (X_t - 1)/2$, which means

$$\mathbb{E}[X_t - X_{t+1} | X_t] \geq \frac{X_t}{2} = \frac{\lceil X_t \rceil}{2}$$

since X_t is integral. Therefore we choose $h(x) = \lceil x \rceil / 2$. Letting $x_{\min} = 1$, we obtain the drift function

$$g(i) = 2 + \int_1^i \frac{2}{\lceil j \rceil} dj = \sum_{j=1}^i 2/j \quad \text{for } i \geq 1 \text{ and } g(0) = 0.$$

For the drift theorem, we have to compute $g(i) - g(X_{t+1})$, given $X_t = i$, and to bound the MGF with respect to this difference. Using that $\sum_{j=1}^k j^{-1} - \ln k$ is a decreasing sequence, we get

$$g(i) - g(X_{t+1}) \leq \begin{cases} 2(\ln(i) - \ln(j)) & \text{for } j = 1, \dots, i - 1, \text{ each with probability } 1/i, \\ 2(\ln(i) + 1) & \text{with probability } 1/i. \end{cases}$$

Let $X_t = i$. For $\lambda > 0$, we bound the MGF

$$\mathbb{E}[e^{\lambda(g(i) - g(X_{t+1}))}] \leq \frac{1}{i} \cdot e^{2\lambda} e^{2\lambda \ln i} + \frac{1}{i} \sum_{j=1}^{i-1} e^{2\lambda(\ln i - \ln j)} = \frac{1}{i} e^\eta i^\eta + \frac{1}{i} i^\eta \sum_{j=1}^{i-1} j^{-\eta},$$

where $\eta = 2\lambda$. Now assume η constant and $\eta < 1$. Then

$$\begin{aligned} \mathbb{E}[e^{\lambda(g(i) - g(X_{t+1}))}] &\leq i^{\eta-1} e^\eta + i^{\eta-1} \left(1 + \int_1^{i-1} j^{-\eta} dj\right) \\ &\leq i^{\eta-1} e^\eta + i^{\eta-1} \left(1 + \left(\frac{1}{1-\eta} ((i-1)^{1-\eta} - 1)\right)\right) \\ &\leq i^{\eta-1} (e^\eta + 1) + \frac{1}{1-\eta} - i^{\eta-1} \\ &= i^{\eta-1} e^\eta + \frac{1}{1-\eta} \\ &= 1 + i^{\eta-1} e^\eta + \frac{\eta}{1-\eta} \\ &\leq e^{e^\eta i^{\eta-1} + \eta/(1-\eta)} \\ &=: \beta \end{aligned}$$

using $1 + x \leq e^x$. The factor $e^{e^\eta i^{\eta-1}}$ will be negligible (more precisely, $e^{O((\ln n)^{\eta-1})}$) for $i \geq \ln n$ in the following, which is why we set $a := \ln n$ in Theorem 3.2(iv).

From the theorem, we get $\mathbb{P}[T_a < t] \leq \beta^t e^{-\lambda(g(X_0)-g(a))}$. We work with the lower bound

$$g(X_0) - g(a) = \sum_{j=a}^n \frac{2}{j} \geq 2(\ln(n+1) - \ln(a+1)),$$

which yields

$$\begin{aligned} \mathbb{P}[T_a < t] &< \beta^t e^{-\lambda(2(\ln(n+1)-\ln(a+1)))} \\ &= \beta^t e^{-\eta \ln n + O(\ln \ln n)} \\ &= \exp\left(O(t(\ln n)^{\eta-1}) + \frac{\eta}{1-\eta}t - \eta \ln n + O(\ln \ln n)\right) \\ &= \exp\left(o(t) + O(\ln \ln n) + \frac{\eta}{1-\eta}t - \eta \ln n\right). \end{aligned}$$

Now we concentrate on the difference $d(\varepsilon) = (\eta/(1-\eta))t - \eta \ln n$ that is crucial for the order of growth of the last exponent. We assume $t := (1-\varepsilon) \ln n$ for some constant $\varepsilon > 0$ and set $\eta := \varepsilon/2$ (implying $\varepsilon < 2$); hence $\lambda = \varepsilon/4$. We get

$$d(\varepsilon) = \frac{\varepsilon/2}{1-\varepsilon/2}(1-\varepsilon)(\ln n) - \frac{\varepsilon}{2}(\ln n) = \frac{\varepsilon}{2}(\ln n)\left(\frac{1-\varepsilon}{1-\varepsilon/2} - 1\right) \leq -\frac{\varepsilon^2}{4}(\ln n).$$

Plugging the bound for $d(\varepsilon)$ in the exponent and noting that $\varepsilon > 0$ is constant gives

$$\mathbb{P}[T_a < (1-\varepsilon) \ln n] \leq \exp\left(-\frac{\varepsilon^2}{4}(1-o(1)) \ln n\right),$$

which also bounds T_0 the same way.

To prove the upper tail, we must set $a := 0$ in Theorem 3.2(iii). Using the lower bound on the difference of g -values derived above, we estimate for $X_t = i$ and any $\lambda > 0$

$$\mathbb{E}[e^{-\lambda(g(i)-g(X_{t+1}))}] \leq \frac{1}{i} \sum_{j=0}^{i-1} e^{-\lambda(2(\ln(i+1)-\ln(j+1)))} = \frac{1}{i} \sum_{j=0}^{i-1} \left(\frac{j+1}{i+1}\right)^\eta,$$

where again $\eta = 2\lambda$. Hence, similarly to the estimations for the lower tail,

$$\mathbb{E}[e^{-\lambda(g(i)-g(X_{t+1}))}] \leq \frac{1}{i^{\eta+1}} \int_1^i j^\eta dj \leq \frac{1}{i^{\eta+1}} \frac{1}{\eta+1} i^{\eta+1} = \frac{1}{\eta+1} \leq e^{-\eta/(\eta+1)} =: \beta.$$

From the drift theorem, we get

$$\mathbb{P}[T_0 > t] \leq \beta^t e^{\lambda(g(X_0)-g(0))} \leq e^{-\eta t/(\eta+1)} e^{\lambda(2(\ln(n+1)))} = \exp\left(-\frac{\eta t}{\eta+1} + \eta(\ln n + 1)\right).$$

Setting $t := (1+\varepsilon)(\ln n + 1)$ and $\eta = \varepsilon/2$, the exponent is no more than

$$-\frac{\eta(1+\varepsilon/2+\varepsilon/2)(\ln n + 1)}{1+\varepsilon/2} + \eta(\ln n + 1) \leq -\frac{\varepsilon^2}{4+2\varepsilon}(\ln n + 1).$$

The last fraction is at most $-\varepsilon^2/6$ if $\varepsilon \leq 1$ and at most $-\varepsilon/6$ otherwise (if $\varepsilon > 1$). Altogether,

$$\mathbb{P}[T_0 > t \mid X_0 = n] \leq \exp(-\min\{\varepsilon^2, \varepsilon\}(\ln n + 1)/6). \quad \square$$

5. Conclusions

We have presented a new and versatile drift theorem with tail bounds. It can be understood as a general variable drift theorem and can be specialized into all existing variants of variable, additive

and multiplicative drift theorems we found in the literature as well as the fitness-level technique. Moreover, it provides lower and upper tail bounds, which were not previously available in the context of variable drift. These tail bounds were used to prove sharp-concentration inequalities on the optimization time of the $(1 + 1)$ EA on ONEMAX, linear functions and LEADINGONES. Despite the highly random fashion in which this RSH operates, its optimization time is highly concentrated up to lower-order terms. The drift theorem also leads to tail bounds on the number of cycles in random permutations. We expect further applications of these tail bounds, also to classical randomized algorithms.

References

- [1] Arratia, R. and Tavaré, S. (1992) The cycle structure of random permutations. *Ann. Probab.* **20** 1567–1591.
- [2] Auger, A. and Doerr, B., eds, (2011) *Theory of Randomized Search Heuristics: Foundations and Recent Developments*. World Scientific.
- [3] Coffman, E. G., Feldmann, A., Kahale, N. and Poonen, B. (1999) Computing call admission capacities in linear networks. *Probab. Engrg Inform. Sci.* **13** 387–406.
- [4] Doerr, B., Doerr, C. and Kötzing, T. (2016) The right mutation strength for multi-valued decision variables. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2016)*, pp. 1115–1122. ACM Press.
- [5] Doerr, B., Doerr, C. and Yang, J. (2020) Optimal parameter choices via precise black-box analysis. *Theoret. Comput. Sci.* **801** 1–34.
- [6] Doerr, B., Fouz, M. and Witt, C. (2011) Sharp bounds by probability-generating functions and variable drift. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2011)*, pp. 2083–2090. ACM Press.
- [7] Doerr, B. and Goldberg, L. A. (2013) Adaptive drift analysis. *Algorithmica* **65** 224–250.
- [8] Doerr, B., Jansen, T., Witt, C. and Zarges, C. (2013) A method to derive fixed budget results from expected optimisation times. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2013)*, pp. 1581–1588. ACM Press.
- [9] Doerr, B., Johannsen, D. and Winzen, C. (2012) Multiplicative drift analysis. *Algorithmica* **64** 673–697.
- [10] Doerr, B. and Neumann, F., eds, (2020) *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*. Springer.
- [11] Droste, S., Jansen, T. and Wegener, I. (2002) On the analysis of the $(1 + 1)$ evolutionary algorithm. *Theoret. Comput. Sci.* **276** 51–81.
- [12] Eryilmaz, A. and Srikant, R. (2012) Asymptotically tight steady-state queue length bounds implied by drift conditions. *Queueing Syst. Theory Appl.* **72** 311–359.
- [13] Feldmann, M. and Kötzing, T. (2013) Optimizing expected path lengths with ant colony optimization using fitness proportional update. In *Proceedings of Foundations of Genetic Algorithms (FOGA 2013)*, pp. 65–74. ACM Press.
- [14] Gießen, C. and Witt, C. (2018) Optimal mutation rates for the $(1 + \lambda)$ EA on OneMax through asymptotically tight drift analysis. *Algorithmica* **80** 1710–1731.
- [15] Hajek, B. (1982) Hitting and occupation time bounds implied by drift analysis with applications. *Adv. Appl. Probab.* **14** 502–525.
- [16] He, J. and Yao, X. (2001) Drift analysis and average time complexity of evolutionary algorithms. *Artif. Intell.* **127** 57–85. Erratum in *Artif. Intell.* **140** (2002) 245–248.
- [17] Hwang, H., Panholzer, A., Rolin, N., Tsai, T. and Chen, W. (2018) Probabilistic analysis of the $(1 + 1)$ -evolutionary algorithm. *Evol. Comput.* **26** 299–345.
- [18] Jägersküpper, J. (2011) Combining Markov-chain analysis and drift analysis: the $(1 + 1)$ evolutionary algorithm on linear functions reloaded. *Algorithmica* **59** 409–424.
- [19] Jansen, T. (2013) *Analyzing Evolutionary Algorithms: The Computer Science Perspective*, Natural Computing Series. Springer.
- [20] Jansen, T. (2020) Analysing stochastic search heuristics operating on a fixed budget. In *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization* (B. Doerr and F. Neumann, eds), pp. 249–270. Springer.
- [21] Johannsen, D. (2010) Random combinatorial structures and randomized search heuristics. PhD thesis, Universität des Saarlandes, Germany.
- [22] Karp, R. M. (1994) Probabilistic recurrence relations. *J. Assoc. Comput. Mach.* **41** 1136–1150.
- [23] Kötzing, T. (2016) Concentration of first hitting times under additive drift. *Algorithmica* **75** 490–506.
- [24] Kötzing, T. and Krejca, M. S. (2019) First-hitting times under drift. *Theoret. Comput. Sci.* **796** 51–69.
- [25] Kötzing, T. and Witt, C. (2020) Improved fixed-budget results via drift analysis. In *Proceedings of Parallel Problem Solving from Nature (PPSN 2020)*, Vol. 12270 of Lecture Notes in Computer Science, pp. 648–660. Springer.
- [26] Lehre, P. K. (2011) Negative drift in populations. In *Proceedings of Parallel Problem Solving from Nature (PPSN 2010)*, Vol. 6238 of Lecture Notes in Computer Science, pp. 244–253. Springer.

- [27] Lehre, P. K. (2012) Drift analysis (tutorial). In *Companion to GECCO 2012*, pp. 1239–1258. ACM Press.
- [28] Lehre, P. K. and Witt, C. (2014) Concentrated hitting times of randomized search heuristics with variable drift. In *Proceedings of the 25th International Symposium on Algorithms and Computation (ISAAC 2014)*, Vol. 8889 of Lecture Notes in Computer Science, pp. 686–697. Springer.
- [29] Lehre, P. K. and Witt, C. (2018) General drift analysis with tail bounds. Preprint of this paper, including supplementary material. [arXiv:1211.7184](https://arxiv.org/abs/1211.7184)
- [30] Lengler, J. (2020) Drift analysis. In *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization* (B. Doerr and F. Neumann, eds), pp. 89–131. Springer.
- [31] Lengler, J. and Steger, A. (2018) Drift analysis and evolutionary algorithms revisited. *Combin. Probab. Comput.* **27** 643–666.
- [32] Mitavskiy, B., Rowe, J. E. and Cannings, C. (2009) Theoretical analysis of local search strategies to optimize network communication subject to preserving the total number of links. *Int. J. Intelligent Computing Cybernetics* **2** 243–284.
- [33] Neumann, F. and Witt, C. (2010) *Bioinspired Computation in Combinatorial Optimization: Algorithms and Their Computational Complexity*, Natural Computing Series. Springer.
- [34] Oliveto, P. S. and Witt, C. (2011) Simplified drift analysis for proving lower bounds in evolutionary computation. *Algorithmica* **59** 369–386.
- [35] Oliveto, P. S. and Witt, C. (2012) Erratum: Simplified drift analysis for proving lower bounds in evolutionary computation. [arXiv:1211.7184](https://arxiv.org/abs/1211.7184)
- [36] Rowe, J. E. and Sudholt, D. (2012) The choice of the offspring population size in the $(1, \lambda)$ EA. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2012)*, pp. 1349–1356. ACM Press.
- [37] Sasak, G. H. and Hajek, B. (1988) The time complexity of maximum matching by simulated annealing. *J. Assoc. Comput. Mach.* **35** 387–403.
- [38] Sudholt, D. (2013) A new method for lower bounds on the running time of evolutionary algorithms. *IEEE Trans. Evol. Comput.* **17** 418–435.
- [39] Wegener, I. (2001) Theoretical aspects of evolutionary algorithms. In *Proceedings of the 28th International Colloquium on Automata, Languages and Programming (ICALP 2001)*, Vol. 2076 of Lecture Notes in Computer Science, pp. 64–78. Springer.
- [40] Witt, C. (2013) Tight bounds on the optimization time of a randomized search heuristic on linear functions. *Combin. Probab. Comput.* **22** 294–318. Preliminary version in *STACS 2012*.