

# (1+1) EA on Generalized Dynamic OneMax

Timo Kötzing  
Friedrich-Schiller-Universität  
Jena, Germany

Andrei Lissovoi  
DTU Compute, Technical  
University of Denmark  
Kongens Lyngby, Denmark

Carsten Witt  
DTU Compute, Technical  
University of Denmark  
Kongens Lyngby, Denmark

## ABSTRACT

Evolutionary algorithms (EAs) perform well in settings involving *uncertainty*, including settings with stochastic or dynamic fitness functions. In this paper, we analyze the (1+1) EA on dynamically changing ONEMAX, as introduced by Droste (2003). We re-prove the known results on first hitting times using the modern tool of drift analysis. We extend these results to search spaces which allow for more than two values per dimension.

Furthermore, we make an *anytime analysis* as suggested by Jansen and Zarges (2014), analyzing how closely the (1+1) EA can track the dynamically moving optimum over time. We get tight bounds both for the case of bit strings, as well as for the case of more than two values per position. Surprisingly, in the latter setting, the expected quality of the search point maintained by the (1+1) EA does not depend on the number of values per dimension.

## Categories and Subject Descriptors

F.2 [Theory of Computation]: Analysis of Algorithms and Problem Complexity

## General Terms

Theory, algorithms

## Keywords

evolutionary computation, dynamic optimization, drift, theory

## 1. INTRODUCTION

Randomized search heuristics, such as evolutionary algorithms (EAs), are general purpose optimization algorithms applicable to virtually any (formal) optimization task. In particular, EAs (and other randomized search heuristics) have been applied very successfully in domains featuring *uncertainty*; for example, the objective functions (so-called fitness functions) can be *randomized* or *dynamically changing*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

FOGA'15, January 17–20, 2015, Aberystwyth, UK.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-3434-1/15/01 ...\$15.00.

<http://dx.doi.org/10.1145/2725494.2725502>.

(see Jin and Branke, 2005 for an excellent survey on evolutionary algorithms in settings featuring uncertainty). In this paper we focus on dynamically changing fitness functions.

We call a fitness function *dynamic* if the fitness values of search points depend on the iteration number (but might be *deterministic* for each iteration). For example, the shortest path between two cities might depend on whether it is rush hour or not. The classical task of an optimization algorithm is to find the best solution it can (in terms of fitness); for dynamic optimization, there need not be a single solution which is good at all times: solutions that are good now might be bad later. Thus, algorithms in this domain need to be able to find *and track* the optimal solution (or at least a good solution) over time as the problem changes.

With this paper we contribute to the *theoretical foundations* of randomized search heuristics, for the domain of dynamic fitness functions. While there has been a lot of work on the theory of randomized search heuristics in static settings (see Auger and Doerr, 2011; Neumann and Witt, 2010; Jansen, 2013), there are only a few works on dynamically changing fitness functions. The utility of a population for tracking problems was studied in evolutionary computation by Jansen and Schellbach (2005), while different mechanisms for ensuring population diversity have been considered by Oliveto and Zarges (2013). In particular, a mechanism called genotype diversity was proved to be inefficient on a particular dynamic problem. The papers by Kötzing and Molter (2012) and Lissovoi and Witt (2015) consider dynamic pseudo-Boolean functions where the optimum moves slowly from the all-ones to the all-zeros bit string; the papers show that, while the Max-Min Ant System is able to track the changes occurring in this fitness function, an evolutionary algorithm (in Lissovoi and Witt, 2015 using a population) loses track of the optimum. Jansen and Zarges (2014) analyzed the performance of a standard evolutionary algorithm on a dynamically changing fitness function, introducing “anytime analysis”, the expected distance to the optimum at any given point in time.

The oldest theoretical running time analyses of evolutionary algorithms for dynamic fitness functions are probably due to Droste (2002, 2003). Here the fitness function is the (Hamming-) distance to a (dynamically changing) point in the hypercube (so-called *dynamic ONEMAX*). In each iteration, the current optimum is changed by flipping each bit with some fixed probability  $p$ ; from Droste (2003) we know that the standard (1+1) EA is able to *find* the optimum in polynomial time if and only if  $p = O(\log n/n^2)$ .

In this paper we build on the setting of dynamic ONEMAX from Droste (2003). We re-prove the classic results using the modern tools of drift theory and extend them as follows. First, we generalize the domain by allowing not only bit strings, but each position can take any of the values in  $\{0, \dots, r-1\}$ . Fitness is again distance from the current optimum; we measure distances as the sum of the distances of each component, where in each component we measure distance “with wrap around” (giving each component the metric space of a ring, see Section 2 for a detailed definition). We extend the (1+1) EA by letting mutation change any position independently with probability  $1/n$ ; any changed position is randomly increased or decreased by one (with probability  $1/2$  each). Note that similar extensions of the ONEMAX function (without dynamic changes) have been studied by Doerr, Johannsen, and Schmidt (2011) and Doerr and Pohl (2012); they considered arbitrary linear functions over  $\{0, \dots, r\}$ , and a mutation where changing a position means selecting a new value at this position uniformly at random (excluding the old value). We chose the ring topology, as we consider it more natural for a dynamically moving optimum, which can now never run into a boundary.

The second extension to Droste (2003) is that we do not only consider the first hitting time of the optimum but, as suggested by Jansen and Zarges (2014), we give an “anytime analysis”, an analysis of the distance to the optimum at any time of the search process.

We state our setting more formally in Section 2. In Section 3 we give our anytime results, considering cases with  $p = o(1/n)$ . The first part is about the case of bit strings (i. e.  $r = 2$ ), where we show that the distance to the optimum is (in the limit) strongly concentrated at  $\Theta(pn^2)$ . This gives an anytime result as suggested by Jansen and Zarges (2014). The second part shows that, for large  $r$ , the distance to the optimum in each dimension is strongly concentrated at  $O(1)$ , leading to an expected distance of  $O(n)$  from the optimum (again an anytime result). Note that this shows that the distance is independent of  $r$ .

In Section 4 we consider the expected hitting times of the (randomly moving) optimum. Here we re-prove the result of Droste (2003) (who considered the case of  $r = 2$ ) that the first hitting time is polynomial if  $p = O(\log n/n^2)$ . We use modern drift theory, leading to a much shorter and more elegant proof, resulting in a better bound. We extend this result to arbitrary  $r$ .

Droste (2003) also gave a lower bound, which shows that for  $p = \omega(\log n/n^2)$  we do not get polynomial hitting times; in Section 5 we re-prove this result (again with modern drift theory) and extend them to arbitrary values of  $r$ .

As mentioned, we will use modern drift theory to derive our results. In Section 2 we restate known drift theorems, partly in more general form than before, and also present new variants. A new theorem regards variable drift, which allows for *negative* drift close the optimum and shows how stochastic processes can bridge such an area of headwind. For our anytime analysis, the crucial tool is a lemma by Lissovoi and Witt (2015), which we restate as Lemma 6 below, effectively turning expected drift into probabilities about deviating from the target of the drift after having reached that target.

## 2. PRELIMINARIES

In this section we first make our setting formal (see Section 2.1) and then give a number of helpful theorems, both from the literature and new theorems (see Section 2.2).

### 2.1 Setting

Droste (2003) proposes a dynamic version of ONEMAX and analyses the performance of (1+1) EA on this dynamic fitness function in terms of first hitting times of the optimum. We extend this dynamic version of ONEMAX as follows.

For all  $r \in \mathbb{N}$ , let  $[r] = \{0, \dots, r-1\}$ ; for two elements  $x, y \in [r]$ , we let  $d(x, y) = \min((y-x) \bmod r, (x-y) \bmod r)$  (intuitively,  $d$  is the metric of  $[r]$  with wrap-around). We consider the search space  $[r]^n$  (note that  $r = 2$  gives the standard setting of bit strings).

Given a current optimum  $a$ , we let

$$\text{ONEMAX}_a : [r]^n \rightarrow \mathbb{R}, x \mapsto \sum_{i=1}^n d(a_i, x_i).$$

The goal of the (1+1) EA is to evolve and *maintain* bit strings with as small as possible  $\text{ONEMAX}_a$ -value. In particular, in this setting optimization means *minimizing*.

We consider the following mutation operator on  $[r]^n$ , parametrized by  $p \in [0, 1]$ . Given  $x \in [r]^n$ , create mutant  $x'$  by choosing, for each component  $i \leq n$  independently,

$$x'_i = \begin{cases} x_i + 1 \bmod r, & \text{with probability } p/2; \\ x_i - 1 \bmod r, & \text{with probability } p/2; \\ x_i, & \text{with probability } 1 - p. \end{cases}$$

We use this operator with  $p = 1/n$  for the (1+1) EA (see Algorithm 1).

---

#### Algorithm 1: (1+1)-EA

---

```

1 choose  $x \in [r]^n$  uniformly at random;
2 repeat
3    $x' \leftarrow \text{mutate}(x)$ ;
4   if  $\text{fitness}(x') \leq \text{fitness}(x)$  then
5      $x \leftarrow x'$ ;
6 until forever;
```

---

In each iteration, we change the optimum by applying the mutation operator with some fixed  $p$ . This extends the setting of Droste (2003), where only the case of  $r = 2$  was addressed.

### 2.2 Drift Theorems

In this section we first discuss drift theorems regarding first hitting times, and afterwards discuss how one can turn statements about the drift into statements about occupation probabilities (of a random process).

#### 2.2.1 First Hitting Times

As mentioned in the introduction, almost all of our proofs use state-of-the-art drift statements, many of which were not available to Droste (2003). The simplest case is the one of additive drift, as described in the following theorem. It goes back to He and Yao (2001); however, is presented in a more general form here, which is proved in Lehre and Witt (2014).

**THEOREM 1 (ADDITIVE DRIFT, EXPECTED TIME).**

Let  $(X^t)_{t \geq 0}$ , be a stochastic process, adapted to some filtration  $\mathcal{F}_t$ , over a bounded state space  $S \subseteq \mathbb{R}_0^+$ . Let  $T_0 := \min\{t \geq 0 : X^t = 0\}$  denote the first hitting time of 0 and assume that both  $\mathbb{E}(X^0)$  and  $\mathbb{E}(T_0 | X^0)$  are finite. Then:

- (i) If  $\mathbb{E}(X^t - X^{t+1} | \mathcal{F}_t; X^t > 0) \geq \varepsilon$  then  $\mathbb{E}(T_0 | X^0) \leq \frac{X^0}{\varepsilon}$ .
- (ii) If  $\mathbb{E}(X^t - X^{t+1} | \mathcal{F}_t; X^t > 0) \leq \varepsilon$  then  $\mathbb{E}(T_0 | X^0) \geq \frac{X^0}{\varepsilon}$ .

Intuitively, the filtration  $\mathcal{F}_t$  describes the history of the process up to time  $t$ . For Markov processes, it simplifies to the state at time  $t$ ; for instance, the first drift condition would read  $\mathbb{E}(X^t - X^{t+1} | X^t; X^t > 0) \geq \delta$  instead.

Often, the state space  $Z$  of the underlying stochastic process and the support  $S$  of the random variables  $X^t$  are not identical. Obviously, this is the case if the state space is not a subset of  $\mathbb{R}$ , e.g., if we are dealing with bit strings, where we have  $Z := \{0, 1\}^n$ . In particular, even if the state space is the real numbers, it might be convenient to introduce a so-called potential function (also called Lyapunov function)  $g: Z \mapsto S$ , which leads to a new stochastic process  $Y^t := g(X^t)$  on the new state space  $S$ . One reason might be that the drift of  $Y^t$  is easier to compute. We abstract away from this mapping by allowing the random variables  $X^t$  from Theorem 1 to represent a process obtained after any transformation of the original process using a potential function. Such a transformation might turn Markovian processes into non-Markovian ones.

Theorem 1 is only concerned with bounds on the expected value of the first hitting time of the target state 0. Recently, it has been shown in Kötzing (2014) that the first hitting time is sharply concentrated (exhibits so-called tail bounds) if additional assumptions are made on the step size. We restate this in the following theorem.

**THEOREM 2 (ADDITIVE DRIFT, TAIL BOUNDS).**

Let the prerequisites of Theorem 1 hold and assume additionally that  $|X^t - X^{t+1}| < c$  for some  $c > 0$  and all  $t \geq 0$ . Then:

- (i) If  $\mathbb{E}(X^t - X^{t+1} | \mathcal{F}_t; X^t > 0) \geq \varepsilon$  then  $\Pr(T_0 > s) \leq \exp(-s\varepsilon^2/(16c^2))$  for all  $s \geq 2X^0/\varepsilon$ .
- (ii) If  $\mathbb{E}(X^t - X^{t+1} | \mathcal{F}_t; X^t > 0) \leq \varepsilon$  then  $\Pr(T_0 < s) \leq \exp(-(X^0)^2/(16c^2s))$  for all  $s \leq X^0/(2\varepsilon)$ .

The previous two theorems dealt with a drift towards the target state 0. If the drift is directed away from the target, lower bounds on the hitting time can be proved. This is the realm of *negative drift theorems*, several variants of which exist (see Oliveto and Witt, 2011, 2012 for the original version). In this work we use the following version, adapted from Rowe and Sudholt (2014, Theorem 4), which takes into account the probabilities of staying at a state. For technical reasons, it is restricted to Markov processes and the use of a possible potential function is made explicit.

**THEOREM 3 (NEGATIVE DRIFT WITH SELF-LOOPS).**

Let  $(X^t)_{t \geq 0}$ , be a Markov process over a state space  $S$ . Suppose there exist an interval  $[a, b] \subseteq \mathbb{R}_0^+$ , two constants  $\delta, \varepsilon > 0$ , a function  $r(\ell)$  satisfying  $1 \leq r(\ell) = o(\ell/\log(\ell))$ , and a potential function  $g: S \rightarrow \mathbb{R}_0^+$ , such that for all  $t \geq 0$ , the following two conditions hold:

$$(i) \mathbb{E}(\Delta^t | X^t; a < g(X^t) < b) \geq \varepsilon(1 - p_{t,0}),$$

$$(ii) \Pr(|\Delta^t| \geq j | X^t; a < g(X^t)) \leq \frac{r(\ell)(1-p_{t,0})}{(1+\delta)^j} \text{ for } j \in \mathbb{N}_0,$$

where  $\Delta^t = g(X^{t+1}) - g(X^t)$  and  $p_{t,0} := \Pr(\Delta^t = 0 | X^t)$ .

Then there is a constant  $c^* > 0$  such that for  $T^* := \min\{t \geq 0 : g(X^t) \leq a | g(X^0) \geq b\}$  it holds

$$\Pr(T^* \leq 2^{c^* \ell/r(\ell)}) = 2^{-\Omega(\ell/r(\ell))}.$$

Intuitively, drift away from the target makes it difficult to reach the target. Nevertheless, if the drift is negative only for a few states and directed towards the target at the remaining states, the expected first hitting time of the target might still be small. Such a scenario of “headwind drift” on the way towards the target will appear in our analyses if the probability of flipping a bit of the optimum is small, e.g.,  $p = O((\log n)/n^2)$ , resulting in only very few states close to the target having negative drift.

The following novel theorem proves upper bounds in the presence of possibly negative drift. The bounds  $\delta(i)$  are lower bounds on the drift at state  $i$ , pessimistically assuming that all steps towards the target improve only by 1. The  $p^-(i)$  and  $p^+(i)$  are bounds on the probability of improving by at least 1 and worsening by at least 1, respectively. The theorem is general enough to analyze different scenarios, e.g., blind random walks on the hypercube. However, we will mostly apply Corollary 5, which is easier to use.

For notational convenience, we state the theorem only for Markov processes, however, it can easily be generalized to non-Markovian ones. Extensions to continuous search spaces seem also possible; however, these are not straightforward. Therefore, the state space is restricted to be non-negative integers.

**THEOREM 4 (HEADWIND DRIFT, UPPER BOUND).**

Let  $(X^t)_{t \geq 0}$  be a Markov process on  $\{0, \dots, N\}$ . Let bounds

$$p^-(i) \leq \Pr(X^{t+1} \leq i - 1 | X^t = i)$$

and

$$p^+(i) \geq \Pr(X^{t+1} \geq i + 1 | X^t = i),$$

where  $0 \leq i \leq N$ , be given, and define

$$\delta(i) := p^-(i) - \mathbb{E}((X^{t+1} - i) \cdot \mathbf{1}\{X^{t+1} > i\} | X^t = i).$$

Assume that  $\delta(i)$  is monotone increasing w. r. t.  $i$  and let  $\kappa \geq \max\{i \geq 0 | \delta(i) \leq 0\}$  (noting that  $\delta(0) \leq 0$ ). The function  $g: \{0, \dots, N+1\} \rightarrow \mathbb{R}^+$  is defined by

$$g(i) := \sum_{k=i+1}^N \frac{1}{\delta(k)}$$

for  $i \geq \kappa$  (in particular,  $g(N) = g(N+1) = 0$ ), and inductively by

$$g(i) := \frac{1 + (p^+(i+1) + p^-(i+1))g(i+1)}{p^-(i+1)}$$

for  $i < \kappa$ .

Then it holds for the first hitting time  $T := \min\{t \geq 0 | X^t = 0\}$  of state 0 that

$$\mathbb{E}(T | X^0) \leq g(0) - g(X^0).$$

*Remark.*  $\delta(i)$  respects the following simple lower bound:

$$\delta(i) \geq \mathbb{E}((i - X^{t+1}) \cdot \mathbb{1}\{X^{t+1} \geq i - 1\} \mid X^t = i).$$

PROOF. We will prove that  $g(i)$  is a monotone decreasing function and can be used as a potential function to satisfy the drift condition

$$\mathbb{E}(g(X^{t+1}) - g(i) \mid X^t = i \wedge i > 0) \geq 1.$$

Due to the monotonicity of  $g(i)$ , the first hitting time where  $g(X^t) = g(0)$  equals the first hitting time where  $X^t = 0$  for the original  $X^t$ -process. Then the theorem follows by the additive drift theorem (Theorem 1).

To prove the monotonicity of  $g(i)$ , we observe that  $g(i) - g(i+1) \geq 0$  for  $i \geq \kappa$  immediately by definition. For  $i < \kappa$ , we get

$$\begin{aligned} \frac{g(i)}{g(i+1)} &= \frac{\frac{1}{g(i+1)} + (p^+(i+1) + p^-(i+1))}{p^-(i+1)} \\ &\geq \frac{p^+(i+1) + p^-(i+1)}{p^-(i+1)} \geq 1, \end{aligned}$$

where we used  $g(i+1) \geq 0$ . This completes the proof of the monotonicity.

To prove the drift condition, we distinguish between two cases. Suppose  $X^t = i > \kappa$ . The monotonicity of the  $\delta(i)$  implies the ‘‘concavity’’ condition  $g(i-1) - g(i) \geq g(i) - g(i+1)$  for  $i > \kappa$ . We obtain

$$\begin{aligned} &\mathbb{E}(g(X^{t+1}) - g(i) \mid X^t = i) \\ &\geq p^-(i)(g(i-1) - g(i)) \\ &\quad - \left( \sum_{k=1}^N (g(i) - g(i+k)) \Pr(X^{t+1} = i+k) \right) \\ &\geq p^-(i)(g(i-1) - g(i)) \\ &\quad - \left( \sum_{k=1}^N k(g(i-1) - g(i)) \Pr(X^{t+1} = i+k) \right) \\ &= p^-(i)(g(i-1) - g(i)) \\ &\quad - (g(i-1) - g(i)) \mathbb{E}((X^{t+1} - i) \mathbb{1}\{X^{t+1} > i\} \mid X^t = i) \\ &= (g(i-1) - g(i)) \delta(i) = \frac{1}{\delta(i)} \cdot \delta(i), \end{aligned}$$

where the second inequality used the concavity repeatedly. If  $X^t = i \leq \kappa$ , we pessimistically assume all steps towards the target to reach  $i-1$  and all away from it to reach  $N$  (resulting in zero  $g$ -value). Hence, using the definition of  $g(i-1)$ ,

$$\begin{aligned} &\mathbb{E}(g(X^{t+1}) - g(i) \mid X^t = i) \\ &\geq p^-(i)(g(i-1) - g(i)) - p^+(i)g(i) \\ &= p^-(i) \left( \frac{1 + (p^+(i) + p^-(i))g(i)}{p^-(i)} - g(i) \right) - p^+(i)g(i) \\ &= 1, \end{aligned}$$

which proves the bound on the drift and, therefore, the theorem.  $\square$

We now state the announced corollary, which gives us a closed expression for the expected first hitting time  $\mathbb{E}(T \mid X^0)$ . This expression involves the factor  $\sum_{k=\kappa+1}^N \frac{1}{\delta_k}$  that is reminiscent of the formula for the expected first hitting time of state  $\kappa$  under variable drift towards the

target (see, e.g., Rowe and Sudholt, 2014 for a formulation of the variable drift theorem). For the states less than  $\kappa$ , where drift away from the target holds, the product  $\prod_{k=1}^{\kappa} \frac{p^+(k) + p^-(k)}{p^-(k)}$  comes into play. Intuitively, it represents the waiting time for the event of taking  $\kappa$  consecutive steps against the drift. Since the product involves probabilities conditioned on leaving the states, which effectively removes self-loops, another sum of products must be added. This sum, represented by the second line of the expression for  $\mathbb{E}(T \mid X^0)$ , intuitively accounts for the self-loops.

COROLLARY 5. *Let the assumptions of Theorem 4 hold. Then*

$$\begin{aligned} \mathbb{E}(T \mid X^0) &\leq \left( \left( \sum_{k=\kappa+1}^N \frac{1}{\delta_k} \right) \left( \prod_{k=1}^{\kappa} \frac{p^+(k) + p^-(k)}{p^-(k)} \right) \right) \\ &\quad + \left( \sum_{k=1}^{\kappa} \frac{1}{p^-(k)} \prod_{j=1}^{k-1} \frac{p^+(j) + p^-(j)}{p^-(j)} \right). \end{aligned}$$

PROOF. It is sufficient to prove that the right hand side is an upper bound on the  $g(0)$  defined in Theorem 4. We note that  $\sum_{k=\kappa+1}^N \frac{1}{\delta_k} = g(\kappa)$ . The inductive expression for  $g(i)$  yields  $g(i) \leq \frac{1 + (p^+(i+1) + p^-(i+1))g(i+1)}{p^-(i+1)}$  for  $i \leq \kappa - 1$ . Inductively

$$\begin{aligned} g(0) &\leq \left( \prod_{j=0}^{\kappa-1} \frac{p^+(j+1) + p^-(j+1)}{p^-(j+1)} \right) g(\kappa) \\ &\quad + \left( \sum_{k=0}^{\kappa-1} \frac{1}{p^-(1)} \prod_{j=1}^k \frac{p^+(j) + p^-(j)}{p^-(j+1)} \right), \end{aligned}$$

and the corollary follows by index transformations and re-grouping terms.  $\square$

## 2.2.2 Occupation Probabilities

In this section we move away from analyses of the first hitting time of a state and direct our attention to so-called occupation probabilities. For the anytime analysis in Section 3, we want to make statements about how far from the optimum the (1+1) EA will stray, and with what probability. In particular, we want to know the probability that the current search point is more than  $j$  away from the optimum in iteration  $t$ , for large  $t$ . The idea is that a stochastic process  $(X^t)_{t \geq 0}$  on  $\mathbb{R}$  which has a drift towards 0 will, after hitting 0 for the first time, likely stay in the proximity of 0 and stray off only with a low probability. This is what is meant by ‘‘occupation probabilities’’.

Hajek (1982), in his third section, already gives some general bounds on these probabilities. This is also the idea of another lemma regarding occupation probabilities given in Lissovoi and Witt (2015, Lemma 13) (restated below as Lemma 6), from which we will here derive a simple version (Theorem 7), tailored to the case of additive drift and Markov processes with self-loops.

LEMMA 6 (LISSOVOI AND WITT, 2015, LEMMA 13). *Let  $(X^t)_{t \geq 0}$ , be a stochastic process, adapted to a filtration  $(\mathcal{F}_t)_{t \geq 0}$ , over some state space  $S \subseteq \{0\} \cup [x_{\min}, x_{\max}]$ , where  $x_{\min} > 0$ . Let  $a, b \in \{0\} \cup [x_{\min}, x_{\max}]$ ,  $b > a$ . Let  $h: [x_{\min}, x_{\max}] \rightarrow \mathbb{R}^+$  be such that  $1/h$  is integrable on  $[x_{\min}, x_{\max}]$  and define  $g: \{0\} \cup [x_{\min}, x_{\max}] \rightarrow \mathbb{R}^{\geq 0}$  by  $g(x) := \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^x \frac{1}{h(y)} dy$  for  $x \geq x_{\min}$  and  $g(0) := 0$ .*

If there exist  $\lambda > 0$ ,  $\beta < 1$  and  $D > 0$  such that

$$\begin{aligned} \mathbb{E}\left(e^{-\lambda(g(X^t)-g(X^{t+1}))} \cdot \mathbb{1}\{X^t > a\} \mid \mathcal{F}_t\right) &\leq \beta \\ \text{and } \mathbb{E}\left(e^{-\lambda(g(a)-g(X^{t+1}))} \cdot \mathbb{1}\{X^t \leq a\} \mid \mathcal{F}_t\right) &\leq D \end{aligned}$$

then

$$\Pr(X^t \geq b \mid X^0) < \beta^t \cdot e^{\lambda(g(X^0)-g(b))} + \frac{1-\beta^t}{1-\beta} D e^{\lambda(g(a)-g(b))}$$

for  $t > 0$ .

We give two definitions regarding Markov processes before giving our theorem regarding occupation probabilities. Let a Markov process  $(X^t)_{t \geq 0}$  on  $\mathbb{R}_0^+$  be given. We say that  $(X^t)_{t \geq 0}$  has step size at most  $c \in \mathbb{R}$  if, for all  $t$ ,  $|X^t - X^{t+1}| \leq c$ . We say that  $(X^t)_{t \geq 0}$  has self-loop probability at least  $p_0 \in \mathbb{R}$  iff, for all  $t$  such that  $X^t > 0$  we have  $\Pr(X^t = X^{t+1} \mid X^t) \geq p_0$ . From Lemma 6 we derive the following statement on occupation probabilities for the case of bounded step sizes.

**THEOREM 7 (OCCUPATION PROBABILITIES).**

Let a Markov process  $(X^t)_{t \geq 0}$  on  $\mathbb{R}_0^+$  with additive drift of at least  $d$  towards 0 be given, starting at 0 (i.e.  $X_0 = 0$ ), with step size at most  $c$  and self-loop probability at least  $p_0$ . Then we have, for all  $t \in \mathbb{N}$  and  $b \in \mathbb{R}_0^+$ ,

$$\Pr(X^t \geq b) \leq 2e^{\frac{2d}{3c(1-p_0)}(1-b/c)}.$$

**PROOF.** First, we define a new Markov process  $(Y^t)_{t \geq 0}$  obtained from  $X^t$  by omitting all steps that do not change the current state; formally, since we consider Markov chains, we have  $Y^t - Y^{t+1} = X^t - X^{t+1}$  in the conditional space where  $X^{t+1} \neq X^t$ . By definition of conditional probability and expectation, we obtain

$$\begin{aligned} \mathbb{E}(Y^t - Y^{t+1} \mid X^t) &= \frac{\mathbb{E}(X^t - X^{t+1} \mid X^t)}{\Pr(X^{t+1} \neq X^t \mid X^t)} \\ &\geq \frac{\mathbb{E}(X^t - X^{t+1} \mid X^t)}{1-p_0} \end{aligned}$$

since the probability of changing the state is at most  $1-p_0$ .

The theorem makes a statement for all  $t$ , however, to prove it, it is enough to consider steps that actually change state. Hence, in the following, the aim is to analyze the  $Y$ -process using Lemma 6 with  $a := x_{\min} := 0$  and the constant function  $h(x) := 1$ . From this we obtain the trivial potential function  $g(x) = x$ . From our prerequisites and the previous paragraph, we get

$$\begin{aligned} \mathbb{E}(g(Y^t) - g(Y^{t+1}) \mid Y_t; Y^t > 0) & \\ &= \mathbb{E}(Y^t - Y^{t+1} \mid Y_t; Y^t > 0) \\ &\geq \frac{\mathbb{E}(X^t - X^{t+1} \mid X^t; X^t > 0)}{1-p_0} \\ &\geq \frac{d}{1-p_0}. \end{aligned}$$

Let  $d^* := d/(1-p_0)$ . To bound the moment-generating function of the drift, we abbreviate  $\Delta^t := Y^t - Y^{t+1}$ . We already know that  $\mathbb{E}(\Delta^t \mid Y^t; Y^t > 0) \geq d^*$  and argue

$$\mathbb{E}\left(e^{-\lambda \Delta^t} \cdot \mathbb{1}\{Y^t > 0\} \mid Y^t\right) \leq \mathbb{E}\left(e^{-\lambda \Delta^t} \mid Y^t; Y^t > 0\right)$$

In the following, we condition on  $Y^t; Y^t > 0$  everywhere but omit this from the formulas for the sake of readability. Using the Taylor expansion of the exponential function, we get

$$\mathbb{E}\left(e^{-\lambda \Delta^t}\right) \leq 1 - \lambda \mathbb{E}(\Delta^t) + \sum_{k=2}^{\infty} \frac{\lambda^k \mathbb{E}(|\Delta^t|^k)}{k!},$$

which for any  $\eta \geq \lambda$  is at most

$$1 - \lambda \mathbb{E}(\Delta^t) + \frac{\lambda^2}{\eta^2} \sum_{k=2}^{\infty} \frac{\eta^k \mathbb{E}(|\Delta^t|^k)}{k!}.$$

Now, by setting  $\eta := 1/c$ ,  $\lambda := 2d^*/(3c^2)$  and noting that  $\Delta^t \leq c$  (also for the  $Y$ -process), we get the bound

$$1 - \lambda \mathbb{E}(\Delta^t) + \lambda \frac{2d^*}{3c^2 \cdot (1/c^2)} \cdot \sum_{k=2}^{\infty} \frac{1}{k!} \leq 1 - \lambda \mathbb{E}(\Delta^t) + \lambda \frac{d^*}{2}$$

where the last inequality used that  $\sum_{k=2}^{\infty} \frac{1}{k!} = e - 2 \leq 3/4$ . Altogether, using  $\mathbb{E}(\Delta^t) \geq d^*$ , we get

$$\begin{aligned} \mathbb{E}\left(e^{-\lambda \Delta^t} \cdot \mathbb{1}\{Y^t > 0\}\right) &\leq 1 - \frac{\lambda d^*}{2} \leq e^{-\lambda d^*/2} \\ &= e^{-\frac{(d^*)^2}{3c^2}} =: \beta < 1. \end{aligned}$$

Moreover, in order to apply Lemma 6, we need to bound

$$\mathbb{E}\left(e^{-\lambda(a-Y^{t+1})} \cdot \mathbb{1}\{Y^t = 0\} \mid Y_t\right) \leq e^{c\lambda} = e^{\frac{2d^*}{3c}} =: D > 1$$

using  $a := 0$  and the bounded step size. Altogether, from the lemma we get

$$\begin{aligned} \Pr(Y^t \geq b) &\leq \left(\beta^t + \frac{1-\beta^t}{1-\beta} D\right) e^{-\lambda b} \\ &\leq (1+D)e^{-\lambda b} \leq 2e^{\frac{2d^*}{3c}} e^{-\frac{2bd^*}{3c^2}} \\ &= 2e^{\frac{2d}{3c(1-p_0)} - \frac{2bd}{3c^2(1-p_0)}}, \end{aligned}$$

and the last expression is also a bound on  $\Pr(X^t \geq b)$  as it does not depend on  $t$ .  $\square$

### 3. AN ANYTIME ANALYSIS

In this section we give our anytime analysis, separately for the cases of  $r = 2$  and for large  $r$ . We will start in Section 3.1 with the classical case of  $r = 2$ , i.e., bit strings. In Section 3.2 we consider large  $r$ . We restrict ourselves to  $p = o(1/n)$ , i.e., in expectation less than one bit of the optimum is changed.

#### 3.1 The Case of $r = 2$

We fix  $r = 2$  and start by computing the expected change (drift) in the search point. We expect the (1+1) EA, starting from a random string, to make some progress towards the optimum until the number of incorrect bits is lower than the drift caused by the dynamically changing optimum.

More precisely, we consider the process  $X^t$  given by the current ONEMAX-value (i.e., the number of incorrect bits) and assume a current ONEMAX-value of  $X^t = i < n/2$ . We identify a forward drift

$$\Delta^-(i) := (i - X^{t+1}) \cdot \mathbb{1}\{X^{t+1} < i\}$$

caused by the selection mechanism of the (1+1) EA and a backward drift

$$\Delta^+(i) := (X^{t+1} - i) \cdot \mathbb{1}\{X^{t+1} > i\}$$

caused by the random movement of the optimum. The total drift  $\Delta^t = (X^t - X^{t+1})$  under  $X_t = i$  satisfies  $\Delta^t = \Delta^-(i) - \Delta^+(i)$ , and also  $E(\Delta^t) = E(\Delta^-(i)) - E(\Delta^+(i))$ . We bound the forward and backward drift. Progress is made when one of the incorrect bits flips and neither the rest of the bits nor the optimum flips, and can only be made by flipping incorrect bits of the current string or the optimum. The total expected number of flipping bits among  $i$  bits is  $i(p + 1/n)$ . We obtain

$$(1-p)^n \left(1 - \frac{1}{n}\right)^{n-i} \frac{i}{n} \leq E(\Delta^-(i)) \leq \frac{i}{n} + ip,$$

and, since  $p = o(1/n)$ ,

$$\frac{i}{e^2 n} \leq E(\Delta^-(i)) \leq \frac{2i}{n}.$$

Similarly, since the ONEMAX-value can only increase (move away from the optimum) by flipping bits of the optimum, we get

$$(1 - 1/n)^n (n-i)p(1-p)^i \leq E(\Delta^+(i)) \leq np,$$

implying, since  $i < n/2$

$$\frac{np}{4e^2} \leq E(\Delta^+(i)) \leq np.$$

We solve  $E(\Delta^t) = 0$  to find an  $i^*$  where we have a drift of zero, and get from the inequalities above that

$$\frac{1}{8e^2} n^2 p \leq i^* \leq e^2 n^2 p.$$

If  $i > e^2 n^2 p$ , there is certainly a drift towards the optimum; and if  $i < \frac{1}{8e} n^2 p$ , there is certainly a drift away from the optimum. In the region of  $i^* = \Theta(n^2 p)$  there is an equilibrium with zero drift, and we would expect the (1+1) EA to approach this region and not to move significantly away from it afterwards. This is made precise in the following theorem.

**THEOREM 8.** *Let  $r = 2$ ,  $p = o(1/n)$  and  $1/p = n^{O(1)}$ . Let  $(x^t, a^t)_{t \in \mathbb{N}}$  be the sequence of random variables denoting the pair of current search point and current optimum as given by running the (1+1) EA on dynamic ONEMAX. Then, for any  $t \geq 0$  and any  $\alpha = \omega(\ln n)$  there is  $b_t := n - \frac{tpn}{7}$  such that*

$$\Pr(d(x^t, a^t) \geq \max\{\alpha b_t, 2e^2 n^2 p + \alpha\}) \leq e^{-\Omega(\alpha)}.$$

Moreover, for all  $t \geq \alpha/p$ ,

$$\Pr(d(x^t, a^t) \geq 2e^2 n^2 p + \alpha) \leq e^{-\Omega(\alpha)}.$$

**PROOF.** Still,  $X^t := \text{ONEMAX}_{a^t}(x^t) = d(x^t, a^t)$ . We recall that there is a drift towards the target if  $X^t > i^*$ . More precisely, from the estimations presented before this theorem we obtain

$$\begin{aligned} E(i - X^{t+1} \mid X^t = i) &= E(\Delta^-(i)) - E(\Delta^+(i)) \\ &\geq \frac{i}{e^2 n} - np \geq np \end{aligned}$$

for  $i \geq 2e^2 n^2 p$ . In the following, we analyze the  $X^t$ -process using Lemma 6 with  $a := 2e^2 n^2 p$ ,  $b = \max\{\alpha b_t, a + \alpha\}$ ,  $x_{\min} := 0$  and the constant function  $h(x) := 1$ . From this we obtain the trivial potential function  $g(x) = x$ . To bound the moment-generating function of the drift if  $X^t > a$ , we use  $\Delta := g(X^t) - g(X^{t+1}) = X^t - X^{t+1}$  and argue

$$E(e^{-\lambda \Delta} \cdot \mathbb{1}\{X^t > a\} \mid X^t) \leq E(e^{-\lambda \Delta} \mid X^t; X^t > a).$$

In the following, we condition on  $X^t; X^t > a$  in all expectations unless stated otherwise but omit this for the sake of readability. Using the Taylor expansion of the exponential function, we get

$$E(e^{-\lambda \Delta}) \leq 1 - \lambda E(\Delta) + \sum_{k=2}^{\infty} \frac{\lambda^k E(|\Delta|^k)}{k!},$$

which for any  $\eta \geq \lambda$  is at most

$$\begin{aligned} 1 - \lambda E(\Delta) + \frac{\lambda^2}{\eta^2} \sum_{k=2}^{\infty} \frac{\eta^k E(|\Delta|^k)}{k!} \\ = 1 - \lambda E(\Delta) + \frac{\lambda^2}{\eta^2} \underbrace{\left( E(e^{\eta|\Delta|}) - \eta E(|\Delta|) - 1 \right)}_{\Psi}. \end{aligned}$$

The aim now is to bound the term in parentheses such that  $\Psi = O(\eta^2(pn + i/n))$ . To this end, note that  $|\Delta|$  is stochastically dominated by a sum of two independent random variables

$$Z \sim \text{Bin}(n, p) + \text{Bin}(i, 1/n)$$

since it is necessary to flip a bit of the optimum or a wrong bit of the current state to change the state. The sum of the two random variables overestimates the change of distance since the two types of flips might cancel each other.

The moment-generating function of the binomial distribution is well known and, for our parameters, given by

$$\begin{aligned} E(e^{\eta Z}) &= E(e^{\eta \text{Bin}(n, p)}) \cdot E(e^{\eta \text{Bin}(i, 1/n)}) \\ &= (pe^\eta + 1 - p)^n \cdot \left( \frac{1}{n} e^\eta + \left(1 - \frac{1}{n}\right) \right)^i \\ &\leq (p(1 + \eta + \eta^2) + 1 - p)^n \cdot \left( \frac{1}{n} (1 + \eta + \eta^2) + 1 - \frac{1}{n} \right)^i \\ &= ((\eta + \eta^2)p + 1)^n \left( \frac{\eta + \eta^2}{n} + 1 \right)^i \end{aligned}$$

for  $\eta \leq 1$  as  $e^x \leq 1 + x + x^2$  for  $x \leq 1$ . Using  $1 + x \leq e^x$ , we obtain from this

$$\begin{aligned} E(e^{\eta Z}) &\leq (e^{(\eta + \eta^2)p})^n (e^{\eta/n + \eta^2/n})^i \\ &= e^{(pn + i/n)\eta + (pn + i/n)\eta^2} \end{aligned}$$

Introducing  $q := pn + i/n \leq 2$ , we have

$$\begin{aligned} E(e^{\eta Z}) &\leq 1 + (q\eta + q\eta^2) + (q\eta + q\eta^2)^2 \\ &= 1 + q\eta + q\eta^2 + q^2\eta^2 + 2q^2\eta^3 + q^2\eta^4 \\ &\leq 1 + q\eta + q\eta^2 + 2q\eta^2 + 2q\eta^3 + 2q\eta^4 \\ &\leq 1 + q\eta + 7q\eta^2, \end{aligned}$$

where we assumed  $q\eta + q\eta^2 \leq 1$ , which holds for  $\eta \leq 1/4$ . Since  $E(Z) = q$ , we have established

$$\begin{aligned} \Psi &= \sum_{k=2}^{\infty} \frac{\eta^k E(|\Delta|^k)}{k!} \leq \sum_{k=2}^{\infty} \frac{\eta^k E(Z^k)}{k!} = E(e^{\eta Z}) - \eta E(Z) - 1 \\ &\leq (1 + q\eta + 7q\eta^2) - q\eta - 1 = 7q\eta^2 \end{aligned}$$

since  $\eta \leq 1$ .

Plugging this into the above bound on  $E(e^{-\lambda \Delta})$ , we get

$$E(e^{-\lambda \Delta}) \leq 1 - \lambda E(\Delta) + 7\lambda^2 q$$

for  $\lambda \leq \eta \leq 1/4$ . Since  $E(\Delta) \geq \frac{i}{e^2 n} - np$ , we choose  $\lambda = \frac{1}{21e^2}$  (and  $\eta = 1/4$ ) to get

$$\begin{aligned} E(e^{-\lambda\Delta}) &\leq 1 - \lambda \left( \frac{i}{e^2 n} - np \right) + 7\lambda \frac{1}{21e^2} q \\ &= 1 - \lambda \left( \frac{i}{e^2 n} - np \right) + \frac{\lambda}{3e^2} \left( \frac{i}{n} + pn \right) \\ &\leq 1 - \lambda \left( \frac{2i}{3e^2 n} - \frac{22pn}{21} \right) =: \beta \leq 1 - \frac{6pn}{441e^2} \end{aligned}$$

where the final inequality holds since  $i \geq a = 2e^2 n^2 p$ . In addition, we have then

$$\beta \leq e^{-6pn/(441e^2)}.$$

So far, we have bounded the moment-generating function of the drift by less than 1. We are left with a bound on

$$E\left(e^{-\lambda(a-X^{t+1})} \cdot \mathbb{1}\{X^t \leq a\}\right).$$

Noting that the exponent is positive for  $X^{t+1} \geq a \geq X^t$ , we bound the expression by

$$\begin{aligned} E\left(e^{\lambda|X^t - X^{t+1}|} \mid X^t \leq a\right) &\leq E\left(e^{\lambda Z} \mid X_t \leq a\right) \\ &\leq 1 + 8\lambda q \leq 1 + 16\lambda =: D \end{aligned}$$

using the estimations that bounded  $E(e^{\eta Z})$  further above. Applying Lemma 6, we get

$$\begin{aligned} \Pr(X^t \geq b) &\leq \beta^t \cdot e^{\lambda(X^0 - b)} + \frac{1}{1 - \beta} D e^{-\lambda\alpha} \\ &\leq e^{-\frac{6t pn}{441e^2}} e^{\frac{n-b}{21e^2}} + \frac{441e^2}{6pn} \left(1 + \frac{16}{21e^2}\right) e^{-\frac{\alpha}{21e^2}}. \end{aligned}$$

As  $b \geq b_t = n - \frac{t pn}{7}$ , the first term is  $e^{-b/(882e^2)}$ . Hence, if  $b \geq \alpha(n - \frac{t pn}{7})$ , it is  $e^{-\Omega(\alpha)}$ . Assuming  $\alpha = \omega(\ln n)$ , the second term is  $e^{-\Omega(\alpha)} = n^{-\omega(1)}$ , which makes the polynomial  $\frac{441e^2}{6pn}$  negligible. This proves the first statement from the theorem. The second one follows for  $t \geq n\alpha/(pn) = \alpha/p$  since then the first term is clearly  $e^{-\Omega(\alpha)}$ .

Altogether, for  $t$  large enough, we have

$$\Pr(X^t \geq b) \leq e^{-\Omega(\alpha)}. \quad \square$$

Theorem 8 shows that after a polynomial amount of time, the distance is very likely to be not by much above the equilibrium state. We can also show a somewhat symmetrical statement, showing that it is very likely to be not by much below the equilibrium state. This is proven in the following theorem.

**THEOREM 9.** *Let  $r = 2$ ,  $p = o(1/n)$  and  $1/p = n^{O(1)}$ . Let  $(x^t, a^t)_{t \in \mathbb{N}}$  be the sequence of random variables denoting the pair of current search point and current optimum as given by running the (1+1) EA on dynamic ONEMAX. Then, for any  $\alpha = \omega(\ln n)$  and all  $t \geq 0$*

$$\Pr(d(x^t, a^t) \leq n^2 p / (16e^2) - \alpha) \leq e^{-\Omega(n)} + e^{-\Omega(\alpha)}.$$

**PROOF.** We essentially follow the analysis from the proof of Theorem 8, but focus on a region close to the target where the negative drift is stronger than the positive one. To match the drift theorem, we flip the orientation of the space and let  $X^t = n - \text{ONEMAX}_{a^t}(x^t)$ . We recall that there is a drift away from the optimum if  $n - X^t < i^*$ . More precisely, from

the estimations presented at the beginning of this subsection we obtain

$$\begin{aligned} E((n-i) - X^{t+1} \mid X^t = n-i) &= E(\Delta^+(i)) - E(\Delta^-(i)) \\ &\geq \frac{np}{4e^2} - \frac{2i}{n} \geq \frac{np}{8e^2} \end{aligned}$$

for  $i \leq n^2 \frac{p}{16e^2}$ . This corresponds to  $X^t \geq n - n^2 \frac{p}{16e^2}$ . In the following, the aim is to analyze the  $X^t$ -process using Lemma 6 with  $a := n - n^2 p / (16e^2)$ ,  $b = a + \alpha$ ,  $x_{\min} := 0$  and the constant function  $h(x) := 1$ . From this we obtain the trivial potential function  $g(x) = x$ . We define  $\Delta$  and bound the moment-generating with the same procedure as in the proof of Theorem 8. Then (on  $X^t; X^t > a$ , which means  $i < n^2 p / (16e^2)$ )

$$E(e^{-\lambda\Delta}) \leq 1 - \lambda E(\Delta) + 7\lambda^2 q.$$

Since  $E(\Delta) \geq \frac{np}{8e^2}$ , we choose  $\lambda = \frac{1}{112e^2}$  to get

$$E(e^{-\lambda\Delta}) \leq 1 - \lambda \left( \frac{np}{17e^2} \right) = 1 - \frac{np}{1904e^4} =: \beta < 1.$$

The bound

$$E\left(e^{-\lambda(a-X^{t+1})} \cdot \mathbb{1}\{X^t \leq a\}\right) \leq 1 + 16\lambda =: D$$

is the same as in the proof of Theorem 8 since it only takes into account the worst-case distribution of  $\Delta$ . Altogether,

$$\begin{aligned} \Pr(X^t \geq b) &\leq e^{-\frac{t pn}{1904e^4}} e^{\frac{X^0 - b}{112e^2}} + \frac{1904e^4}{np} \left(1 + \frac{16}{112e^2}\right) e^{-\frac{\alpha}{112e^2}}. \end{aligned}$$

If  $X^0 \leq 2n/3$ , which happens with probability  $1 - 2^{-\Omega(n)}$  according to Chernoff bounds, the first term is  $e^{-\Omega(n)}$  since  $b = n - O(n^2 p) = n - o(n)$ . The second term is  $e^{-\Omega(\alpha)}$  by the same arguments as in the proof of Theorem 8. Hence, turning back to the original state space,

$$\Pr(d(x^t, a^t) \leq n^2 p / (16e^2) - \alpha) \leq e^{-\Omega(n)} + e^{-\Omega(\alpha)}$$

as suggested.  $\square$

## 3.2 The Case of Large $r$

In this section we consider large values of  $r$ . With the next theorem we show that, even for exponentially large  $r$ , the (1+1) EA maintains search points which differ from the optimum only by a constant in each dimension (in expectation)! This holds after an initial mixing phase, the length of which depends linearly on  $r$ .

**THEOREM 10.** *Let  $p = o(1/n)$  and let  $(x^t, a^t)_{t \in \mathbb{N}}$  be the sequence of random variables denoting the pair of current search point and current optimum as given by running the (1+1) EA on dynamic ONEMAX. Then there are  $k_0, k_1 > 1$  such that, for all  $t \geq k_0 r n^2$ ,*

$$\forall b \geq 4 : \Pr(\text{ONEMAX}_{a^t}(x^t) \geq bn) \leq n 2^{-k_1 b} \quad (1)$$

and

$$E(\text{ONEMAX}_{a^t}(x^t)) = O(n). \quad (2)$$

*In particular, this bound is independent of  $r$ . In addition, for all  $i \leq n$  and all  $t \geq k_0 r n^2$ ,*

$$\forall b \geq 4 : \Pr(d(a_i^t, x_i^t) \geq b) \leq 2^{-k_1 b}. \quad (3)$$

PROOF. We start by showing Equation (3). Fix a bit position  $i \leq n$ . We reason with drift on  $d^t = d(a_i^t, x_i^t)$  and show that it leads towards 0. Note that this value can change by at most two per iteration (one movement step of the algorithm, one of the optimum). Let some time  $t$  be given and suppose  $d^t \neq 0$ .

Let  $E$  be the event that the (1+1) EA keeps all bits other than  $i$  unchanged. We bound the expectation of moving in the *wrong* direction conditional on  $\overline{E}$  as

$$\mathbb{E}(d^{t+1} - d^t \mid \overline{E}) = o(1/n),$$

as the optimum might move away with probability  $o(1/n)$  but, if the new solution is accepted at all, it is more likely to be accepted if the  $i$  was changed in the right direction than when it was changed in the wrong direction. We further bound expectation of moving in the *right* direction conditional on  $E$  as

$$\mathbb{E}(d^t - d^{t+1} \mid E) = \Omega(1/n),$$

as we will not accept a worsening, but do accept an improvement, which will happen with probability  $\Omega(1/n)$ . Using that  $P(E)$  approaches  $1/e$  as  $n$  approaches infinity, we get a drift of  $\Omega(1/n)$  towards the optimum.

With the use of the additive drift theorem (Theorem 1), this shows that the first time we have  $d^t = 0$  is expected to be at most  $k_1 rn$  iterations, for some  $k_1$  large enough. Let  $T$  be the random variable denoting the smallest  $t$  with  $d^t = 0$ . Using concentration bounds for additive drift (Theorem 2), we get

$$\forall s \geq 2k_1 rn : \Pr(T \geq s) \leq \exp\left(-\frac{s}{64n^2}\right).$$

Let  $t_0 = 2k_1 rn^2$ . Thus, we do not have  $d^{t_0} = 0$  for the first time within the first  $t$  steps with probability  $2^{-\Omega(r)}$ .

We now set up to use Theorem 7 to derive bound for straying from the optimum after reaching it for the first time. In the notation of that lemma, our process has a drift of  $d = O(1/n)$ , a self-loop probability of  $p_0 = 1 - O(1/n)$  and a step size of at most  $c = 2$ . Thus, Lemma 7 gives some  $k$  such that, for all  $t$ ,

$$\forall b \geq 4 : \Pr(d(a_i^t, x_i^t) \geq b \mid t > T) \leq 2^{-kb}.$$

We have, for all  $t \geq t_0$ ,

$$\begin{aligned} \Pr(X^t \geq b) &\leq \Pr(d(a_i^t, x_i^t) \geq b \mid t > T) + \Pr(t \leq T) \\ &\leq 2^{-kb} + 2^{-\Omega(r)}. \end{aligned}$$

This gives the existence of a  $k_1$  as desired, which shows Equation (3). Equation (1) now follows from the union bound, while Equation (2) follows from linearity of expectation (and the trivial bound on the expectation of exponentially decaying random variables).  $\square$

#### 4. UPPER BOUND ON HITTING TIME OF TARGET

In this section, we re-prove the upper bound given by Droste (2003) in Theorem 12 and then extend it to the case of arbitrary  $r$  in Theorem 14. We start with the case of  $r = 2$ . Let  $X^t$ ,  $t \geq 0$ , be the Hamming distance of the current optimum string and the current search point of the (1+1) EA at time  $t$ . Hence, we get a process on  $\{0, \dots, n\}$  with target state 0. We lower bound the parameter  $p^-(i)$  and the “drift”  $\delta(i)$  in the sense of Theorem 4.

LEMMA 11. For  $i > 0$ ,  $\Pr(X^{t+1} = i - 1 \mid X^t = i) \geq (1 - p)^n \frac{i}{en}$ ,  $\Pr(X^{t+1} \geq i + 1 \mid X^t = i) \leq pn$  and  $\delta(i) \geq (1 - p)^n \frac{i}{en} - pn$ .

PROOF. The distance to the optimum decreases if the optimum does not move (probability  $(1 - p)^n$ ) and exactly one wrong bit flips (probability  $(1 - 1/n)^{n-1} \frac{i}{n} \geq \frac{i}{en}$ ), which proves the bound on  $\Pr(X^{t+1} = i - 1 \mid X^t = i)$ . The distance to the optimum can only increase if the dynamic component flips a bit. By a union bound, the probability is at most  $pn$ , which proves the bound on  $\Pr(X^{t+1} \geq i + 1 \mid X^t = i)$ .

To bound  $\mathbb{E}((X^{t+1} - i) \cdot \mathbf{1}\{X^{t+1} > i\} \mid X^t = i)$ , which appears in the definition of  $\delta(i)$ , we pessimistically assume that each change of the optimum string increases the distance to the current search point. The expected number of bits changed by the dynamic component equals  $pn$ , which altogether leads to the bound on  $\delta(i)$ .  $\square$

Hereinafter, we work with  $p^-(i) = (1 - p)^n \frac{i}{en}$  and  $p^+(i) = pn$ . We get the following polynomial upper bound in the case of bit strings.

THEOREM 12. Let  $r = 2$  and  $p \leq c \frac{\ln n}{n^2}$  for some constant  $c$ . Then the expected optimization time of the (1+1) EA on the dynamic ONEMAX is  $O(n^{4.8c+2} \ln^2 n)$ .

PROOF. By solving the equation  $\delta(i) = 0$  with the bounds from Lemma 11, we are allowed to set  $\kappa := \frac{pen^2}{(1-p)^n}$ . Using the assumption on  $p$ ,  $\kappa = (1 + o(1))ec \ln n \leq 3c \ln n$  for  $n$  large enough. Moreover, we have  $p^+(i) + p^-(i) \leq \frac{c \ln n}{n} + \frac{c \ln n}{n} \leq \frac{2c \ln n}{n}$  for  $i \leq \kappa$ . Then

$$\begin{aligned} \mathbb{E}(T \mid X^0) &\leq \left( \left( \sum_{k=3c \ln n+1}^n \frac{1}{\frac{k}{en} - (c \ln n)/n} \right) \prod_{k=1}^{3c \ln n} \frac{\frac{2c \ln n}{n}}{(1-p)^n \frac{k}{en}} \right) \\ &\quad + \left( \sum_{k=1}^{3c \ln n} \frac{1}{(1-p)^n \frac{k}{en}} \prod_{j=1}^{k-1} \frac{\frac{2c \ln n}{n}}{(1-p)^n \frac{j}{en}} \right) \\ &\leq \sum_{k=3c \ln n+1}^n \frac{1}{k/(10n \ln n)} \cdot P + 3c \ln n \cdot P, \end{aligned}$$

where  $P := en(1-p)^{-n} \prod_{k=2}^{3c \ln n} \frac{2ce \ln n}{k}$ . If  $n$  is not too small, we have

$$\mathbb{E}(T \mid X^0) \leq (10n \ln^2 n) + 3c \ln n)P \leq (11n \ln^2 n)P.$$

Now,

$$\begin{aligned} P &\leq en \left(1 - \frac{c \ln n}{n^2}\right)^{-n} \frac{(2ce \ln n)^{3c \ln n}}{(3c \ln n)!} \\ &\leq en(1 + o(1)) \left(\frac{2ce^2}{3c}\right)^{3c \ln n} \\ &\leq 2ene^{3c \ln(2e^2/3) \ln n} \leq 2en^{4.8c+1} \end{aligned}$$

using  $k! \geq (k/e)^k$ . Altogether,

$$\mathbb{E}(T \mid X^0) \leq 22en^{4.8c+2} \ln^2 n$$

for  $n$  large enough.  $\square$

For comparison, Droste (2003) proves the upper bound  $\mathbb{E}(T) = O(n^{4ce/\ln(2)+1} \ln n)$ , i.e., the exponent is almost



12c. Hence, state-of-the-art drift analysis yields more precise results, is more versatile and leads to cleaner and shorter proofs than the previous analysis by Droste (2003).

It is not too difficult to generalize Theorem 12 to arbitrary  $r$  if we replace the prerequisite on  $p$  by  $p \leq c \frac{\ln n}{rn^2}$ . Basically, a factor of  $r$  is lost if we work under the worst case assumption that wrong positions have distance  $r$  from the optimum, resulting in only  $\text{ONEMAX}_{a^t}(x^t)/r$  wrong positions. To increase the regime of polynomial hitting times, we have to prove this worst case to be unlikely. Fortunately, the anytime analysis from Theorem 10 can be used here to show that we lose a factor of at most  $O(\log n)$  regardless of  $r$ . To this end, we will use the following lemma, which immediately follows from Theorem 10.

**LEMMA 13.** *Let  $p = o(1/n)$ . Then there are constants  $a > 0, b > 1$  such that for any  $i \in [n], j \in \mathbb{N}$  and all  $t \geq rn^2 \ln n$  it holds  $\Pr(d(a_i^t, x_i^t) \geq j) \leq ab^{-j}$ .*

We now state the theorem concerned with polynomial hitting times for large  $r$ . To ease the statement, we only consider polynomial-sized  $r$ .

**THEOREM 14.** *Let  $r \leq n^k$  for some constant  $k$  and  $p \leq \frac{c \ln n}{\min\{r, \ln n\}n^2}$  for some sufficiently small constant  $c$  (possibly depending on  $k$ ). Then the expected optimization time of the (1+1) EA on the dynamic  $\text{ONEMAX}$  is polynomial in  $n$ .*

**PROOF.** Let  $t^* = rn^2 \ln n$ . Pessimistically ignoring the case that the optimum is hit in less than  $t^*$  steps, we apply Lemma 13. Choosing  $j^* = \ln(an^2 t^*) / (\ln b) = O(\log n)$ , we obtain that  $\Pr(d(a_i^t, x_i^t) \geq j^*) \leq 1/(t^* n^2)$  for any  $j$  and  $t \geq t^*$ . By a union bound, the probability that for all  $i \in [n]$  we have  $d(a_i^t, x_i^t) \leq j^*$  is  $1 - O(1/(t^* n))$ . From now on, we assume this to hold in a phase of length  $t^*$ , starting from time  $t^*$  up to time  $2t^* - 1$ . Again by a union bound, the probability that within  $t^*$  steps all positions have distance at most  $j^*$  from the target is still  $1 - O(1/n) - n^{-\omega(1)} = 1 - o(1)$ . If  $j^* \geq r$ , the assumption holds trivially, i. e., with probability 1.

Under our assumption, we conduct a drift analysis with respect to  $X^t := \text{ONEMAX}_{a^t}(x_t)$ . Similarly to Lemma 11,  $\Pr(X^{t+1} \geq i+1 \mid X^t = i) \leq pn$  since each changing position of the target increases the distance by at most 1 and also

$$\mathbb{E}((X^t - X^{t+1}) \cdot \mathbb{1}\{X^{t+1} > X^t\} \mid X^t = i) \leq pn.$$

If  $X^t = i$ , then there are at least  $\frac{i}{\min\{r, j^*\}}$  wrong positions, hence  $p^-(i) = \Pr(X^{t+1} = i-1 \mid X^t = i) \geq (1-p)^n \frac{i}{\min\{r, j^*\}en}$  and

$$\begin{aligned} \delta(i) &\geq (1-p)^n \frac{i}{\min\{r, j^*\}en} - pn \\ &\geq (1-p)^n \frac{i}{\min\{r, j^*\}en} - \frac{c \ln n}{\min\{r, \ln n\}n}. \end{aligned}$$

By our assumptions,  $j^* \leq kc_1 \ln n$  for some constant  $c_1$  (depending on  $a$  and  $b$ ) for large enough  $n$ . Using our assumption on  $p$ , we get that  $\delta(i) \geq 0$  for  $i \geq c_3 \ln n$ , where  $c_3$  is a constant such that  $c_3 = cc_1 k + c_2$  for another constant  $c_2$ . Hence, we work with  $\kappa := c_3 \ln n$ . If  $c_2$  is chosen appropriately, then we also have  $p^+(i) + p^-(i) \leq \frac{c_3 \ln n}{2n}$  for  $i \leq \kappa$ . Similarly as in the proof of Theorem 12, we get for small

enough  $c$  that

$$\mathbb{E}(T \mid X^0) \leq \sum_{\ell=c_3 \ln n}^{j^* n} \frac{1}{\ell / (2ekc_1 n \ln n)} \cdot P + c_3 \ln n \cdot P,$$

where

$$P := en \min\{r, j^*\} (1-p)^{-n} \prod_{\ell=2}^{c_3 \ln n} \frac{(c_3/2) \ln n}{\ell}.$$

If  $n$  is not too small,  $\mathbb{E}(T \mid X^0) \leq 3ekc_1 c_3 n (\ln^2 n) P$ , so we are left with an estimate for  $P$ . We get

$$P \leq O(n \ln^2 n) \left( \frac{c_3 e \ln n}{2c_3 \ln n} \right)^{c_3 \ln n} \leq n^{c_3}$$

for  $n$  large enough. Altogether,

$$\mathbb{E}(T \mid X^0) \leq n^{cc_1 k + c_2}.$$

If  $n$  is large enough and  $c$  is sufficiently small but still constant, then  $n^{cc_1 k + c_2} \leq t^*/2$ . Hence, by Markov's inequality, the probability that a phase of  $t^*$  steps is successful, i. e., the optimum is hit, is at least  $1/2$ ; still conditioning on maximum distance  $j^*$  for all positions. By the considerations from above, the unconditional probability of a successful phase is at least  $1/2 - o(1)$ . In case of a failure, we consider the subsequent phase of  $t^*$  steps. The expected number of phases is at most  $2 + o(1)$ , hence the overall expected first hitting time of the target is at most  $t^* + (2 + o(1))t^* = (3 + o(1))t^*$ , i. e., polynomial.  $\square$

We conjecture that the assumption on  $p$  in Theorem 14 can be replaced by  $p \leq \frac{c \ln n}{n^2}$ , i. e., that the same regime for polynomial first hitting time holds regardless of  $r$ . However, we cannot prove this at the moment since the processes describing the distance from the target for different positions are not independent.

## 5. LOWER BOUND ON HITTING TIME OF TARGET

When the mutation probability applied to the optimum is asymptotically larger than  $\log n/n^2$ , Droste (2003) shows that the first hitting time of (1+1) EA on  $\text{ONEMAX}$  for  $r = 2$  is polynomial only with super-polynomially small probability. We re-prove this result for any  $r \geq 2$  and  $p \leq 1/n$  using drift analysis.

**THEOREM 15.** *With  $p \in \omega(\log n/n^2)$  and  $p \leq 1/n$ , the first hitting time of the (1+1) EA on the dynamic  $\text{ONEMAX}$  for any  $r \geq 2$  is polynomial only with super-polynomially small probability.*

**PROOF.** To prove the result, we let  $X^t$  be the current solution of the EA and apply Theorem 3 using potential function  $g(X^t) = \sum_{i=1}^n [a_i^t \neq x_i^t]$ , i. e., the number of characters the individual and the optimum differ by at time  $t$ .

Consider the effects of mutating the optimum and the mutation/selection step separately. When  $g(X^t) \leq n/2$ , the effect of optimum mutation on  $g(X^t)$  non-matching characters is countered by the effect of mutation on  $g(X^t)$  matching characters, leaving  $n - 2g(X^t)$  matching characters which cause a drift away from the optimum:

$$\mathbb{E}(g(X^{t+1}) - g(X^t) \mid S^t) \geq (n - 2g(X^t)) \cdot p,$$

where  $S^t$  is the event that no mutation occurs during the mutation/selection step of iteration  $t$ .

For the mutation/selection step, the expected increase in the number of matching characters is at most the number of mutated non-matching characters. We can consider the mutation/selection occurring after the optimum is mutated, so the number of non-matching characters is in expectation increased by at most 1 for  $p \leq 1/n$ , leading to a combined drift of:

$$\mathbb{E}(g(X^{t+1}) - g(X^t)) \geq (n - 2g(X^t)) \cdot p - (g(X^t) + 1) \cdot 1/n.$$

Let  $p = \alpha(n) \cdot \log n/n^2 \leq 1/n$ , where  $\alpha(n) \in \omega(1) \leq n/\log n$ ; limiting  $g(X^t) < b = \alpha(n)^c \log n$ , where  $c < 1$  is a constant, reveals a drift away from the optimum:

$$\begin{aligned} & \mathbb{E}(g(X^{t+1}) - g(X^t) \mid g(X^t) < b) \\ & \geq (n - 2g(X^t)) \cdot p - (g(X^t) + 1) \cdot 1/n \\ & \geq \frac{\alpha(n) \log n}{n} - \frac{\alpha(n)^c \log n}{n} \left( \frac{2\alpha(n) \log n}{n} + 1 \right) - \frac{1}{n} \\ & \in \Omega\left(\frac{\alpha(n) \log n}{n}\right). \end{aligned}$$

With low  $p$  and  $g(X^t)$ , a large number of iterations might not alter the value of  $g(X^t)$  (causing a ‘‘self-loop’’). The probability  $p_0$  of an iteration resulting in a self-loop can be bounded by considering the probability that none of the characters of the optimum mutate, and none of the non-matching characters in the current individual mutate:

$$\begin{aligned} p_0 &= \Pr(g(X^{t+1}) = g(X^t) \mid X^t) \\ &\geq (1-p)^n \cdot (1-1/n)^{g(X^t)} \\ &\geq 1 - \frac{n\alpha(n) \log n}{n^2} - \frac{\alpha(n)^c \log n}{n} \\ (1-p_0) &\in O\left(\frac{\alpha(n) \log n}{n}\right). \end{aligned}$$

Thus, there exists a constant  $\varepsilon > 0$  satisfying the first requirement of Theorem 3. We then need to bound the probabilities of  $g(X^t)$  changing significantly in a single iteration. Throughout the following, let  $M$  be the event that a self-loop does not occur, i.e.,  $g(X^{t+1}) \neq g(X^t)$ .

Let  $c_1 = n - g(X^t) \leq n$  be the number of matching characters in the optimum and the current individual (for which  $d(a_i, x_i) = 0$ ). We note that mutating such a character in the optimum would increase  $g(X^{t+1})$  unless it is also mutated in the current individual; let  $C_1$  be the number of such mutations that occur:

$$\begin{aligned} \Pr(M) &\geq c_1 \cdot p(1-p)^{n-1} \cdot (1-1/n)^n \\ &\geq c_1 p/4e \\ \Pr(C_1 \geq j) &\leq \binom{c_1}{j} p^j \\ \Pr(C_1 \geq j \mid M) &\leq \frac{(c_1 p)^{j-1} 4e}{2^{j-1}} \leq \frac{8e}{2^j} \end{aligned}$$

for  $n \geq 2$ .

Let  $c_2 \leq g(X^t) \leq n$  be the number of characters for which  $d(a_i, x_i) \geq 2$ , i.e., those that would not transform into matching characters even if improved by mutation; notably,  $c_2 = 0$  if  $r = 2$ . When both, a matching character mutation, and a mutation improving such a character occurs

in the current individual,  $g(X^{t+1})$  increases without reducing fitness, allowing the mutated individual to be accepted; let  $C_2$  be the number of such mutations that occur:

$$\begin{aligned} \Pr(M) &\geq c_1 c_2 / (8en^2) \\ \Pr(C_2 \geq j) &\leq \binom{c_1}{j} \binom{c_2}{j} n^{-2j} \\ \Pr(C_2 \geq j \mid M) &\leq \frac{(c_1 c_2)^j 8en^2}{2^{2(j-1)} n^{2j} c_1 c_2} \\ &\leq \frac{(c_1 c_2)^{j-1} 32e}{n^{2(j-1)} 4^j} < \frac{32e}{4^j} \end{aligned}$$

as  $c_1 c_2 \leq n^2$ .

The increase in potential value is at most the sum of these two effects, and hence:

$$\Pr(g(X^{t+1}) - g(X^t) \geq j \mid M) \leq \Pr(C_1 + C_2 \geq j \mid M).$$

By Lemma 16, there exists a choice of  $r(\ell)$  and  $\delta$  that satisfies the second condition of Theorem 3 for jumps away from the optimum.

For jumps toward the optimum, let  $k_1$  be the number of characters for which  $d(a_i, x_i) = 1$ , i.e., those that can be corrected by a mutation in either the optimum or the current individual; let  $J_1$  be the number of such characters corrected in a given iteration. Proceeding as before,

$$\begin{aligned} \Pr(J_1 \geq j \mid M) &\leq \binom{2k_1}{j} \frac{1}{n^j} \frac{8en}{k_1} \\ &\leq \frac{(2k_1)^{j-1} 8e}{n^{j-1} 3^{j-2}} \leq \frac{36e}{1.5^j}, \end{aligned}$$

noting that if  $r > 2$ , a mutation in a specific direction is required while considering  $P(M)$ , while a mutation either direction is acceptable to upper-bound  $P(J_1 \geq j)$ .

Furthermore, let  $k_2$  be the number of characters for which  $d(a_i, x_i) = 2$ , i.e., those that can match if they are mutated appropriately in both the optimum and the current individual; let  $J_2$  be the number of such characters corrected in a given iteration. Similarly,

$$\begin{aligned} \Pr(J_2 \geq j \mid M) &\leq \binom{k_2}{j} \left(\frac{p}{4n}\right)^j \frac{16en}{k_2 p} \\ &\leq \frac{k_2^{j-1} 4e}{8^{j-1}} \left(\frac{p}{n}\right)^{j-1} \leq \frac{32e}{8^j} \end{aligned}$$

as  $k_2 \leq n \leq 1/p$ .

The reduction in potential value is at most the sum of these two effects, and so:

$$\Pr(g(X^{t+1}) - g(X^t) \leq -j \mid M) \leq \Pr(J_1 + J_2 \geq j \mid M).$$

Per Lemma 16, there exists a choice of  $r(\ell)$  and  $\delta$  that satisfies the second condition of Theorem 3 for jumps toward the optimum.

Thus, there exists a choice of  $r(\ell)$  and  $\delta$  that satisfies the second requirement of Theorem 3 both for jumps away from and jumps toward the optimum.

Finally, we note that the probability of a randomly initialized character matching the optimum is  $1/r \leq 1/2$ . Using Chernoff’s inequality, the probability that more than  $3n/4$  characters are initialized correctly is at most  $e^{-n/12}$ , and therefore  $g(X^0) \geq n/4 > b$  with high probability.

By applying Theorem 3 with  $g(X^t)$ ,  $b = \alpha(n)^c \log n$  and  $a = 0$ , and hence  $\ell = \omega(\log n)$ , we can conclude that if

$p \in \omega(\log n/n^2) \leq 1/n$ , the (1+1) EA finds the optimum in polynomial time with only super-polynomially small probability.  $\square$

While proving that large jumps are exponentially unlikely even after removing self-loops from the process, we used the following lemma to combine upper bounds for different kinds of jumps.

LEMMA 16. *Let  $J = J_1 + J_2$ ; if there exist constants  $r_1, r_2 \geq 1$ , and  $d_1, d_2 > 1$ , s.t. for some event  $E$ ,*

$$\Pr(J_1 \geq j \mid E) \leq r_1/d_1^j$$

$$\Pr(J_2 \geq j \mid E) \leq r_2/d_2^j$$

*and it holds that  $\Pr(J_2 \geq j_2 \mid E, J_1 \geq j_1) \leq \Pr(J_2 \geq j_2 \mid E)$ , then there also exist constants  $r, d > 1$ , s.t.*

$$\Pr(J \geq j \mid E) \leq r/d^j.$$

PROOF. Let  $r_* = \max(r_1, r_2)$  and  $d_* = \min(d_1, d_2)$ ; given the conditions, it is the case that:

$$\begin{aligned} \Pr(J \geq j \mid E) &\leq \sum_{i=0}^j \Pr(J_1 \geq i \mid E) \Pr(J_2 \geq j-i \mid E, J_1 \geq j_1) \\ &\leq \sum_{i=0}^j \Pr(J_1 \geq i \mid E) \Pr(J_2 \geq j-i \mid E) \\ &\leq r_*^2 d_*^{-j} (j+1). \end{aligned}$$

We note that  $(j+1)d_*^{-j} \leq \sqrt{d_*}^{-j}$  for  $j \geq 16/(\ln^2 d_*)$ . It is possible to pick a constant  $c = \sqrt{d_*}^{16/(\ln^2 d_*)}$ , ensuring that  $c/\sqrt{d_*}^j \geq 1$  for  $j \leq 16/(\ln^2 d_*)$ , which proves the lemma with  $r = r_*c$  and  $d = \sqrt{d_*}$ .  $\square$

## 6. CONCLUSION

In this paper we revisited the setting of dynamic ONE-MAX as introduced by Droste (2003), where the optimum moves by flipping the bit of each position with some fixed probability  $p$ .

We showed that his results, both the upper and the lower bound, extend to versions of dynamic ONEMAX where each dimension has  $r$  different possible values. By using modern drift analysis, the proof is shorter and more elegant.

Furthermore, we made an analysis of how far from the optimum the (1+1) EA strays after getting close for the first time. For the case of bit strings, this value is concentrated around  $\Theta(pn^2)$  (for  $p = o(1/n)$ ), which shows that the optimum is very elusive unless  $p$  is small. On the other hand, we showed that only the *dimension*, and not the *size*, of the search space has an impact on the ability of the (1+1) EA to track good solutions. We did this by considering search spaces with  $r$  possible values in each dimension, and saw that  $r$  does not influence the resulting bounds, i. e., the distance is bounded by a constant in expectation in each dimension if  $r$  is large (see Theorem 10).

We believe that the methods we used, especially the statements about the occupation probabilities as given in Lemma 13 of Lissovoi and Witt (2015) or in our Theorem 7, will be beneficial in many more settings, especially those aiming at an anytime analysis for dynamic problems.

## Acknowledgments

The authors would like to thank Benjamin and Carola Doerr for interesting discussions on the topic of this paper. The anonymous reviewers of the pre-conference reviews made many valuable suggestions which improved the paper. Furthermore, Golnaz Badkobeh, Per Kristian Lehre and Dirk Sudholt gave many useful suggestions as part of the FOGA post-conference reviews. Financial support from the Danish Council for Independent Research (grant no. DFF-4002-00542) is gratefully acknowledged.

## References

- Auger, Anne and Doerr, Benjamin (2011). *Theory of Randomized Search Heuristics – Foundations and Recent Developments*. World Scientific Publishing.
- Doerr, Benjamin, Johannsen, Daniel, and Schmidt, Martin (2011). Runtime analysis of the (1+1) evolutionary algorithm on strings over finite alphabets. In *Proc. of FOGA'11*, 119–126. ACM Press.
- Doerr, Benjamin and Pohl, Sebastian (2012). Run-time analysis of the (1+1) evolutionary algorithm optimizing linear functions over a finite alphabet. In *Proc. of GECCO'12*, 1317–1324. ACM Press.
- Droste, Stefan (2002). Analysis of the (1+1) EA for a dynamically changing OneMax-variant. In *Proc. of CEC'02*, 55–60. IEEE Press.
- Droste, Stefan (2003). Analysis of the (1+1) EA for a dynamically bitwise changing OneMax. In *Proc. of GECCO'03*, 909–921. Springer.
- Hajek, Bruce (1982). Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability*, **13**, 502–525.
- He, Jun and Yao, Xin (2001). Drift analysis and average time complexity of evolutionary algorithms. *Artificial Intelligence*, **127**, 57–85.
- Jansen, Thomas (2013). *Analyzing Evolutionary Algorithms – The Computer Science Perspective*. Natural Computing Series. Springer.
- Jansen, Thomas and Schellbach, Ulf (2005). Theoretical analysis of a mutation-based evolutionary algorithm for a tracking problem in the lattice. In *Proc. of GECCO'05*, 841–848. ACM Press.
- Jansen, Thomas and Zarges, Christine (2014). Evolutionary algorithms and artificial immune systems on a bi-stable dynamic optimisation problem. In *Proc. of GECCO'14*, 975–982. ACM Press.
- Jin, Yaochu and Branke, Jürgen (2005). Evolutionary optimization in uncertain environments—A survey. *IEEE Transactions on Evolutionary Computation*, **9**, 303–317.
- Kötzing, Timo (2014). Concentration of first hitting times under additive drift. In *Proc. of GECCO'14*, 1391–1398. ACM Press.
- Kötzing, Timo and Molter, Hendrik (2012). ACO beats EA on a dynamic pseudo-boolean function. In *Proc. of PPSN'12*, 113–122. Springer.

- Lehre, Per Kristian and Witt, Carsten (2014). Concentrated hitting times of randomized search heuristics with variable drift. In *Proc. of ISAAC'14*, 686–697. Springer. Extended version at <http://arxiv.org/abs/1307.2559>.
- Lissovoi, Andrei and Witt, Carsten (2015). MMAS versus population-based EA on a family of dynamic fitness functions. *Algorithmica*. In press, final version online at <http://dx.doi.org/10.1007/s00453-015-9975-z>.
- Neumann, Frank and Witt, Carsten (2010). *Bioinspired Computation in Combinatorial Optimization – Algorithms and Their Computational Complexity*. Natural Computing Series. Springer.
- Oliveto, Pietro Simone and Witt, Carsten (2011). Simplified drift analysis for proving lower bounds in evolutionary computation. *Algorithmica*, **59**, 369–386.
- Oliveto, Pietro Simone and Witt, Carsten (2012). Erratum: Simplified drift analysis for proving lower bounds in evolutionary computation. Tech. rep., <http://arxiv.org/abs/1211.7184>.
- Oliveto, Pietro Simone and Zarges, Christine (2013). Analysis of diversity mechanisms for optimisation in dynamic environments with low frequencies of change. In *Proc. of GECCO'13*, 837–844. ACM Press.
- Rowe, Jonathan E. and Sudholt, Dirk (2014). The choice of the offspring population size in the  $(1, \lambda)$  evolutionary algorithm. *Theoretical Computer Science*, **545**, 20–38.