Algorithmica (2011) 59: 369–386
DOI 10.1007/s00453-010-9387-z

Simplified Drift Analysis for Proving Lower Bounds in Evolutionary Computation

Pietro S. Oliveto · Carsten Witt

Received: 6 February 2009 / Accepted: 5 January 2010 / Published online: 14 January 2010

Abstract Drift analysis is a powerful tool used to bound the optimization time of evolutionary algorithms (EAs). Various previous works apply a drift theorem going back to Hajek in order to show exponential lower bounds on the optimization time of EAs. However, this drift theorem is tedious to read and to apply since it requires two bounds on the moment-generating (exponential) function of the drift. A recent work identifies a specialization of this drift theorem that is much easier to apply. Nevertheless, it is not as simple and not as general as possible. The present paper picks up Hajek's line of thought to prove a drift theorem that is very easy to use in evolutionary computation. Only two conditions have to be verified, one of which holds for virtually all EAs with standard mutation. The other condition is a bound on what is really relevant, the drift. Applications show how previous analyses involving the complicated theorem can be redone in a much simpler and clearer way. In some cases even improved results may be achieved. Therefore, the simplified theorem is also a didactical contribution to the runtime analysis of EAs.

Keywords Randomized search heuristics · Evolutionary algorithms · Computational complexity · Runtime analysis · Drift analysis

1 Introduction

Theoretical studies of the computational complexity of Evolutionary Algorithms (EAs) have appeared since the 1990s (see Oliveto, He and Yao [12]). Since then

P.S. Oliveto
Centre of Excellence for Research in Computational Intelligence and Applications (CERCIA), School of Computer Science, University of Birmingham, Birmingham, UK
e-mail: p.s.oliveto@cs.bham.ac.uk

C. Witt (✉)
DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark
e-mail: cfw@imm.dtu.dk

various mathematical techniques for the analysis of EAs have been constructed. An overview of many important tools can be found in Wegener [14].

Recently *drift analysis*, a technique that goes back to the 1940s (cf. the introduction in [5]), was introduced for the analysis of EAs by He and Yao [7, 8]. The authors concentrated on the obtainment of both lower and upper bounds on the expected runtime of EAs. Concerning lower bounds, Giel and Wegener [4] point out that a drift theorem on the success probability may also be obtained rather than only the expected waiting time. In this form the drift theorem has been used several times (e.g., Giel and Wegener [4] for maximum matching, Oliveto, He and Yao [11] for vertex cover, Friedrich, Oliveto, Sudholt and Witt [2] for analyzing population-based EAs with diversity mechanisms etc.) to prove exponential lower bounds on optimization times that even hold with probabilities exponentially close to 1. Although the mentioned drift theorem has turned out to be very useful, it often leads to tedious and complicated calculations. This seems to be the price to pay for the sake of keeping the drift theorem as general as possible. However, by considering the characteristics of the stochastic processes defined by EAs, it is possible to derive conditions which are more restrictive but considerably easier to verify. In fact, with similar motivations, Happ, Johannsen, Klein and Neumann [6] have recently introduced a simplified drift theorem called “Global Gambler’s Ruin”.

In this paper we present a further simplification of the drift theorem which is particularly suited for the analysis of EAs. Our proof resembles the argumentation used by Hajek to verify the conditions of his complicated but general theorem. It seems that, to a certain extent, many applications of the complicated theorem rely on a historical accident. Hajek himself states simpler but more restrictive conditions which he claims to be useful in applications. We only slightly tweak these conditions to make them even easier to verify in the analysis of EAs.

The rest of the paper is structured as follows. Some background on drift analysis is given in Sect. 2. In Sect. 3 the simplified drift theorem is presented. Afterwards, we study some exemplary applications to show the strength and elegance of the new approach. Section 4 contains a warm-up example. In Sect. 5 we show that the simplified drift theorem can also be used in the setting of Happ et al. [6] and that even significantly stronger results are obtained with shorter proofs. In Sect. 6 we study the maximum matching problem as an advanced application to show that proofs are considerably simplified. We finish with some conclusions.

2 Drift Analysis

Hajek introduced drift analysis to provide a flexible technique for proving the stability of processes frequently encountered in queuing systems [5]. This technique also turned out to be useful in the analysis of the computational complexity of simulated annealing for the maximum matching problem [13]. However, only in recent years was drift analysis adapted to the study of the computational complexity of EAs [7]. Following Sasaki and Hajek’s ideas for the analysis of simulated annealing, He and Yao first modelled the process underlying an EA as a Markov chain and then gave *drift* conditions for proving upper and lower bounds on the expected runtime of the EA.

Given a Markov process $\{X_t\}_{t\geq 0}$ over a search space S and a *distance function* $g\colon S \to \mathbb{R}_0^+$ mapping each state to a non-negative real number, the *one-step drift* at time t of the Markov process is defined as

$$\Delta(t) := g(X_t) - g(X_{t+1}).$$

The drift $\Delta(t)$ represents the random decrease in distance to the optimum obtained by the algorithm in one step at time t. The idea behind drift analysis is quite simple. If the current process is at distance d from the optimum and at each step there is an improvement (i.e., a positive drift) towards the optimum of at least $\delta > 0$, then the optimal solution will be found in at most d/δ steps.

From this idea the following drift theorem for obtaining upper bounds on the runtime of EAs is derived.

Theorem 1 (Drift Theorem for Upper Bounds) *Let $\{X_t\}_{t\geq 0}$ be a Markov process over a set of states S, and $g : S \to \mathbb{R}_0^+$ a function that assigns to every state a non-negative real number. Let the time to reach the optimum be $T := \min\{t \geq 0 : g(X_t) = 0\}$. If there exists $\delta > 0$ such that at any time step $t \geq 0$ and at any state X_t with $g(X_t) > 0$ the following condition holds*:

$$E(g(X_t) - g(X_{t+1}) \mid g(X_t) > 0) \geq \delta \tag{1}$$

then

$$E(T \mid X_0, g(X_0) > 0) \leq \frac{g(X_0)}{\delta}$$

and

$$E(T) \leq \frac{E(g(X_0))}{\delta}.$$

The first statement was proved by He and Yao [7, 8]. The second statement of the theorem follows from the first one by using the law of total expectation. Let n be the length of the strings representing the candidate solutions to an optimisation problem. If the drift condition, i.e. Condition (1), can be proved for $\delta = 1/\text{poly}(n)$, then a polynomial upper bound on the expected runtime can be achieved by applying Theorem 1.

By using the same idea the other way round, conditions for proving lower bounds can be obtained. If the expected drift is negative, then the algorithm is moving in expectation away from the optimum rather than towards it. However, this would not be enough to prove an exponential lower bound on the runtime. The probability that the process may perform large jumps towards the optimum must also be low. Following Hajek's previous work [5], He and Yao derived two conditions for proving exponential lower bounds on the expected runtime of an evolutionary algorithm [7]. Later on, Giel and Wegener noticed that from He and Yao's proof, a statement on the success probability could be obtained rather than just a bound on the expected runtime [4]. The latter version of the theorem follows.

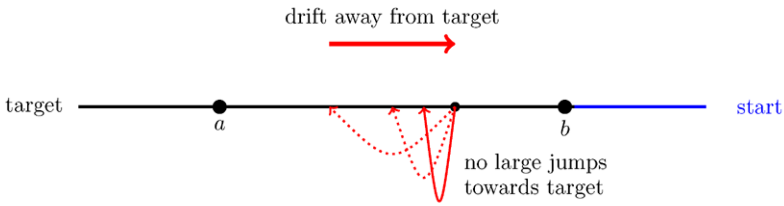

Fig. 1 Illustration of the scenario underlying the drift theorems for lower bounds

Theorem 2 (Drift Theorem for Lower Bounds [4]) *Let $\{X_t\}_{t\geq 0}$ be a Markov process over a set of states S, and $g: S \to \mathbb{R}_0^+$ a function that assigns to every state a non-negative real number. Pick two real numbers $a(\ell)$ and $b(\ell)$ which depend on a parameter $\ell \in \mathbb{R}^+$ such that $0 < a(\ell) < b(\ell)$ holds and let the random variable T denote the earliest point in time $t \geq 0$ where $g(X_t) \leq a(\ell)$ holds.*

If there are constants $\lambda > 0$ and $D \geq 1$ and a polynomial $p(\ell)$ taking only positive values, for which the following four conditions hold

1. $\Pr[g(X_0) \geq b(\ell)] = 1$
2. $b(\ell) - a(\ell) = \Omega(\ell)$
3. $\forall t \geq 0 : E[e^{-\lambda(g(X_{t+1})-g(X_t))} \mid X_t, a(\ell) < g(X_t) < b(\ell)] \leq 1 - 1/p(\ell)$
4. $\forall t \geq 0 : E[e^{-\lambda(g(X_{t+1})-b(\ell))} \mid X_t, b(\ell) \leq g(X_t)] \leq D$,

then for all time bounds $B \geq 0$, the following upper bound on probability holds for random variable T

$$\text{Prob}(T \leq B) \leq e^{\lambda(a(\ell)-b(\ell))} \cdot B \cdot D \cdot p(\ell).$$

Here $a(\ell)$ and $b(\ell)$ identify an interval where the drift $g(X_t) - g(X_{t+1})$ typically is negative, i.e., $g(X_{t+1}) > g(X_t)$. Condition 1 means that the starting point X_0 must be on the other side of the interval $[a(\ell), b(\ell)]$ compared to the target point, while Condition 2 states that the length of the interval should be at least of order ℓ. Then, Condition 3 guarantees that when the process is inside the interval it drifts away from the target, and Condition 4 assures that there are very low chances that the algorithm performs long jumps to the other side of the interval. See Fig. 1 for an illustration. If the four conditions hold then an upper bound on the success probability in B steps is obtained by applying Theorem 2. If the interval $[a(\ell), b(\ell)]$ depends on the problem size n and B is exponential in n, then the probability that the runtime is not exponential can be proved to be overwhelmingly small.

This version of the theorem was first used by Giel and Wegener to prove exponential runtime for the (1 + 1)-EA on maximum matching. Since then it has been used several times for the obtainment of exponential lower bounds (e.g., [2, 11]). Although the theorem is very powerful, the main drawback in its application is that an exponential term, namely the moment-generating function of the one-step drift $g(X_t) - g(X_{t+1})$, has to be bounded to prove whether the conditions are satisfied. This often leads to very tedious calculations. In the next section the conditions will be considerably simplified.

3 The Simplified Drift Theorem

As seen in the previous section, with the aim of proving exponential lower bounds on first hitting times, usually four conditions to be fulfilled are listed. Interestingly, in essence, there is only a single inequality, namely the bound on the moment-generating function of the one-step drift, to be checked for the final statement of the theorem to hold. By carefully looking at the original presentation by Hajek [5], it follows that the remaining conditions can be either rephrased or removed. In particular, there is no need for the following values $\lambda(\ell)$ and $D(\ell)$ to be constant or $p(\ell)$ to be polynomial. We only have to make sure that $D(\ell)$ is defined since otherwise the theorem becomes meaningless. Hajek's theorem reads in its most general form for Markov processes as follows.

Theorem 3 (Hajek [5]) *Let $X_0, X_1, X_2, \dots$ be the random variables describing a Markov process over a state space S and $g\colon S \to \mathbb{R}_0^+$ a function mapping each state to a non-negative real number. Pick two real numbers $a(\ell)$ and $b(\ell)$ depending on a parameter ℓ such that $0 \le a(\ell) < b(\ell)$ holds. Let $T(\ell)$ be the random variable denoting the earliest point in time $t \ge 0$ such that $g(X_t) \le a(\ell)$ holds. If there are $\lambda(\ell) > 0$ and $p(\ell) > 0$ such that the condition*

$$E\left(e^{-\lambda(\ell)\cdot(g(X_{t+1})-g(X_t))} \mid a(\ell) < g(X_t) < b(\ell)\right) \le 1 - \frac{1}{p(\ell)} \quad \text{for all } t \ge 0 \qquad (*)$$

holds then for all time bounds $L(\ell) \ge 0$

$$\mathrm{Prob}\left(T(\ell) \le L(\ell) \mid g(X_0) \ge b(\ell)\right) \le e^{-\lambda(\ell)\cdot(b(\ell)-a(\ell))} \cdot L(\ell) \cdot D(\ell) \cdot p(\ell),$$

where $D(\ell) = \max\{1, E(e^{-\lambda(\ell)\cdot(g(X_{t+1})-b(\ell))} \mid g(X_t) \ge b(\ell))\}$.

In the typical applications of Theorem 3 cited above, the main drift Condition (∗) is proved with $p(\ell)$ being a polynomial. Having accomplished this, it often easily follows that $D(\ell)$ does not grow with ℓ. The values $a(\ell)$ and $b(\ell)$ are frequently chosen linear in the dimension of the search space n such that $b(\ell) - a(\ell) = \Omega(n)$ and $\ell = \Omega(n)$ while $\lambda(\ell)$ is chosen constant. Consequently, choosing $L(\ell) = 2^{cn}$, where c is a sufficiently small constant, the final statement of the theorem boils down to $\mathrm{Prob}(T(\ell) \le 2^{cn}) \le 2^{-\Omega(n)}$. This is as desired: even given exponential time, the probability of finding the optimum (i.e., $g(X_t) \le a$) is exponentially small w.r.t. the problem dimensionality.

Happ et al. [6] present a simplified version of the drift theorem called "Global Gambler's Ruin" with conditions that are much easier to check. The main simplification introduced to prove Condition (∗) of the original theorem is as follows: assuming $S = \mathbb{N}_0$ and $g = \mathrm{id}$, they demand the existence of a constant $\delta > 1$ such that, given $X_t = i$, the condition $\mathrm{Prob}(X_{t+1} = i + j) \ge \delta^j\, \mathrm{Prob}(X_{t+1} = i - j)$ holds for all $j \ge 1$. Intuitively, this means that for every step length j, there is a bias (drift) towards increasing the state by j compared to decreasing it by j; moreover, this bias increases exponentially w.r.t. j. In an application to an EA with fitness-proportional selection, it turns out that the new condition is relatively easy to verify. The drawback is that $a(\ell)$ and $b(\ell)$ have to be chosen carefully to establish the exponential bias δ^j

for all j. Moreover, the new theorem by Happ et al. [6] contains an additional condition on—in essence—the moment-generating function $E(\delta^{-(X_{t+1}-X_t)} \mid X_t \geq b(\ell))$ in order to bound the value $D(\ell)$ of the original theorem. Despite being relatively easy to verify, both conditions seem stronger than needed for our purpose.

Our main contribution is another simplification of the drift theorem, which is particularly suited for the stochastic processes described by evolutionary algorithms and even easier to apply than the version by Happ et al. [6]. With the aim of proving that the process does not pass the interval $[a, b]$ within the state space $\mathbb{N}_0$ in exponential time if started above state b, we intuitively need the following two conditions:

- Assuming to be in the interval at time t, there must be a drift, an expected displacement, towards increasing the state, more precisely, there must be some constant $\varepsilon > 0$ such that $\sum_{j\in\mathbb{Z}} j \cdot \text{Prob}(X_{t+1} = i + j \mid X_t = i) \geq \varepsilon$ for all i in the interval. There seems to be no need for the drift to be bounded in the same manner for every j or even to increase with j.
- Drift alone is not enough. Considering exponentially long phases, the probability must be exponentially small to leave the interval towards the optimum using large jumps. The random step length towards the optimum has to exhibit an exponential decay. This follows from $\text{Prob}(X_{t+1} = i - j \mid X_t = i) \leq r/(1+\delta)^j$ for two constants $\delta, r > 0$ and all $i > a$, i.e., within and outside the interval. We will see that this form of the second condition always holds for standard bit flip mutations.

Besides, we will need a technical condition regarding the absolute convergence of the power series appearing in the following proof. Since we usually consider finite search spaces, we restrict the state space of the Markov process to a finite set and obtain such convergence for free. Weaker conditions could be proven if applications in infinite search spaces are desired.

Finally, we introduce a simplification which is simply a matter of notation. Instead of considering arbitrary state spaces S and mapping them to non-negative reals via the distance function $g\colon S \to \mathbb{R}_0^+$, we identify the random variables X_t, $t \geq 0$, behind the Markov process with the random outcomes of the distance functions themselves. In order to capture a broad class of distance functions, we allow any finite subset of the non-negative reals rather than only integral values. Now we are ready to state and prove our simplified drift theorem. Note that unlike the perspectives in Theorems 1 and 2, we are interested in bounding $X_{t+1} - X_t$, the random *increase* in distance; hence positive values drive us away from the target.

Theorem 4 (Simplified Drift Theorem) *Let X_t, $t \geq 0$, be the random variables describing a Markov process over a finite state space $S \subseteq \mathbb{R}_0^+$ and denote $\Delta_t(i) := (X_{t+1} - X_t \mid X_t = i)$ for $i \in S$ and $t \geq 0$. Suppose there exist an interval $[a, b]$ in the state space, two constants $\delta, \varepsilon > 0$ and, possibly depending on $\ell := b - a$, a function $r(\ell)$ satisfying $1 \leq r(\ell) = o(\ell/\log(\ell))$ such that for all $t \geq 0$ the following two conditions hold:*

1. $E(\Delta_t(i)) \geq \varepsilon$ *for* $a < i < b$,
2. $\text{Prob}(\Delta_t(i) \leq -j) \leq \frac{r(\ell)}{(1+\delta)^j}$ *for* $i > a$ *and* $j \in \mathbb{N}_0$.

Then there is a constant $c^* > 0$ such that for $T^* := \min\{t \geq 0\colon X_t \leq a \mid X_0 \geq b\}$ it holds $\text{Prob}(T^* \leq 2^{c^*\ell/r(\ell)}) = 2^{-\Omega(\ell/r(\ell))}$.

In the conference version of this paper [10], $r(\ell)$ was only allowed to be a constant, i.e., $r(\ell) = O(1)$. In this case, the final statement of the theorem is even simpler as we get

$$\text{Prob}(T^* \le 2^{c^*\ell}) = 2^{-\Omega(\ell)}.$$

In fact, the new and slightly more general version of the simplified drift theorem as presented here has been recently found useful for the analysis of a population based evolutionary algorithm using fitness-proportional selection [9].

Proof We will apply Theorem 3 for suitable choices of its variables, some of which might depend on the parameter $\ell = b - a$ denoting the length of the interval $[a, b]$. Moreover, we set $g :=$ id. The following argumentation is also inspired by Hajek's work [5].

Fix $t \ge 0$ and some i such that $a < i < b$. Let $Z := \{s - i \mid s \in S\}$ be the finite set containing all possible changes of the state number (in particular negative and fractional values are possible) and denote $p_j := \text{Prob}(\Delta_t(i) = j)$ for $j \in Z$. To prove Condition (∗), it is sufficient to identify values $\lambda := \lambda(\ell) > 0$ and $p(\ell) > 0$ such that

$$S(\lambda) := \sum_{j \in Z} e^{-\lambda j} p_j \le 1 - \frac{1}{p(\ell)}.$$

Using the series expansion for $e^{\lambda j} = \sum_{k=0}^{\infty} (\lambda j)^k / k!$, we have

$$S(\lambda) = \sum_{j \in Z} \sum_{k=0}^{\infty} \frac{(-\lambda j)^k}{k!} p_j = 1 + \sum_{j \in Z} (-\lambda j) \cdot p_j + \sum_{k=2}^{\infty} \sum_{j \in Z} \frac{(-\lambda j)^k}{k!} p_j,$$

where all series converge absolutely for any $\lambda > 0$ since we are dealing with a finite state space and thus a finite Z; however, their limits might depend on the largest state number. The first sum in the last expression equals $-\lambda E(\Delta_t(i))$, and the double sum is non-negative since $e^x - (1 + x) \ge 0$ for every (also negative) $x \in \mathbb{R}$. Using this together with the first condition of the theorem, i.e., the bound on the drift, we obtain for all $\gamma \ge \lambda$

$$S(\lambda) \le 1 - \lambda E(\Delta_t(i)) + \frac{\lambda^2}{\gamma^2} \sum_{k=2}^{\infty} \sum_{j \in Z} \frac{(-\gamma j)^k}{k!} p_j \le 1 - \lambda\varepsilon + \lambda^2 \cdot \underbrace{\frac{\sum_{j \in Z} e^{-\gamma j} p_j}{\gamma^2}}_{=:C(\gamma)}.$$

Given any $\gamma > 0$, choosing $\lambda := \min\{\gamma, \varepsilon/(2C(\gamma))\}$ results in

$$S(\lambda) \le 1 - \lambda\varepsilon + \lambda \cdot \frac{\varepsilon}{2C(\gamma)} \cdot C(\gamma) = 1 - \frac{\lambda\varepsilon}{2} = 1 - \frac{1}{p(\ell)}$$

with $p(\ell) := 2/(\lambda\varepsilon) = \Theta(1/\lambda)$ since only λ but not ε is allowed to depend on ℓ. Choosing $\gamma := \ln(1+\delta/2)$, which does not depend on ℓ since δ is a constant, yields

$$C'(\gamma) := \sum_{j\in Z} e^{-\gamma j} p_j = \sum_{\substack{j\in Z:\\ j\geq 0}} (1+\delta/2)^{-j}\cdot p_j + \sum_{\substack{j\in Z:\\ j<0}} (1+\delta/2)^{-j}\cdot p_j$$

$$\leq \sum_{\substack{j\in Z\\ j\geq 0}} p_j + \sum_{\substack{j\in Z:\\ j<0}} (1+\delta/2)^{-j}\cdot p_j \leq 1 + \sum_{j\in\mathbb{N}_0} (1+\delta/2)^{j+1}\cdot \mathrm{Prob}(\Delta_t(i)\leq -j),$$

where the last inequality follows by means of $\sum_{j\in Z} p_j = 1$ and $\sum_{j'\in Z\cap[-j-1,-j]} p_{j'} \leq \mathrm{Prob}(\Delta_t(i)\leq -j)$ for any integral j as well as $(1+\delta/2)^{j'} \leq (1+\delta/2)^{j+1}$ for $j'\leq j+1$. Now, exploiting the second condition, we get

$$C'(\gamma) \leq 1 + r(\ell)\cdot \sum_{j\in\mathbb{N}_0} \frac{(1+\delta/2)^{j+1}}{(1+\delta)^j} = 1 + r(\ell)\cdot\left(1+\frac{\delta}{2}\right)\cdot \sum_{j\in\mathbb{N}_0}\left(1-\frac{\delta/2}{1+\delta}\right)^j$$

$$= 1 + r(\ell)\cdot\left(1+\frac{\delta}{2}\right)\cdot\left(2+\frac{2}{\delta}\right) \leq r(\ell)\cdot(4+\delta+2/\delta),$$

where $r(\ell)\geq 1$ is used. Hence $C(\gamma)\leq r(\ell)(4+\delta+2/\delta)/\ln^2(1+\delta/2)$, which means $C(\gamma) = O(r(\ell))$ since δ is a constant. By our choice of λ, we have $\lambda \geq \varepsilon/(2C(\gamma)) = \Omega(1/r(\ell))$ since also ε is a constant. Since $p(\lambda) = \Theta(1/\lambda)$, we know $p(\ell) = O(r(\ell))$, too. Condition ($*$) of Theorem 3 has been established along with these bounds on $p(\ell)$ and $\lambda = \lambda(\ell)$.

To bound the probability of a success within $L(\ell)$ steps, we still need a bound on $D(\ell) = \max\{1, E(e^{-\lambda(X_{t+1}-b)} \mid X_t \geq b)\}$. If 1 maximizes the expression then $D(\ell) \leq r(\ell)$ follows. Otherwise,

$$D(\ell) = E(e^{-\lambda(X_{t+1}-b)} \mid X_t\geq b) \leq E(e^{-\lambda(X_{t+1}-X_t)} \mid X_t \geq b)$$

$$= \sum_{j\in Z} e^{-\lambda j}\cdot \mathrm{Prob}(\Delta_t(i) = j \mid X_t\geq b) \leq 1$$

$$+ \sum_{\substack{j\in Z\\ j<0}} e^{-\lambda j}\cdot \mathrm{Prob}(\Delta_t(i) = j \mid X_t \geq b)$$

$$\leq 1 + \sum_{\substack{j\in Z\\ j<0}} e^{-\gamma j}\cdot \mathrm{Prob}(\Delta_t(i) = j \mid X_t\geq b)$$

for $i\geq b$, where the first inequality follows from $X_t\geq b$, the second one since we are dealing with a probability distribution, and the third one since $\gamma\geq\lambda$. The last term can be bounded as in the above calculation leading to $C'(\gamma) = O(r(\ell))$ since that estimation holds for arbitrary $i\geq a$ due to the second condition. Hence, in any case also $D(\ell) = O(r(\ell))$. Altogether, we have

$$e^{-\lambda(\ell)\cdot\ell} \cdot D(\ell) \cdot p(\ell) = e^{-\Omega(1/r(\ell))\cdot\ell} \cdot O((r(\ell))^2)$$
$$= e^{-\Omega(\ell/r(\ell))+O(\log(r(\ell)))} = 2^{-\Omega(\ell/r(\ell))},$$

where the last simplification follows since $r(\ell) = o(\ell/\log(\ell))$ by prerequisite. Choosing $L(\ell) = 2^{c^*\ell/r(\ell)}$ for some sufficiently small constant $c^* > 0$, Theorem 3 yields

$$\text{Prob}(T(\ell) \le L(\ell)) \le L(\ell) \cdot 2^{-\Omega(\ell/r(\ell))} = 2^{-\Omega(\ell/r(\ell))},$$

which proves the theorem. □

We repeat that $r(\ell)$, which is only relevant for the second condition of the theorem, can be bounded from above by a constant in most applications to the well-known $(1+1)$-EA. In the context of population-based EAs, a union bound over the number of individuals can nevertheless entail values for $r(\ell)$ which grow slowly with the population size. This was the case in the analysis of a population based evolutionary algorithm using fitness-proportional selection [9].

In the latter paper, there was also need for a fractional state space $S \subseteq [0, N]$. If the state space is integral, i.e., $S = \{0, 1, \ldots, N\}$, then the first condition of the theorem follows if

$$\text{Prob}(\Delta_t(i) = j) \ge (1+\varepsilon) \cdot \text{Prob}(\Delta_t(i) = -j) \quad \text{for all } j \ge 1$$

and additionally $\text{Prob}(\Delta_t(i) = 0) \le 1 - \varepsilon$. Hence, we can obtain a corollary from our Simplified Drift Theorem that reminds of the "Global Gambler's Ruin" by Happ et al. [6] by using a bias towards increasing states separately for every step size j. However, in contrast to the latter theorem, we do not demand the bias to grow with j since we use a factor $(1+\varepsilon)$ for every j instead of $(1+\varepsilon)^j$ as in [6].

Our Simplified Drift Theorem can therefore easily be applied to Randomized Local Search (RLS) on the search space $\{0, 1\}^n$, which flips only one bit per iteration. Then Condition 2 is trivially fulfilled for $r(\ell) = O(1)$ and the theorem breaks down to the classical *Gambler's Ruin* Theorem where only steps of size 1 are allowed (see also [6]). However, the generalized drift technique was previously used to obtain lower bounds on the first hitting time of the $(1+1)$-EA, which can flip several bits in a step. Then the original Gambler's Ruin Theorem does not apply. For maximization problems, the $(1+1)$-EA is defined as follows.

(1 + 1)-EA
- Choose uniformly at random an initial bit string $x \in \{0, 1\}^n$;
- Repeat the following steps until a termination criterion is satisfied:

1. Create x' by flipping each bit in x with probability $p := 1/n$;
2. Replace x with x' if $f(x') \ge f(x)$.

In the rest of the paper we will show that proofs regarding lower bounds on the runtime of the $(1+1)$-EA that hold with overwhelming probability $1 - 2^{-\Omega(b-a)}$ are really easy to obtain by using the proposed drift theorem. Our proofs are universal enough to apply, after some tiny changes, also for RLS.

4 An Application for the (1 + 1)-EA

In this section we present a first application of Theorem 4 as a warm-up example. We choose the Needle-in-a-haystack function which is well known to be hard for EAs [3]. However, we also show by simple arguments that the (1 + 1)-EA is even at distance almost $n/2$ from its optimum for an exponential number of steps. The whole search space consists of a plateau except for one point representing the global optimum. W.l.o.g. we choose the optimum to be the point represented by the bit string of all ones. The function is the following:

$$\textsc{Needle}_n(x) = \begin{cases} 1 & \text{if } x = 1^n, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 5 *Let $\eta > 0$ be constant. Then there is a constant $c > 0$ such that with probability $1 - 2^{-\Omega(n)}$ the (1 + 1)-EA on $\textsc{Needle}_n$ creates only search points with at most $n/2 + \eta n$ ones in 2^{cn} steps.*

Proof Let X_t denote the number of zeroes in the bit string at time step t. We set $a := n/2 - 2\gamma n$ and $b := n/2 - \gamma n$, where $\gamma := \eta/2$. Such a value for b is suitable because by Chernoff bounds the probability that the initial bit string has less than $n/2 - \gamma n$ zeroes is $2^{-\Omega(n)}$. Now we use the proposed Simplified Drift Theorem for the rest of the proof. It therefore remains to check that the two conditions of Theorem 4 hold.

Given a string in state $i < n/2 - \gamma n$, i.e., with i zeroes, let $\Delta(i)$ denote the random increase of the number of zeroes. Condition 1 holds if $E(\Delta(i)) \geq \varepsilon$ for some constant $\varepsilon > 0$. Since the (1 + 1)-EA flips 0-bits and 1-bits independently, an expected number of i/n 0-bits and $(n-i)/n$ 1-bits is flipped. Hence,

$$E(\Delta(i)) = \frac{n-i}{n} - \frac{i}{n} = \frac{n-2i}{n} \geq 2\gamma.$$

So we can choose $\varepsilon = 2\gamma$ for Condition 1 to hold.

Condition 2 is: $\text{Prob}(\Delta(i) \leq -j) \leq r(b-a)/(1+\delta)^j$ for $j \in \mathbb{N}_0$. In order to reach state $i - j$ or less from state i, at least j bits have to flip. Hence

$$\text{Prob}(\Delta(i) \leq -j) \leq \binom{n}{j}\left(\frac{1}{n}\right)^j \leq \frac{1}{j!} \leq 2 \cdot \left(\frac{1}{2}\right)^j,$$

which proves the condition for $\delta = 1$ and $r(b-a) = 2$ even independently of i and of selection. So from Theorem 4 it follows for a constant $c^* > 0$ that the global optimum is found in $2^{c^*(b-a)} = 2^{cn}$ steps, where $c := c^*(b-a)/n > 0$ is a different constant, with probability at most $2^{-\Omega(b-a)} = 2^{-\Omega(n)}$. □

Previous analyses of the (1 + 1)-EA on $\textsc{Needle}_n$ by direct approaches [3] or arguments from black-box complexity [1] did not show that the optimization process is at distance almost $n/2$ for an exponential number of steps.

5 An Application for the (1 + 1)-EA with Fitness-proportional Selection

Recently Happ et al. [6] have presented a simplified drift theorem called *Global Gambler's Ruin*. They introduced the new theorem to prove that the (1 + 1)-EA using fitness-proportional selection requires exponential runtime for optimizing OneMax and linear functions in general. The algorithm considered by theses authors works as follows:

(1 + 1)-EA with Fitness-proportional Selection (**(1 + 1)-EA$_{\text{prop}}$**)
- Choose uniformly at random an initial bit string $x \in \{0,1\}^n$;
- Repeat the following steps until a termination criterion is satisfied:

1. Create x' by flipping each bit in x with probability $p := 1/n$;
2. Replace x with x' with probability $f(x')/(f(x') + f(x))$.

A pseudo-boolean function $f\colon \{0,1\}^n \to \mathbb{R}$ is called *linear* if it can be written as $f(x_1,\dots,x_n) = w_0 + w_1x_1 + \dots + w_nx_n$ with coefficients $w_i \geq 0$, $0 \leq i \leq n$. In the special case $w_1 = \dots = w_n = 1$ and $w_0 = 0$ we obtain the OneMax function counting the number of ones of the bit string. Concerning linear functions, Happ et al. [6] prove that with overwhelming probability only search points with at most $0.97n$ ones are created by the (1 + 1)-EA$_{\text{prop}}$ after an exponential number of steps. We show that Theorem 4 can be used for this purpose and that it can lead to significantly stronger results. We remark that the following proof also holds for fitness-proportional RLS, where the stronger statement is already known [6].

Theorem 6 *Let $0 < \eta \leq 1/4$ and η be constant. Then there is a constant $c > 0$ such that with probability $1 - 2^{-\Omega(n)}$ the (1 + 1)-EA$_{\text{prop}}$ for linear functions only creates search points with at most $2n/3 + \eta n$ ones in 2^{cn} steps.*

In the following lemmas we will bound the drift before and after selection separately. Afterwards, we will add the results of the lemmas together and prove Theorem 6 by applying our Simplified Drift Theorem.

For this purpose, we set $a := n/3 - 2\gamma n$ and $b := n/3 - \gamma n$, where $\gamma := \eta/2 \leq 1/8$. Given a current number of $a < i < b$ zeroes, let $\Delta(i)$ and $\Delta^{\text{sel}}(i)$ denote the random change in this number before and after the application of the fitness-proportional selection operator, respectively. Furthermore, let $\Delta^+(i) := \Delta(i) \cdot \mathbb{1}\{\Delta(i) > 0\}$ and $\Delta^-(i) := -\Delta(i) \cdot \mathbb{1}\{\Delta(i) < 0\}$ be the positive and the negative contributions to the drift such that $\Delta(i) = \Delta^+(i) - \Delta^-(i)$. The notion $\mathbb{1}\{E\}$ for an event E represents the indicator random variable of the event.

We first note that both the drift before and after selection are mostly determined by "small" steps of size at most $r = \gamma n/4$. To this end, we prove in the following lemma that considering only such steps introduces an exponentially small absolute error. Here the abbreviation $\mathbb{1}_r := \mathbb{1}\{|\Delta(i)| \leq r\}$ is useful.

Lemma 1 *Let $r := \gamma n/4$. Then $E(\Delta(i) \cdot \mathbb{1}_r) \geq E(\Delta(i)) - 2^{-\Omega(n)}$ and also $E(\Delta^{sel}(i) \cdot \mathbb{1}_r) \geq E(\Delta^{sel}(i)) - 2^{-\Omega(n)}$.*

Proof Flipping at least r bits in a step has a probability of at most

$$\binom{n}{r}\left(\frac{1}{n}\right)^r \le \frac{1}{r!},$$

and, altogether, at most n bits can flip. Using Stirling's formula and that γ is a positive constant, we obtain

$$E(\Delta(i)\cdot \mathbb{1}_r) \ge E(\Delta(i)) - \frac{1}{r!}\cdot n = E(\Delta(i)) - \frac{1}{(\gamma n/4)!}\cdot n = E(\Delta(i)) - 2^{-\Omega(n)}$$

and accordingly

$$E(\Delta^{\text{sel}}(i)\cdot \mathbb{1}_r) \ge E(\Delta^{\text{sel}}(i)) - 2^{-\Omega(n)}.$$ □

Now, in Lemma 2, the expected negative drift before the application of the selection operator will be considered, and the ratio of positive and negative drift will be bounded from below. Then, in Lemma 3, the effects of the fitness-proportional selection will be considered.

Lemma 2 $E(\Delta^-(i)) \ge 1/36$. *Moreover,* $E(\Delta^+(i)) \ge (2+6\gamma)\cdot E(\Delta^-(i))$.

Proof Using the arguments from the proof of Theorem 5, we start with

$$E(\Delta(i)) = \frac{n-2i}{n} \ge \frac{n-2(n/3-\gamma n)}{n} = \frac{n/3-2\gamma n}{n} = \frac{1}{3}+2\gamma$$

as a bound on the drift before selection.

We proceed by bounding $E(\Delta^-(i))$ from below and (as needed later) also from above. By considering only the expected number of flipping bits among the $i \le b$ 0-bits, we get

$$E(\Delta^-(i)) \le \frac{b}{n} = \frac{n/3-\gamma n}{n} = \frac{1}{3}-\gamma.$$

On the other hand, we get a lower bound by considering the $i \ge a$ different mutations flipping only one 0-bit and no other bit. Each of these mutations has probability $(1/n)\cdot(1-1/n)^{n-1}$. Hence,

$$E(\Delta^-(i)) \ge \frac{a}{n}\cdot\left(1-\frac{1}{n}\right)^{n-1} = \left(\frac{1}{3}-2\gamma\right)\left(1-\frac{1}{n}\right)^{n-1} \ge \frac{1/3-2\gamma}{e} \ge \frac{1}{36},$$

where the last inequality follows from $\gamma \le 1/8$. This proves the first statement of the lemma.

To prove the second statement, we use the bounds $E(\Delta^-(i)) \le 1/3$ and $E(\Delta(i)) \ge 1/3+2\gamma$ derived above and the linearity of expectation. Hence,

$$\frac{E(\Delta^+(i))}{E(\Delta^-(i))} = \frac{E(\Delta(i)+\Delta^-(i))}{E(\Delta^-(i))} = \frac{E(\Delta(i))}{E(\Delta^-(i))} + \frac{E(\Delta^-(i))}{E(\Delta^-(i))} \ge 2+6\gamma,$$

or, equivalently, $E(\Delta^+(i)) \ge (2+6\gamma)\cdot E(\Delta^-(i))$. □

Lemma 3 *Let x' be the random offspring produced by mutation of an arbitrary search point x. The selection probability $s(x') := f(x')/(f(x) + f(x'))$ for x' is lower bounded by* $1/2 - \gamma/4 - 2^{-\Omega(n)}$ *and upper bounded by* 1.

Proof The upper bound 1 is trivial. For the lower bounds, we pessimistically assume all zeroes of the current search point x to have coefficients 0. This implies $f(x') \le f(x)$ for all offspring x' of x. Hence, the selection probability for arbitrary but fixed x' is at least

$$s(x') := \frac{f(x')}{f(x') + f(x)} \ge \frac{f(x')}{2f(x)},$$

which is linear in $f(x')$. We now turn to considering x' as a random object created by the probabilistic mutation operator. The aim is to relate the expectation of $f(x')$ with the value $f(x)$. Again we consider at most $r = \gamma n/4$ flipping bits, which, as shown in Lemma 1, introduces an error of at most $2^{-\Omega(n)}$. Let us pessimistically assume exactly r flipping bits and denote by F the corresponding event. Moreover, all flipping bits are imagined as 1-bits. If a random subset of r out of $n-i$ bits flips, each bit flips (not independently) with probability $r/(n-i)$. Let us pessimistically assume $w_0 = 0$ for the constant coefficient of f. Then each coefficient with contribution to $f(x)$ is derated by an expected factor of $r/(n-i)$. Using the linearity of expectation and of the fitness function f, the expected offspring value $e_F(x') := E(f(x') \mid F)$ under assumption F is therefore at least $f(x)(1 - \frac{r}{n-i})$.

Our aim is to obtain a lower bound on the selection probability $s(x' \mid F)$ for a random offspring x' assuming the event F. To this end, we apply the law of total probability and decompose the probability space according to the events $\{x' = y\}$, $y \in \{0,1\}^n$. Hence,

$$s(x' \mid F) = \sum_{y \in \{0,1\}^n} s(x' \mid x' = y) \cdot \text{Prob}(x' = y \mid F).$$

We already know the lower bound $s(x' \mid x' = y) \ge f(x' \mid x' = y)/(2f(x))$. Thereby,

$$s(x' \mid F) = \sum_{y \in \{0,1\}^n} \frac{f(x' \mid x' = y)}{2f(x)} \cdot \text{Prob}(x' = y \mid F) = \frac{e_F(x')}{2f(x)},$$

and we are ready to apply the lower bound on the expectation $e_F(x')$ developed above. Thus,

$$s(x' \mid F) \ge \frac{e_F(x')}{2f(x)} \ge \frac{1}{2}\left(1 - \frac{r}{n-i}\right) = \frac{1}{2} - \frac{r}{2(n-i)} = \frac{1}{2} - \frac{\gamma n/4}{2(n-i)} \ge \frac{1}{2} - \frac{\gamma}{4},$$

where the last inequality follows from $n - i \ge n/2$. Taking into consideration the error brought in by assuming F, we have $s(x') = s(x' \mid F) - 2^{-\Omega(n)} = 1/2 - \gamma/4 - 2^{-\Omega(n)}$. □

Now we are ready to apply the Simplified Drift Theorem to bound the total drift.

Proof of Theorem 6 First we prove Condition 1 of the Simplified Drift Theorem. Applying Lemma 3 yields

$$E(\Delta^{\text{sel}}(i)) \geq \left(\frac{1}{2} - \frac{\gamma}{4}\right) \cdot E(\Delta^+(i) \cdot \mathbb{1}_r) - E(\Delta^-(i) \cdot \mathbb{1}_r) - 2^{-\Omega(n)}$$

using the upper and lower bounds on the selection probability and taking into consideration the error brought in by flipping at most r bits. We proceed by filling in the lower bound $E(\Delta^-(i)) \geq 1/36$ as well as the inequality $E(\Delta^+(i)) \geq (2+6\gamma)E(\Delta^-(i))$ obtained in Lemma 2 together with our prerequisite $\gamma \leq 1/8$. Hence,

$$\begin{aligned} E(\Delta^{\text{sel}}(i)) &\geq \left(\frac{1}{2} - \frac{\gamma}{4}\right) \cdot (2+6\gamma) \cdot E(\Delta^-(i) \cdot \mathbb{1}_r) - E(\Delta^-(i) \cdot \mathbb{1}_r) - 2^{-\Omega(n)} \\ &\geq \left(\frac{5\gamma}{2} - \frac{3\gamma^2}{2}\right) \cdot E(\Delta^-(i) \cdot \mathbb{1}_r) - 2^{-\Omega(n)} \geq \frac{37\gamma}{16} \cdot \frac{1}{36} - 2^{-\Omega(n)} \geq \frac{\gamma}{16} \end{aligned}$$

for n large enough. This bounds the drift for general linear functions by a constant as desired for the satisfaction of Condition 1 of Theorem 4. The proof of Condition 2 carries over from the proof of Theorem 5. □

For the linear function OneMax, stronger results can be obtained as stated in the following theorem.

Theorem 7 *Let $0 < \eta \leq 1/4$ and η be constant. Then there is a constant $c > 0$ such that with probability $1 - 2^{-\Omega(n)}$ the $(1+1)$-EA$_{\text{prop}}$ for* OneMax *only creates search points with at most $n/2 + \eta n$ ones in 2^{cn} steps.*

Proof We follow the proof idea of Theorem 6. With OneMax, the situation is simpler. Since f equals the number of ones, we can bound from above the probability of accepting a string x' with up to r more ones than x by

$$s(x' \mid F) \leq \frac{f(x' \mid F)}{2f(x)} \leq \frac{f(x) + r}{2f(x)} \leq \frac{f(x)(1+\gamma/2)}{2f(x)} = \frac{1}{2} + \frac{\gamma}{4}$$

for all search points x such that $f(x) \geq n/2$, i.e., search points with at most $n/2$ zeroes.

Setting $a := n/2 - 2\gamma n$ and $b := n/2 - \gamma n$, similar calculations as in Lemma 2 of Theorem 6 yield the following three bounds for the drift before selection:

$$\begin{aligned} E(\Delta(i)) &= \frac{n-2i}{n} \geq \frac{n - 2(n/2 - \gamma n)}{n} = \frac{2\gamma n}{n} = 2\gamma, \\ E(\Delta^-(i)) &\leq \frac{b}{n} = \frac{n/2 - \gamma n}{n} = \frac{1}{2} - \gamma \leq \frac{1}{2} \end{aligned}$$

and

$$\frac{E(\Delta^+(i))}{E(\Delta^-(i))} = \frac{E(\Delta(i))}{E(\Delta^-(i))} + \frac{E(\Delta^-(i))}{E(\Delta^-(i))} \geq \frac{2\gamma}{1/2} + 1 \geq 1 + 4\gamma.$$

Finally, by using the lower bounds on $s(x')$ and $E(\Delta^-(i))$ presented in the proof of Theorem 6 (recall that $\gamma \le 1/8$), we obtain

$$E(\Delta^{\text{sel}}(i)) \ge \left(\left(\frac{1}{2} - \frac{\gamma}{4}\right)(1+4\gamma) - \left(\frac{1}{2} + \frac{\gamma}{4}\right)\right) \cdot E(\Delta^-(i)\cdot \mathbb{1}_r) - 2^{-\Omega(n)}$$
$$\ge \left(\frac{3\gamma}{2} - \gamma^2\right) \cdot E(\Delta^-(i)\cdot \mathbb{1}_r) - 2^{-\Omega(n)} \ge \gamma \cdot \frac{1}{36} - 2^{-\Omega(n)} \ge \frac{\gamma}{40}$$

for n large enough in the same manner as in the proof of Theorem 6.

Again, the proof of Condition 2 of Theorem 4 carries over from the proof of Theorem 5. □

6 An Advanced Application: Maximum Matching

This section presents an application of the Simplified Drift Theorem in combinatorial optimization, more precisely for the analysis of the (1 + 1)-EA on the well-known *maximum matching* problem. Giel and Wegener [4] considered the graph depicted in Fig. 2 consisting of ℓ "columns" and h "rows". They proved that the (1 + 1)-EA has an expected optimization time which is exponential in the number of graph edges in the worst case. One of the crucial parts of their proof is represented by the following theorem.

Theorem 8 *Starting with an almost perfect matching with an augmenting path of length ℓ, the probability that the (1 + 1)-EA finds the perfect matching of the $G_{h,\ell}$ graph within $2^{c\ell}$ steps, $c > 0$ an appropriate constant, is bounded by $2^{-\Omega(\ell)}$ if $h \ge 3$.*

Proof An *almost perfect matching* is just one fitness level away from the global optimum. In order to find the maximum matching, the edges of the only augmenting path in the graph have to be either inverted or the path has to be shortened to its minimum (i.e., three adjacent edges not belonging to the matching are obtained). If the latter case happens, then the extra edge may be added by just using one bit flip. Given an almost perfect matching, the length of the augmenting path changes if at least two adjacent edges flip on either side of the augmenting path. The augmenting path may be *lengthened* or *shortened*. In the former case the process drifts away from the optimum while in the latter case it heads towards it. To apply Theorem 4, we set $a := 0$ and $b := \lceil \ell/2 \rceil - 1$. The random variable X_t, $t \ge 0$, is obtained by taking the random length of the augmenting path at time t, dividing it by 2 and rounding the result up. In this way, we obtain a process on the state space $\{0, 1, \ldots, \lceil \ell/2 \rceil\}$.

We consider a current X_t-value of i, where $i \le \lceil \ell/2 \rceil - 1$. Usually there are $2h$ edges adjacent to the augmenting path, h at each side, that flipped together with the

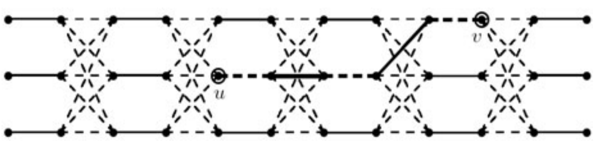

Fig. 2 The $G_{h,\ell}$ graph (in this case $h = 3$ and $\ell = 11$) with an almost perfect matching and its augmenting path between u and v

first edge belonging to the path would *lengthen* it. However, if the augmenting path starts at the beginning of the graph (or at the other end), then there are only h such edges (actually this shows that the length of the augmenting path is not enough to describe the underlying Markov process exactly, yet it gives good enough bounds). In this case, the probability of increasing the X_t-value by 1, i.e., lengthening the augmenting path of length $2i-1$ by 2, is only bounded by

$$p_1(i) \geq \frac{h}{m^2}\left(1-\frac{1}{m}\right)^{m-2},$$

where m is the number of edges of the graph. Here we use that $i \leq \lceil \ell/2 \rceil - 1$, i.e., the augmenting path can still be lengthened. On the other hand, the probability to shorten the augmenting path with a move of length 1 is bounded from above by (see [4])

$$p_{-1}(i) \leq \frac{2}{m^2}\left(1-\frac{1}{m}\right)^{m-2} + \frac{3}{m^4}.$$

Since most other mutations of the $(1+1)$-EA will be rejected in this setting due to worse fitness, we use the condition C_{rel} that a step is *relevant*, meaning it is accepted and changes the current state. The probability p_{rel} of a relevant step is bounded according to

$$\frac{1}{m^2}\left(1-\frac{1}{m}\right)^{m-2} \leq p_{\mathrm{rel}} \leq \frac{2h+2}{m^2}.$$

The lower bound holds because, unless the optimum has been found, there always are two edges that when flipped lenghten or shorten the augmenting path (i.e., the edges at the extremities of the augmenting path). The upper bound holds because there are at most $2h+2$ couples of edges that, if flipped, lenghten or shorten the augmenting path (i.e., h at each extremity lengthening the path and 1 at each extremity shortening it). The probability that more than two bits flip and the step is relevant is lower because at least one of the $2h+2$ couples considered in the bound has to be flipped anyway.

Let $R(i) = (\Delta(i) \mid C_{\mathrm{rel}})$ denote the random increase of the X_t-value in relevant steps, given a current value of i. It suffices to concentrate on the contribution of steps of length 1, i.e., we consider $R_1(i) := R(i) \cdot \mathbb{1}\{|R(i)| \leq 1\}$. We obtain

$$\begin{aligned} E(R_1(i)) &= \frac{p_1(i)}{p_{\mathrm{rel}}} - \frac{p_{-1}(i)}{p_{\mathrm{rel}}} \geq \frac{\frac{h}{m^2}\cdot(1-\frac{1}{m})^{m-2}}{\frac{2h+2}{m^2}} - \frac{\frac{2}{m^2}\cdot(1-\frac{1}{m})^{m-2} + \frac{3}{m^4}}{\frac{2h+2}{m^2}} \\ &= \frac{(h-2)(1-\frac{1}{m})^{m-2}}{2h+2} - \frac{3}{m^2(2h+2)} \geq \frac{1}{8\cdot e} - O(m^{-2}) \end{aligned}$$

since $h \geq 3$ while the unconditional decrease $\Delta^-_{>1}(i) = -\Delta(i) \cdot \mathbb{1}\{\Delta(i) < -1\}$, for negative steps of length greater than 1, in expectation is at most

$$E(\Delta^-_{>1}(i)) \leq \sum_{j=2}^{\infty} j \cdot p_{-j}(i) \leq \sum_{j=2}^{\infty} j \cdot (j+1) \cdot \frac{1}{m^{2j}} \leq \frac{6}{m^4} + \sum_{j=3}^{\infty} \frac{2m^2}{m^{2j}} = O(m^{-4})$$

because $p_{-j} \leq (j+1)/m^{2j}$ [4]. Hence, the total conditional drift is

$$E(R(i)) \geq E(R_1(i)) - \frac{E(\Delta^{-}_{>1}(i))}{p_{\text{rel}}} \geq \frac{1}{8 \cdot e} - O(m^{-2}) - O(m^{-4}) \cdot em^2$$
$$= \frac{1}{8 \cdot e} - O(m^{-2})$$

and Condition 1 is proved. Condition 2, with $\delta = 1$ and $r = 8$, follows from

$$\frac{p_{-j}}{p_{\text{rel}}} \leq \min\left\{1, \frac{j+1}{m^{2j}} \cdot em^2\right\} \leq \min\left\{1, \frac{1}{m^{2j-7}}\right\} \leq 8 \cdot \left(\frac{1}{2}\right)^j$$

for $m \geq 2$. From Theorem 4, the proof follows. □

The bounds on $p_j(i)$ by Giel and Wegener [4] do not imply $p_j(i) \geq p_{-j}(i)$ for every j, hence the theorem by Happ et al. [6] does not apply with these bounds. Without further work on the bounds for $p_j(i)$, it is crucial but also sufficient to focus on the effect of steps of length 1.

7 Conclusion

A simplified drift-analysis theorem has been introduced for proving lower bounds on the runtime of EAs that hold with high probability. The two hypotheses of the theorem are easy to check for stochastic processes such as those described by EAs. The *first condition* holds if the distance to the optimum increases in expectation by at least a constant amount. In other terms, there is a *drift* leading away from the optimum. The *second condition* describes an *exponential decay* in the probabilities of advancing towards the optimum that depends on the step size. Such a condition is trivially fulfilled for the (1 + 1)-EA with standard mutation and many other EAs with a mutation operator that exhibits enough locality. The simplified drift theorem allowed us to redo previous analyses with significantly reduced effort.

For scenarios where bounding the drift directly is more intricate a corollary of the simplified theorem might be used. It is sufficient to decompose the drift into the effects of steps of a given length and to prove a bias leading away from the optimum for every step length. In fact, also Happ et al. [6] exploited a similar idea. Our corollary, though, seems to be easier to verify since we do not require the bias to increase with the step length. Moreover, compared to the latter work, we do not require that the length of the drift interval $[a, b]$ is $\Omega(n)$. Our generalization is necessary, for example, in the study by Friedrich et al. [2] where $b - a = \sqrt[3]{n}$. To the best of our knowledge all previous applications of drift analysis to evolutionary computation can be proven in a considerably simpler shape with the proposed simplified drift theorem. As a result, not only is Theorem 4 considered as an important didactical contribution to the runtime analysis of EAs, but we also believe it will turn out to be useful in future work.

Acknowledgements Pietro Oliveto was supported by an EPSRC grant (EP/C520696/1). Carsten Witt was supported by the Deutsche Forschungsgemeinschaft (DFG) as a part of the Collaborative Research Center "Computational Intelligence" (SFB 531).

References

1. Droste, S., Jansen, T., Wegener, I.: Upper and lower bounds for randomized search heuristics in black-box optimization. Theory Comput. Syst. **39**(4), 525–544 (2006)
2. Friedrich, T., Oliveto, P.S., Sudholt, D., Witt, C.: Theoretical analysis of diversity mechanisms for global exploration. In: Proc. of GECCO '08, pp. 945–952. ACM, New York (2008)
3. Garnier, J., Kallel, L., Schoenauer, M.: Rigorous hitting times for binary mutations. Evol. Comput. **7**(2), 173–203 (1999)
4. Giel, O., Wegener, I.: Evolutionary algorithms and the maximum matching problem. In: Proc. of STACS '03, pp. 415–426. Springer, Berlin (2003)
5. Hajek, B.: Hitting-time and occupation-time bounds implied by drift analysis with applications. Adv. Appl. Probab. **13**(3), 502–525 (1982)
6. Happ, E., Johannsen, D., Klein, C., Neumann, F.: Rigorous analyses of fitness-proportional selection for optimizing linear functions. In: Proc. of GECCO '08, pp. 953–960. ACM, New York (2008)
7. He, J., Yao, X.: Drift analysis and average time complexity of evolutionary algorithms. Artif. Intell. **127**(1), 57–85 (2001)
8. He, J., Yao, X.: A study of drift analysis for estimating computation time of evolutionary algorithms. Nat. Comput. **3**(1), 21–35 (2004)
9. Neumann, F., Oliveto, P.S., Witt, C.: Theoretical analysis of fitness-proportional selection: landscapes and efficiency. In: Proc. of GECCO '09, pp. 835–842. ACM, New York (2009)
10. Oliveto, P.S., Witt, C.: Simplified drift analysis for proving lower bounds in evolutionary computation. In: Proc. of PPSN'08. LNCS, vol. 5199, pp. 82–91. Springer, Berlin (2008)
11. Oliveto, P.S., He, J., Yao, X.: Evolutionary algorithms and the vertex cover problem. In: Proc. of CEC '07, pp. 1430–1438 (2007)
12. Oliveto, P.S., He, J., Yao, X.: Time complexity of evolutionary algorithms for combinatorial optimization: a decade of results. Int. J. Autom. Comput. **4**(3), 281–293 (2007)
13. Sasaki, G.H., Hajek, B.: The time complexity of maximum matching by simulated annealing. J. Assoc. Comput. Mach. **35**(2), 387–403 (1988)
14. Wegener, I.: Methods for the analysis of evolutionary algorithms on pseudo-Boolean functions. In: Sarker, R., Mohammadian, M., Yao, X. (eds.) Evolutionary Optimization. Kluwer Academic, Dordrecht (2001)