TECHNISCHE UNIVERSITÄT DORTMUND

REIHE COMPUTATIONAL INTELLIGENCE

COLLABORATIVE RESEARCH CENTER 531

Design and Management of Complex Technical Processes and Systems by means of Computational Intelligence Methods

Simplified Drift Analysis for Proving Lower Bounds in Evolutionary Computation

Pietro S. Oliveto and Carsten Witt

No. CI-247/08

Technical Report   ISSN 1433-3325   April 2008

Simplified Drift Analysis for Proving Lower Bounds in Evolutionary Computation

Pietro S. Oliveto[1]
School of Computer Science
University of Birmingham
Birmingham, UK

Carsten Witt[2]
Fakultät für Informatik, LS 2
Technische Universität Dortmund
Dortmund, Germany

April 23, 2008

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. Therefore, the simplified theorem is also a didactical contribution to the runtime analysis of EAs.

1 Introduction

Theoretical studies of the computational complexity of Evolutionary Algorithms (EAs) have appeared since the 1990s (see Oliveto, He and Yao [8]). Since then various mathematical techniques for the analysis of EAs have been constructed. An overview of many important tools can be found in Wegener [10].

Recently *drift analysis*, a technique that goes back to the 1940s (cf. the introduction in [4]), was introduced for the analysis of EAs by He and Yao [6]. The authors concentrated on using drift analysis for the obtainment of both

---

1

lower and upper bounds on the expected runtime of various EAs for different problems. Also other applications of the technique are presented by He and Yao [7]. Concerning lower bounds, Giel and Wegener [3] point out that a result 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 [3] for maximum matching, Oliveto, He and Yao [9] for vertex cover, Friedrich, Oliveto, Sudholt and Witt [2] for analyzing 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 [5] have recently introduced a simplified drift theorem.

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 its 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. Section 2 presents some background on drift analysis and the simplified drift theorem. A first example of the usage of the technique is given in Section 3. In Section 4 we show that the simplified drift theorem can also be used in the same setting of Happ et al. [5] and that even significantly stronger results than theirs are obtained for the (1+1)-EA with fitness-proportional selection for the optimization of linear functions. In Section 5 we choose the maximum matching problem as an advanced application to show that proofs are considerably simplified. We finish with some conclusions.

2 Previous Work and the Simplified Drift Theorem

The following theorem goes back to Hajek [4] and since then, has been restated in different forms several times (e.g., He and Yao [6] and Giel and Wegener [3]). 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 a bound on the moment-generating function of the one-step drift, to be checked for the final statement of the theorem to hold. By analyzing the original proof, it follows that the remaining conditions can be either rephrased or removed. In particular, there is no need for the values $\lambda(\ell)$

and $D(\ell)$ to be constant or $p(\ell)$ to be polynomial. In any case, for the theorem to be meaningful, it has to be assured that $D(\ell)$ is defined.

Theorem 1 (Hajek [4]) *Let $X_0, X_1, X_2, \ldots$ 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*

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

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

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

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

In the typical applications of Theorem 1 cited above, the main drift Condition $(*)$ is proved with $p(\ell)$ being a polynomial. Having accomplished this, it often easily follows that $D(\ell)$ is at most a constant. 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. [5] 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 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. [5] contains an additional condition on – in essence – the moment-generating function $E(\delta^{-(X_{t+1}-X_t)} \mid X_t \ge b(\ell))$ 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. [5]. With the aim of proving that the process does not pass the interval $[a, b]$ in exponential time if started above state b, we intuitively need the following two conditions:

3

- 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 $\sum_{j \in \mathbb{Z}} j \operatorname{Prob}(X_{t+1} = i + j \mid X_t = i) \geq \varepsilon$ 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, it must be exponentially unlikely to leave the interval towards the optimum by using large jumps. The random step length towards the optimum has to exhibit some exponential decay. This follows from $\operatorname{Prob}(X_{t+1} = i - j \mid X_t = i) \leq 1/(1+\delta)^{j-r}$ for constant $\delta, r > 0$ and all $i > a$, i.e., within and outside the interval. We will see that this 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 $\{0, 1, \ldots, N\}$ for an arbitrarily large integer N and obtain such convergence for free. Weaker conditions could be proven if applications in infinite search spaces are desired.

We are ready to state and prove our simplified drift theorem. Note an additional difference to the version by Happ et al. [5]: a and b do not need to be linear in the dimension of the search space.

Theorem 2 (Simplified Drift Theorem) *Let X_t, $t \geq 0$, be the random variables describing a Markov process over the state space $S := \{0, 1, \ldots, N\}$ 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 $[a, b]$ of the state space (of asymptotically growing length $b - a$) and three constants $\delta, \varepsilon, r > 0$ such that for all $t \geq 0$*

1. $E(\Delta_t(i)) \geq \varepsilon$ for $a < i < b$

2. $\operatorname{Prob}(\Delta_t(i) = -j) \leq 1/(1+\delta)^{j-r}$ for $i > a$ and $j \geq 1$

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

Proof: Abbreviating $\ell := b - a$, we apply Theorem 1 for suitable parameters $\lambda(\ell), D(\ell), p(\ell), L(\ell)$ and the given interval $[a, b]$. As ε and δ are not allowed to depend on ℓ, this will not be needed for λ, D and p either, hence we omit the index ℓ. Moreover, we set $g := \operatorname{id}$. The following argumentation is also inspired by Hajek's work [4].

Fix $t \geq 0$ and some i such that $a < i < b$ and denote $p_j := \operatorname{Prob}(\Delta_t(i) = j)$. To prove Condition $(*)$, it is sufficient to identify a constant $\lambda > 0$ such that

$$S(\lambda) := \sum_{j \in \mathbb{Z}} e^{\lambda j} p_{-j} < 1.$$

4

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

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

where all series converge absolutely for any $\lambda > 0$ since $p_j = 0$ for $|j| > N$; however, their limits might depend on N. Identifying $E(\Delta_t(i))$ and using the first condition of the theorem, the drift, we obtain for all $\gamma \geq \lambda$

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

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

$$S(\lambda) \leq 1 - \lambda \varepsilon + \lambda \cdot \frac{\varepsilon}{2C(\gamma)} \cdot C(\gamma) = 1 - \frac{\lambda \varepsilon}{2} < 1$$

as desired. Choosing $\gamma := \ln(1 + \delta/2)$, which does not depend on N, and exploiting the second condition yields

$$C'(\gamma) := \sum_{j \in \mathbb{Z}} e^{\gamma j} p_{-j} \leq \sum_{j \leq 0} (1 + \delta/2)^j + \sum_{j \geq 1} \frac{(1 + \delta/2)^j}{(1 + \delta)^{j-r}}$$

$$\leq \left(1 + \frac{2}{\delta}\right) + (1 + \delta)^r \cdot \left(2 + \frac{2}{\delta}\right) \leq (1 + \delta)^r \left(3 + \frac{4}{\delta}\right),$$

hence $C(\gamma) \leq (1 + \delta)^r (3 + 4/\delta)/\ln^2(1 + \delta/2)$, which does not depend on N either. Since, moreover, ε, δ, r do not depend on $\ell = b - a$, neither will $C(\gamma)$, γ, λ, nor our bound on $S(\lambda)$. This establishes Condition $(*)$ of Theorem 1 for $p = p(\ell) = O(1)$ and $\lambda = \lambda(\ell) = \Omega(1)$.

To bound the probability of a success within $L(\ell)$ steps, we still need a bound on $D = D(\ell) = \max\{1, E(e^{-\lambda(X_{t+1}-b)} \mid X_t \geq b)\}$. Since $\lambda \leq \gamma$ and $X_t \geq b$, the expectation in the maximum is at most $C'(\gamma) = O(1)$, hence also $D = O(1)$. Altogether, we have $e^{-\lambda \ell} Dp = 2^{-\Omega(\ell)} = 2^{-\Omega(b-a)}$. Choosing $L(\ell) = 2^{c^*(b-a)}$ for some sufficiently small constant $c^* > 0$, Theorem 1 yields $\text{Prob}(T(\ell) \leq L(\ell)) \leq L(\ell) \cdot 2^{-\Omega(b-a)} = 2^{-\Omega(b-a)}$, and the theorem follows. $\square$

Our drift theorem can easily be applied to Randomized Local Search (RLS), which flips only one bit per iteration. Then Condition 2 is trivial and the theorem resembles the well-known *Gambler's Ruin* Theorem [1]. 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 $p_m := 1/n$;
    2. Replace x with x' iff $f(x') \geq f(x)$;

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 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. This is however not everywhere made explicit in this preliminary technical report.

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

In this section we present a first application of Theorem 2. We choose the Needle-in-a-haystack function which is well known to be hard for EAs. 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:

$$\text{NEEDLE}_n(x) = \left\{ \begin{array}{ll} 1 & \text{if } x = 1^n, \text{ and} \\ 0 & \text{otherwise.} \end{array} \right.$$

Theorem 3 *With probability at most $2^{-\Omega(n)}$ the (1+1)-EA finds the optimum of the NEEDLE_n function in less than 2^{cn} steps for some constant $c > 0$.*

Proof: Let X_t denote the number of zeroes in the bit string at time step t. We set $a := 0$, the global optimum and $b := n/2 - \gamma n$ for an arbitrarily small constant $\gamma > 0$. 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 2 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 $\varepsilon = 2\gamma$.

Condition 2 is: $\text{Prob}(\Delta(i) = -j) \leq 1/(1+\delta)^{j-r}$. In order to reach state $i-j$ from state i, at least j bits have to flip. Hence $\text{Prob}(\Delta(i) = -j) \leq \binom{n}{j}(1/n)^j \leq 1/j! \leq (1/2)^{j-1}$, which proves the condition for $\delta = 1$ and $r = 1$ even independently of i and of acceptance. So from Theorem 2 it follows for a constant $c^* > 0$

6

that the global optimum is found in $2^{c^*b} = 2^{cn}$ steps, where $c := c^*b/n > 0$ is a different constant, with probability at most $2^{-\Omega(b)} = 2^{-\Omega(n)}$. □

In the previous proof also $a = n/2 - 2\gamma n$ could have been used, implying that the (1+1)-EA most of the time is at distance almost $n/2$ from the optimum.

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

Recently Happ et al. [5] 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 works as follows:

(1+1)-EA with Fitness-proportional Selection

- 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 $p_m := 1/n$;
    2. Replace x with x' with probability $f(x')/(f(x') + f(x))$;

A function $f \colon \{0, 1\}^n \to \mathbb{R}$ is *linear* if it can be written as $f(x_1, \ldots, x_n) = w_0 + w_1 x_1 + \cdots + w_n x_n$ with coefficients $w_i \geq 0$, $0 \leq i \leq n$. In the special case $w_1 = \cdots = w_n = 1$ and $w_0 = 0$ we obtain the ONEMAX function counting the number of ones of the bit string. Concerning linear functions, Happ et al. [5] prove that with overwhelming probability at most $0.97n$ bits are set correctly by the (1+1)-EA with fitness-proportional selection after an exponential number of steps. We show that Theorem 2 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 [5].

Theorem 4 *Let $0 < \eta \leq 1/4$ be an arbitrarily small constant. Then there is a constant $c > 0$ such that the (1+1)-EA with fitness-proportional selection for linear functions (for ONEMAX) sets at most $2n/3 + \eta n$ (resp. at most $n/2 + \eta n$) bits correctly in 2^{cn} steps with probability at most $2^{-\Omega(n)}$.*

Proof: Setting $a := n/3 - 2\gamma n$ and $b := n/3 - \gamma n$, where $\gamma := \eta/2 \leq 1/8$, and given a current number of $a < i < b$ zeroes, let $\Delta(i)$ and $\Delta^{\mathrm{sel}}(i)$ denote the random change in this number before and after selection, respectively. Using the arguments from the proof of Theorem 3, we get $E(\Delta(i)) = (n-2i)/n \geq 1/3 + 2\gamma$.

$E(\Delta(i))$ is mostly determined by small steps. Choosing $r := \gamma n/4$, define $\mathbb{1}_r := \mathbb{1}\{|\Delta(i)| \leq r\}$ as the indicator r. v. for the event $|\Delta(i)| \leq r$. Since flipping at least k bits in a step has probability at most $1/k!$ and at most n bits flip,

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

7

and accordingly for $E(\Delta^{\mathrm{sel}}(i))$. By concentrating on steps of length at most r, we therefore introduce only an exponentially small error.

$\Delta(i)$ can be decomposed according to $\Delta(i) := \Delta^+(i) - \Delta^-(i)$, where $\Delta^+(i) := \Delta(i) \cdot \mathbb{1}\{\Delta(i) > 0\}$ and $\Delta^-(i) := -\Delta(i) \cdot \mathbb{1}\{\Delta(i) < 0\}$. By considering only the flipping 0-bits, we get $E(\Delta^-(i)) \leq 1/3 - \gamma$. Using $E(\Delta(i)) \geq 1/3 + 2\gamma$, we obtain $E(\Delta^+(i))/E(\Delta^-(i)) = (E(\Delta(i)) + E(\Delta^-(i)))/E(\Delta^-(i)) \geq 2 + 3\gamma$.

We get a lower bound on $E(\Delta^{\mathrm{sel}}(i))$ by weighting $\Delta^-(i)$ with upper bounds (here 1) on the selection probability and $\Delta^+(i)$ with lower bounds. For the lower bounds, we pessimistically assume all zeroes of the current string x to have coefficients 0 and $w_0 = 0$. Then $f(x') \leq f(x)$ for all offspring x' of x and the selection probability is at least $\frac{f(x')}{f(x')+f(x)} \geq \frac{f(x')}{2f(x)}$, which is linear w.r.t. $f(x')$. We assume at most r flipping bits. If a random subset of r out of $n - i \geq \frac{n}{2}$ ones flips, each bit flips with probability $\frac{r}{n-i}$. Using the linearity of expectation and of f, the expected offspring value $e(x')$ is at least $f(x)(1 - \frac{r}{n-i})$. Thus, using the law of total probability, the selection probability for the random x' is at least $e(x')/(2f(x)) \geq \frac{1}{2} - \frac{r}{2(n-i)} \geq \frac{1}{2} - \frac{\gamma}{4}$. Since the bound is independent of $\Delta(i) \cdot \mathbb{1}_r$,

$$
\begin{aligned}
E(\Delta^{\mathrm{sel}}(i)) &\geq \left(\frac{1}{2} - \frac{\gamma}{4}\right) E(\Delta^+(i) \cdot \mathbb{1}_r) - E(\Delta^-(i) \cdot \mathbb{1}_r) - 2^{-\Omega(n)} \\
&\geq \left(\frac{1}{2} - \frac{\gamma}{4}\right)(2 + 3\gamma)E(\Delta^-(i) \cdot \mathbb{1}_r) - E(\Delta^-(i) \cdot \mathbb{1}_r) - 2^{-\Omega(n)} \\
&\geq \left(\gamma - \frac{3\gamma^2}{4}\right) E(\Delta^-(i) \cdot \mathbb{1}_r) - 2^{-\Omega(n)} \geq \frac{29\gamma}{32} \cdot \frac{1}{36} - 2^{-\Omega(n)} \geq \frac{\gamma}{40}
\end{aligned}
$$

for n large enough, where we have used that $\gamma \leq 1/8$ along with $E(\Delta^-(i)) \geq (1/3 - 2\gamma)(1 - 1/n)^{n-1} \geq (1/3 - 2\gamma)/e \geq 1/36$, which follows by considering only 1-bit mutations. This bounds the drift for general linear functions by a constant.

With ONEMAX, the situation is even simpler. Since then f equals the number of ones, we can bound the probability of accepting a string x' with up to r more ones than x by $f(x')/(2f(x)) \leq (f(x) + r)/(2f(x)) \leq f(x)(1 + \gamma/2)/(2f(x)) = 1/2 + \gamma/4$ using $f(x) \geq n/2$. Setting $a := n/2 - 2\gamma n$ and $b := n/2 - \gamma n$, a similar calculation as in the third paragraph of this proof yields $E(\Delta^+(i))/E(\Delta^-(i)) \geq 1 + 2\gamma$. Finally, we obtain $E(\Delta^{\mathrm{sel}}(i)) \geq ((1/2 - \gamma/4)(1 + 2\gamma) - (1/2 + \gamma/4)) \cdot E(\Delta^-(i)) - 2^{-\Omega(n)} \geq \gamma/100$ for n large enough in the same manner as above.

The rest of the argumentation, in particular the proof of Condition 2 of Theorem 2 carries over from the proof of Theorem 3. $\qquad\square$

5 An Advanced Application: Maximum Matching

Giel and Wegener [3] considered the graph depicted in Figure 1 to prove that the (1+1)-EA has an expected runtime which is exponential in the number of
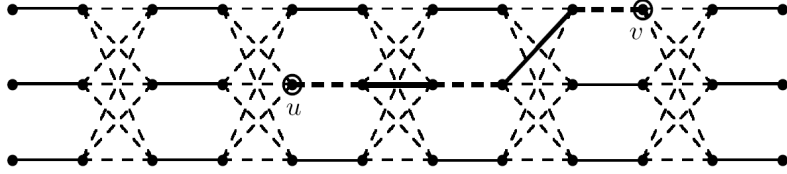
Figure 1: The $G_{h,\ell}$ graph (i.e. $h = 3$ and $\ell = 11$) with an almost perfect matching and its augmenting path.

graph edges for the well known *maximum matching* problem in the worst case. One of the crucial parts of their proof is represented by the following theorem.

Theorem 5 *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 $h \geq 3$.*

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, a move of length $j = 1$ occurs 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 2, we set $a := 0$, the minimum augmenting path length and $b := \ell - 1$ where ℓ is its maximum length.

Usually there are $2h$ edges adjacent to the augmenting path, h at each side, that flipped together with the 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 performing a move of length 1 *lengthening* the augmenting path of length i is only bounded by $p_1(i) \geq (h/m^2)(1 - 1/m)^{m-2}$, where m is the number of edges of the graph. On the other hand, the probability to shorten the augmenting path with a move of length 1 is bounded from above by $p_{-1}(i) \leq (2/m^2)(1 - 1/m)^{m-2} + 3/m^4$ (see [3]). Since most other mutations of the (1+1)-EA due to worse fitness will be rejected in this setting, we use the condition R that a step is *relevant* meaning it is accepted and changes the current state. The probability p_{rel} of a relevant step is bounded by $(1/m^2)(1 - 1/m)^{m-2} \leq p_{\mathrm{rel}} \leq (2h+2)/m^2$.

Let $R(i) = (\Delta(i) \mid R)$ denote the random increase of the path length in relevant steps for a current length 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

$$E(R_1(i)) = \frac{p_1(i)}{p_{\text{rel}}} - \frac{p_{-1}(i)}{p_{\text{rel}}} \geq \frac{h - 2 - O(m^{-2})}{2h + 2} \geq \frac{1}{8} - O(m^{-2})$$

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

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

because $p_{-j} \leq (j+1)/m^{2j}$ [3].

Hence, the total conditional drift is

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

and Condition 1 is proved. Condition 2, with $\delta = 1$ and $r = 3$, follows from

$$\frac{p_{-j}}{p_{\text{rel}}} \leq \min\left\{1, \frac{j+1}{m^{2j}} \cdot em^2\right\} \leq \min\left\{1, \frac{1}{m^{2j-7}}\right\} \leq \left(\frac{1}{2}\right)^{j-3}$$

for $m \geq 2$. From Theorem 2, the proof follows. $\qquad\square$

The bounds on $p_j(i)$ by Giel and Wegener [3] do not imply $p_j(i) \geq p_{-j}(i)$ for every j, hence the theorem by Happ et al. [5] 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.

6   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 mentioned. 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. [5] 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 2 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.

References

[1] W. Feller. *An introduction to probability theory and its applications*, volume 1. Wiley, 1968.

[2] T. Friedrich, P. S. Oliveto, D. Sudholt, and C. Witt. Theoretical analysis of diversity mechanisms for global exploration. In *Proc. of GECCO '08*, 2008. (to appear).

[3] O. Giel and I. Wegener. Evolutionary algorithms and the maximum matching problem. In *Proc. of STACS '03*, pages 415–426. Springer, 2003.

[4] B. Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability*, 13(3):502–525, 1982.

[5] E. Happ, D. Johannsen, C. Klein, and F. Neumann. Rigorous analyses of fitness-proportional selection for optimizing linear functions. In *Proc. of GECCO '08*, 2008. (to appear).

[6] J. He and X. Yao. Drift analysis and average time complexity of evolutionary algorithms. *Artificial Intelligence*, 127(1):57–85, 2001.

[7] J. He and X. Yao. A study of drift analysis for estimating computation time of evolutionary algorithms. *Natural Computing*, 3(1):21–35, 2004.

[8] P. S. Oliveto, J. He, and X. Yao. Computational complexity analysis of evolutionary algorithms for combinatorial optimization: A decade of results. *International Journal of Automation and Computing*, 4(3):281–293, 2007.

[9] P. S. Oliveto, J. He, and X. Yao. Evolutionary algorithms and the vertex cover problem. In *Proc. of CEC '07*, pages 1430–1438, 2007.

[10] I. Wegener. Methods for the analysis of evolutionary algorithms on pseudo-boolean functions. In R. Sarker, M. Mohammadian, and X. Yao, editors, *Evolutionary Optimization*. Kluwer Academic Publishers, 2001.