



# Exponential upper bounds for the runtime of randomized search heuristics

Benjamin Doerr

## ► To cite this version:

Benjamin Doerr. Exponential upper bounds for the runtime of randomized search heuristics. Theoretical Computer Science, Elsevier, 2021, 851, pp.24-38. 10.1016/j.tcs.2020.09.032 . hal-03354713

**HAL Id: hal-03354713**

**<https://hal.archives-ouvertes.fr/hal-03354713>**

Submitted on 25 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Exponential Upper Bounds for the Runtime of Randomized Search Heuristics\*

Benjamin Doerr  
 Laboratoire d'Informatique (LIX)  
 CNRS, École Polytechnique  
 Institut Polytechnique de Paris  
 Palaiseau  
 France

September 7, 2020

## Abstract

We argue that proven exponential upper bounds on runtimes, an established area in classic algorithms, are interesting also in heuristic search and we prove several such results. We show that any of the algorithms randomized local search, Metropolis algorithm, simulated annealing, and (1+1) evolutionary algorithm can optimize any pseudo-Boolean weakly monotonic function under a large set of noise assumptions in a runtime that is at most exponential in the problem dimension  $n$ . This drastically extends a previous such result, limited to the (1+1) EA, the LeadingOnes function, and one-bit or bit-wise prior noise with noise probability at most  $1/2$ , and at the same time simplifies its proof. With the same general argument, among others, we also derive a sub-exponential upper bound for the runtime of the  $(1, \lambda)$  evolutionary algorithm on the OneMax problem when the offspring population size  $\lambda$  is logarithmic, but below the efficiency threshold. To show that our approach can also deal with non-trivial parent population sizes, we prove an exponential upper bound for the runtime of the mutation-based version of the simple genetic algorithm on the OneMax benchmark, matching a known exponential lower bound.

---

\*Extended version of a paper appearing in the proceedings of PPSN 2020 [Doe20b]. This version contains all proofs and other details that had to be omitted in the conference version for reasons of space (roughly an additional 50% text over the conference version) as well as the new Section 5 on the simple genetic algorithm.

# 1 Introduction

The mathematical analysis of runtimes of randomized search heuristics is an established field of the general area of heuristic search [NW10, AD11, Jan13, DN20]. The vast majority of the results in this area show that a certain algorithm can solve (or approximately solve) a certain problem within some polynomial runtime (polynomial upper bound on the runtime) or show that this is not possible by giving a super-polynomial, often exponential, lower bound on the runtime.

As a rare exception to this rule, in his extensive analysis of how the  $(1+1)$  evolutionary algorithm  $((1+1) \text{ EA})^1$  optimizes the `LEADINGONES` benchmark in the presence of prior noise, Sudholt [Sud20, Theorem 6] showed that for one-bit or bit-wise noise with noise probability at most  $\frac{1}{2}$ , the  $(1+1)$  EA finds the optimum of `LEADINGONES` in time at most  $2^{O(n)}$ . While clearly a very natural result – everyone would agree that also with such noise the unimodal `LEADINGONES` problem should not become harder than the needle-in-the-haystack problem – the technical, long, and problem-specific proof of this result, despite following the intuitive argument just laid out, suggests that such analyses can be harder than one would expect.

In this work, we will argue that such exponential upper bounds are interesting beyond completing a runtime picture of a given problem. We then show that with a different analysis method such uncommon runtime questions can be analyzed relatively easily. As one out of several results, we drastically extend the result in [Sud20] and show that an exponential runtime guarantee holds for

- any of the algorithms randomized local search, Metropolis algorithm, simulated annealing, and  $(1+1)$  EA,
- when optimizing any weakly monotonic objective function, e.g., `ONEMAX`, linear functions, monotone polynomials, `LEADINGONES`, plateau functions, and the needle problem,
- in the presence of all common forms of prior and posterior noise with a noise probability of at most  $1 - \varepsilon$ ,  $\varepsilon > 0$  a constant.

## 1.1 Exponential Runtime Analysis

The area of mathematical runtime analysis, established as a recognized sub-field of the theory of evolutionary algorithms by Ingo Wegener and his research group, seeks to understand the working principles of evolutionary

---

<sup>1</sup>See Section 2 for details on all technical terms used in this introduction.

computation via rigorously proven results on the performance of evolutionary algorithms and other search heuristics in a similar spirit as done in classic algorithms analysis for much longer time.

Adopting the view of classic algorithmics that runtimes polynomial in the problem size are efficient and larger runtimes are inefficient, the vast majority of the results in this field prove polynomial upper bounds or super-polynomial lower bounds. For two reasons, we feel that also super-polynomial and even exponential runtime guarantees are desirable in the theory of evolutionary algorithms.

Our first set of arguments is identical to the arguments made in the classic algorithms field, which led to a shift in paradigms and established the field of exact exponential algorithms [FK10, FK13]. These arguments are that (i) for many important problems nothing better than exponential time algorithms are known, so one cannot just ignore these problems in algorithms research, (ii) with the increase of computational power, also exponential time algorithms can be used for problems of moderate (and interesting) size, and (iii) that the existing research on exponential-time algorithms has produced many algorithms that, while still exponential time, are much faster than naïve exponential-time approaches like exhaustive search.

Our second line of argument is that exponential time algorithms are of additional interest in evolutionary computation for the following reasons.

(i) *To increase our understanding of the working principles of evolutionary algorithms.* For a large number of algorithmic problems in our field an exponential lower bound has been proven, but for essentially none of these problems an upper bound better than the trivial  $n^{O(n)}$  bound exists. It is clear that matching upper and lower bounds tell us most, not only about the runtimes, but also about the working principles of EAs. Tight bounds naturally have to grasp the true way the EA progresses better. For example, the general  $n^{O(n)}$  upper bound for all algorithms using standard bit mutation is based on the simple argument that the optimum can be generated from any search point with probability at least  $n^{-n}$ . Besides being very pessimistic, this argument does not tell us a lot on how really the EA optimizes the problem at hand (except for the very particular case that the EA is stuck in a local optimum in Hamming distance  $n$  to the global optimum). In contrast, as a positive example, the matching  $(1 \pm o(1))en \ln n$  upper [Müh92] and lower [GKS99, DFW11] bound for the runtime of the  $(1 + 1)$  EA on ONEMAX together with their proofs shows that for this optimization process, the effect of mutations flipping more than one bit has no influence on the runtime apart from lower order terms. In a broader sense, this insight suggests that flipping larger number of bits is mainly useful to

leave local optima, but not to make fast progress along easy slopes of the fitness landscape.

(ii) *Because understanding runtimes in the exponential and super-exponential regime is important for the application of EAs.* Many classic evolutionary algorithms can easily have a super-exponential runtime. For example, Witt [Wit05] has shown that the simple  $(1 + 1)$  EA has an expected runtime<sup>2</sup> of  $n^{\Theta(n)}$  on the minimum makespan scheduling problem. Hence knowing that an evolutionary algorithm “only” has an exponential runtime can be interesting.

We note that for problems with exponential-size search spaces (such as the search space  $\{0, 1\}^n$  regarded exclusively in this work) blind random search and exhaustive search are exponential-time alternatives. For that reason, in addition to knowing that an EA has an exponential runtime guarantee (that is, a runtime of at most  $C^n$  for some constant  $C > 1$ ), it would be very desirable to also have a good estimate for the base of the exponential function, that is, the constant  $C$ . Unfortunately, at this moment where we just start reducing the trivial  $n^{O(n)}$  upper bound to exponential upper bounds, we are not yet in the position to optimize the constants in the exponent. We are optimistic though (and give some indication for this in Section 4.1.1) that our methods can be fine-tuned to give interesting values for the base of the exponential function as well. We recall that such an incremental progress is not untypical for the mathematical runtime analysis of EAs – in the regime of polynomial bounds, subject to intensive research since the 1990s, the leading constants for elementary problems such as LEADINGONES and linear functions were only determined from 2010 on [BDN10, Sud13, Wit13].

With this motivation in mind and spurred by the observation that exponential upper bounds are not trivial to obtain, we conduct in this work the first investigation focused on the problem of proving exponential upper bounds for runtimes of EAs.

## 1.2 State of the Art

We are not aware of many previous works on exponential or super-exponential upper bounds on runtimes of EAs. In the, to the best of our knowledge, first work proving an exponential upper bound for the runtime of an EA, Droste, Jansen, and Wegener [DJW98, Theorem 9] show that the  $(1 + 1)$  EA optimizes the NEEDLE function (called peak function there) in expected time at

---

<sup>2</sup>As common both in classic algorithms and in our field, by runtime we mean the worst-case runtime taken over all input instances.

most  $(2\pi)^{-1}n^{1/2}\exp(2n)$ . Only a year later, Garnier, Kallel, and Schoenauer [GKS99, Proposition 3.1] in a remarkably precise analysis showed that the expected runtime of the  $(1+1)$  EA on the NEEDLE function is  $(1 \pm o(1))(1 - \frac{1}{e})^{-1}2^n$ .

A general upper bound of  $n^n$  for the expected runtime of the  $(1+1)$  EA on any pseudo-Boolean function was given in [DJW02, Theorem 6] by arguing that each offspring with probability at least  $n^{-n}$  is the optimum. This bound is tight as witnessed, among others, by the trap function [DJW02, Theorem 8] and the minimum makespan scheduling problem [Wit05]. Similar arguments as used in [DJW02] to prove the  $n^n$  upper bound showed an upper bound of  $4^n \log_2 n$  for the  $(1+1)$  EA using the mutation rates  $2^i/n$ ,  $i = 0, 1, \dots, 2^{\lceil \log_2 n \rceil - 1}$ , in a cyclic fashion [JW06, Theorem 3] and an upper bound of  $O(n^\beta 2^n)$  for the fast  $(1+1)$  EA with (constant) power-law exponent  $\beta > 1$  [DLMN17, Theorem 5.3]. With more complex arguments, the self-adjusting SD- $(1+1)$  EA was shown to optimize the trap function in time  $O((2.34)^n \log n)$  [RW20, Corollary 2]. It is easy to see that when compromising with the base of the exponential function, very similar arguments show an exponential upper bound for the runtime on any function  $f : \{0, 1\}^n \rightarrow \mathbb{R}$ .

There are a few analyses for parameterized problems showing bounds that can become exponential or worse when the problem parameter is chosen in an extreme manner. Here the  $\Theta(n^k)$  runtime bound for the  $(1+1)$  EA optimizing jump functions with jump size  $k \geq 2$  [DJW02, Theorem 25] is the best known example. More interesting results have been shown in the context of parameterized complexity [NS20], but again these results have been derived with small parameter values in mind and thus are most interesting for this case.

In contrast to these sporadic upper bounds, there is a large number of (near-)exponential lower bounds, e.g., for a broad class of non-elitist algorithms with too low selection pressure [Leh10], for some algorithms using fitness-proportionate selection [HJKN08, NOW09, Leh11, Doe20c], for the simple genetic algorithm with an only moderately large population size [OW15], for the optimization of strictly monotonic functions [DJS<sup>+</sup>13, Len18, LZ19], and for various problems in noisy optimization [GK16, QBJT19, Sud20].

Apart from a single exception, for none of these lower bounds it is known how tight they are, not even a result ruling out a runtime of  $n^{\Theta(n)}$ . The exceptional exponential upper bound shown in [Sud20, Theorem 6] reads as follows. Consider optimizing the LEADINGONES benchmark function defined on bit strings of length  $n$  via the  $(1+1)$  EA. Assume that in each iteration, the fitness evaluation of both parent and offspring is subject to stochastically independent prior noise of one of the following two types. (i) With probability

$p \leq \frac{1}{2}$ , not the true fitness is returned, but the fitness of a random Hamming neighbor. (ii) With probability  $p' \in [0, 1]$ , the search point to be evaluated is disturbed by flipping each bit independently with some probability  $q \leq \frac{1}{2}$  and the fitness of this disturbed search point is returned, with probability  $1 - p'$ , the fitness of the original search point is returned; here we assume that  $p' \min\{1, qn\} \leq \frac{1}{2}$ . Then the expected optimization time, that is, the number of iterations until the optimum is sampled, is at most exponential in  $n$ .

With a noise probability of at most  $\frac{1}{2}$  and a weakly monotonic, that is, weakly preferring 1-bits over 0-bits, fitness function one would think that this optimization process in some suitable sense is at least as good as the corresponding process on the NEEDLE function, where absolutely no fitness signal guides the search. This is indeed true, as the proof in [Sud20] shows. Surprisingly, as this proof also shows, it is highly non-trivial to make this intuitive argument mathematically rigorous. The proof in [Sud20] is around four pages long (including the one of the preliminary lemma) and builds on a technical estimate of the mixing time, which heavily exploits characteristics of the LEADINGONES objective function. Consequently, this proof does not easily generalize to other easy benchmark functions such as ONEMAX or linear functions.

### 1.3 Our Results

Observing that the natural approach taken in [Sud20] is unexpectedly difficult, we develop an alternative approach to proving exponential upper bounds. It builds on the following elementary observation. In the, slightly extremal, situation that we aim at an exponential upper bound, we can wait for an exponentially unlikely “lucky” way to generate the optimum. Being at most exponentially unlikely, that is, having a probability of  $p = 2^{-O(n)}$ , it takes  $2^{O(n)}$  attempts until we succeed. Hence if each attempt takes at most exponential time  $T_0$  (all our attempts will only take polynomial time), we obtain an exponential upper bound on the expected runtime, and moreover, the distributional bound that the runtime is stochastically dominated by  $T_0$  times a geometric distribution with success rate  $p$ . This general argument (without the elementary rephrasing in the stochastic domination language) was already used in the proof of the  $\text{poly}(n)e^{2n}$  upper bound on the expected runtime of the  $(1 + 1)$  EA on the NEEDLE function by Droste, Jansen, and Wegener [DJW98] more than twenty years ago. It is apparently not very well known in the community, most likely due to the fact that only one year later, Garnier, Kallel, and Schoenauer [GKS99] presented a much tighter analysis of this runtime via different methods. We are not aware of any other use of



this argument, which might explain why it was overlooked in [Sud20] (and we give in that we also learned it only from a review on an earlier version of this work).

How powerful this simple approach is, naturally, depends on how easy it is to exhibit lucky ways to find the optimum fast. As we shall demonstrate, this is in fact often easy. For example (see Theorem 5 for the details), it suffices that in each iteration the probability to move to a Hamming neighbor one step closer to the optimum is  $\Omega(n^{-1})$ . From this, we can show that from any starting point, the probability to reach the optimum in at most  $n$  iterations is at least  $2^{-O(n)}$ . As argued in the preceding paragraph, this yields an expected runtime of  $n2^{O(n)} = 2^{O(n)}$ . This argument, without noise and for the  $(1 + 1)$  EA only, was also used in the NEEDLE analysis in [DJW98].

Together with some elementary computations, this approach suffices to show that a large number of  $(1 + 1)$ -type algorithms in the presence of a large variety of types of noise with noise probability at most  $1 - \varepsilon$ ,  $\varepsilon > 0$  a constant, optimize any weakly monotonic function (including, e.g., ONEMAX, LEADINGONES, and the needle function) in at most exponential time (Theorem 6).

With similar arguments, we extend this result to the  $(1 + 1)$  EA optimizing jump functions with jump size at most  $\frac{n}{\ln n}$  in Section 4.2. For the particular noise model of bit-wise noise with rate  $q$  (search points are disturbed by flipping each bit independently with probability  $q$  before the fitness evaluation), we show in Section 4.3 that any of the above algorithms even in the presence of extreme noise with  $q$  as high as  $1 - \varepsilon$ ,  $\varepsilon > 0$  a constant, can optimize the ONEMAX benchmark in exponential time. For the exponential lower bound on the runtime of the  $(1 + 1)$  EA with fitness-proportionate selection on linear functions in [HJKN08], we easily derive a matching upper bound in Section 4.4. As an example showing that our approach can also yield sub-exponential upper bounds, we prove in Section 4.5 that the  $(1, \lambda)$  EA with  $\lambda \geq (1 - \varepsilon) \log_{\frac{e}{e-1}}(n)$ , and thus potentially below the threshold for polynomial time, optimizes ONEMAX in time  $\exp(O(n^\varepsilon))$ . To ease the presentation, we only regard single-trajectory algorithms in our general result (Theorem 5). To show that similar arguments can also be used to analyze algorithms with non-trivial parent populations, we show in Section 5 that the mutation-based variant of the simple genetic algorithm finds the optimum of ONEMAX in exponential time. This matches the known exponential lower bound.



## 2 Preliminaries

In this section, we describe the algorithms, the noise models, and the benchmark problems considered in this work. Almost all of this is standard, so we aim at brevity and refer to other works for more detail. We only consider optimization problems defined on the search space  $\{0, 1\}^n$  of bit strings of length  $n$ ; we thus also formulate all algorithms only for this setting. We have no doubt, though, that our methods can also be applied to other discrete optimization problems.

We only use the standard notation of this field. This includes writing  $[a..b] := \{z \in \mathbb{Z} \mid a \leq z \leq b\}$  and denoting by  $H(x, y) := |\{i \in [1..n] \mid x_i \neq y_i\}|$  the *Hamming distance* of two bit strings  $x, y \in \{0, 1\}^n$ . We denote by  $\text{Geom}(p)$  the *geometric distribution* with success rate  $p \in (0, 1]$ . Hence if a random variable  $X$  is geometrically distributed with parameter  $p$ , we write  $X \sim \text{Geom}(p)$  to denote this, then  $\Pr[X = k] = (1 - p)^{k-1}p$  for all  $k \in \mathbb{Z}_{\geq 1}$ . For two random variables  $X, Y$  we write  $X \preceq Y$  to denote that  $Y$  *stochastically dominates*  $X$ , that is, that  $\Pr[X \geq \lambda] \leq \Pr[Y \geq \lambda]$  for all  $\lambda \in \mathbb{R}$ .

### 2.1 Algorithms

We call a randomized search heuristic *single-trajectory* search algorithm if it is an iterative heuristic which starts with a single solution  $x^{(0)}$  and in each iteration  $t = 1, 2, \dots$  updates this solution to a solution  $x^{(t)}$ . We do not make any assumption on how this update is computed. In particular, the next solution may be computed from more than one solution candidate sampled in this iteration. We do, in principle, allow that information other than the search point  $x^{(t-1)}$  is taken into iteration  $t$ . However, in our main technical result we require that the key condition can be checked only from the search point  $x^{(t-1)}$ . Formally speaking, this means that for any possible history of the search process up to this point, when conditioning on this history, the key condition is true. To ease the language, we shall write “regardless of what happened in the first  $t - 1$  iterations” to express this conditioning. Naturally, for algorithms that can be described via a Markov chain a conditioning on the history is not necessary.

Examples for single-trajectory algorithms are (randomized) local search, the Metropolis algorithm, simulated annealing, and evolutionary algorithms working with a parent population of size one, such as the  $(1 + 1)$  EA,  $(1 + \lambda)$  EA,  $(1, \lambda)$  EA,  $(1 + (\lambda, \lambda))$  GA [DDE15], and SSWM algorithm [PHST17].

We call a single-trajectory algorithm  $(1+1)$ -*type algorithm* if in each iteration  $t$  it generates one solution  $y$  and takes as next parent individual  $x^{(t)}$  either  $y$  or  $x^{(t-1)}$ . Among the above examples, (randomized) local search, the Metropolis algorithm, simulated annealing, and the  $(1+1)$  EA are  $(1+1)$ -type algorithms.

We briefly describe the  $(1+1)$ -type algorithms regarded in this work. Usually all these algorithms start with a solution  $x^{(0)}$  chosen uniformly at random from  $\{0,1\}^n$ . Since none of our results relies on this assumption, we allow any kind of initialization, that is, also without explicit mention all our results hold for any initial search point  $x^{(0)}$ . We formulate all algorithms for the maximization of a given objective function  $f : \{0,1\}^n \rightarrow \mathbb{R}$ , which as common in evolutionary computation we call *fitness*.

The **Randomized local search (RLS)** heuristic in each iteration  $t$  generates a new solution  $y$  by copying  $x^{(t-1)}$  and then flipping a single bit, chosen uniformly at random, in  $y$ . If  $f(y) \geq f(x^{(t-1)})$ , then  $x^{(t)} := y$ , else  $x^{(t)} := x^{(t-1)}$ .

The **Metropolis algorithm** is identical to RLS apart from the selection step. If  $f(y) \geq f(x^{(t-1)})$ , then again  $x^{(t)} := y$ . However, if  $f(y) < f(x^{(t-1)})$ , then still with probability

$$\exp\left(-\frac{f(x^{(t-1)}) - f(y)}{T}\right)$$

the algorithm accepts the new solution, that is,  $x^{(t)} := y$ . Only with probability  $1 - \exp(-(f(x^{(t-1)}) - f(y))/T)$  the new solution is discarded, that is,  $x^{(t)} := x^{(t-1)}$ . The *temperature*  $T > 0$  is an algorithm parameter that defines the selection pressure. The Metropolis algorithm with a time-dependent (usually decreasing) temperature  $T(t)$  is called **simulated annealing**.

The **(1+1) EA** is identical to randomized local search except that now the offspring  $y$  is generated by a global mutation operator. In this work, we only consider the most classic choice of mutation, which is standard bit mutation. Here,  $y$  is taken as a copy of  $x^{(t-1)}$  and then each bit of  $y$  is flipped independently with probability  $p$ , the *mutation rate*. The most common choice, and our choice in this work for all static mutation rates, is  $p = \frac{1}{n}$ . When the mutation rate is chosen randomly according to a power-law with exponent  $\beta > 1$ , this algorithm is called **fast (1+1) EA** [DLMN17]. More precisely, here a random  $\alpha \in [1, n/2]$  is chosen such that  $\Pr[\alpha = k] = k^{-\beta} / \sum_{i=1}^{n/2} i^{-\beta}$  and then standard bit mutation is performed with  $p = \frac{\alpha}{n}$ .

From the description of the algorithms, the following property is immediate. In simple words, it says that the algorithms go to any Hamming neighbor that is not worse than the parent with probability  $\Omega(\frac{1}{n})$ .

**Proposition 1.** *For any algorithm  $A$  described above (and any choice of the parameters not fixed above, that is, temperature  $T$ , time-dependent temperature  $T(t)$ , or power-law exponent  $\beta > 1$ ), there is a constant  $c_A > 0$  such that the following holds.*

*For any iteration  $t$  and any  $z$  with  $H(z, x^{(t-1)}) = 1$ , and regardless of what happened in the previous iterations, the offspring  $y$  generated by  $A$  in iteration  $t$  satisfies  $\Pr[y = z] \geq \frac{c_A}{n}$ . If  $f(y) \geq f(x^{(t-1)})$ , then also  $\Pr[x^{(t)} = z] \geq \frac{c_A}{n}$ .*

## 2.2 Noise Models

Optimization in the presence of noise, that is, stochastically disturbed access to the problem instance, is an important topic in the optimization of real-world problems. The most common form are noisy objective functions, that is, that the optimization algorithm does not always learn the correct quality (fitness) of a search point. Randomized search heuristics are generally believed to be reasonably robust to noise, see, e.g., [JB05, BDGG09], which differs from problem-specific deterministic algorithms, which often cannot cope with any noise. Some theoretical work exists on how randomized search heuristics cope with noise, started by the seminal paper of Droste [Dro04] and, quite some time later, continued with, among others, [ST12, DL15, AMT15, FKKS15, GK16, FKKS17, Sud20, QYZ18, BQT18, DNDD<sup>+</sup>18, QBJT19, DS19]. We refer to the later papers or the survey [NPR20] for a detailed discussion of the state of the art.

In theoretical studies on how randomized search heuristics cope with noise, the usual assumption is that all fitness evaluations are subject to independently sampled noise. Also, it is usually assumed that whenever the fitness of a search point is used, say in a selection step, then it is evaluated anew. This was not done, e.g., in [ST12], but as [DHK12] points out, a much better performance can be obtained when reevaluating solutions.

Various models of noise have been investigated so far. In *prior noise* models, the search point to be evaluated is subject to a stochastic modification and the algorithm learns the fitness of the disturbed search point (but not the disturbed search point itself). The following prior noise models from the literature will be covered in this work. In this description, we always assume that we try to learn the fitness  $f$  of a search point  $x$ .

**One-bit noise with probability  $p$ :** With probability  $p$ , the fitness of a random Hamming neighbor of  $x$  is returned, that is,  $y$  is obtained from  $x$  by flipping a bit chosen uniformly at random and then  $f(y)$  is returned; otherwise, the correct fitness  $f(x)$  is returned. An asymmetric version of this

noise was considered in [QYZ18]. We do not regard this here, but note that all our results hold for this noise as well.

**Independent bit-flip noise with rate  $q$ :** From  $x$ , a search point  $y$  is obtained by flipping each bit independently with probability  $q$ ; then  $f(y)$  is returned.

**$(p, q)$ -noise:** With probability  $p$ , a search point  $y$  is obtained from  $x$  by flipping each bit independently with probability  $q$  and  $f(y)$  is returned; otherwise,  $f(x)$  is returned. Note that the probability that the fitness of the original search point is returned is  $p + (1 - p)(1 - q)^n$ . The  $(p, q)$ -noise model contains as special case  $(1, q)$  the independent bit-flip noise with rate  $q$ .

Very roughly speaking, the typical results for these types of noise are that  $(1 + 1)$ -type algorithms remain efficient when noise occurs with a low probability, e.g., of  $O(\frac{\log n}{n})$  for the ONEMAX benchmark or  $O(\frac{\log n}{n^2})$  for the LEADINGONES benchmark, but that asymptotically higher noise levels lead to super-polynomial runtimes. By using non-trivial populations, resampling, or resorting to other algorithms such as estimation-of-distribution algorithms, the robustness to noise can be considerably improved.

In the *posterior noise* model, the search point  $x$  is first correctly evaluated, but then the obtained fitness  $f(x)$  is disturbed. The most common posterior noise is **additive noise**, that is, the returned fitness is  $f(x) + X$ , where  $X$  is a random variable sampled from some given distribution, which does not depend on  $x$  (that is, for all search points the difference between the true and the noisy fitness is identically distributed). The most common setting is that  $X$  follows a Gaussian distribution. We note that regardless of  $X$ , this noise gives a correct comparison of two search points of different quality with probability at least  $\frac{1}{2}$ .

**Proposition 2.** *Let  $x, y \in \{0, 1\}^n$  with  $f(x) > f(y)$  (resp.  $f(x) \geq f(y)$ ). Let  $X, Y$  be independent and identically distributed real-valued random variables. Then  $\Pr[f(x) + X > f(y) + Y] \geq \frac{1}{2}$  (resp.  $\Pr[f(x) + X \geq f(y) + Y] \geq \frac{1}{2}$ ).*

*Proof.* By symmetry, we have  $\Pr[X \geq Y] = \Pr[Y \geq X]$ . Hence

$$\begin{aligned} 1 &= \Pr[(X \geq Y) \vee (Y \geq X)] \\ &\leq \Pr[X \geq Y] + \Pr[Y \geq X] = 2\Pr[X \geq Y]. \end{aligned}$$

This shows  $\Pr[X \geq Y] \geq \frac{1}{2}$  and thus  $\Pr[f(x) + X > f(y) + Y] \geq \Pr[X \geq Y] \geq \frac{1}{2}$  (resp.  $\Pr[f(x) + X \geq f(y) + Y] \geq \Pr[X \geq Y] \geq \frac{1}{2}$ ).  $\square$

Since our aim is showing that also in the presence of extreme noise we still have at most exponential runtimes, we also consider the following **unrestricted adversarial noise with probability  $p$** . In this model, with

probability  $1 - p$  the true fitness is returned. With probability  $p$ , however, an all-powerful adversary decides the returned fitness value. This adversary knows the algorithm, the optimization problem, and the full history of the optimization process. He does not know, though, the outcome of future random events (both concerning the algorithm and the noise).

The following basic observation gives an estimate for the probability that a noisy fitness comparison gives the right result. A corresponding result for additive posterior noise, namely that regardless of the noise distribution the fitnesses are compared correctly with probability at least  $\frac{1}{2}$ , was already shown in Proposition 2.

**Proposition 3.** *Let  $\varepsilon > 0$ . Let  $f : \{0, 1\}^n \rightarrow \mathbb{R}$ . Let  $x, y \in \{0, 1\}^n$  such that  $f(x) \leq f(y)$ . Consider any noise model described above except the one of additive posterior noise. Assume that  $p \leq 1 - \varepsilon$  in the case of one-bit noise or unrestricted adversarial noise,  $(1 - q)^n \geq \varepsilon$  in the case of bit-wise noise,  $1 - p(1 - (1 - q)^n) \geq \varepsilon$  in the case of  $(p, q)$ -noise. Denote by  $\tilde{f}$  the noisy version of  $f$  with our convention that each noise evaluation of  $f$  uses fresh independent randomness. Then  $\Pr[\tilde{f}(x) \leq \tilde{f}(y)] \geq \varepsilon^2$ .*

*Proof.* Under the conditions named above, with probability at least  $\varepsilon$  the noisy fitness returns the true fitness value. Consequently, with probability at least  $\varepsilon^2$  this happens for both  $x$  and  $y$  and we have thus  $\tilde{f}(x) \leq \tilde{f}(y)$ .  $\square$

## 2.3 Benchmark Problems

We now describe the benchmark problems used in this paper. They are all well-known and extensively used, so we refer to the literature [NW10, AD11, Jan13, DN20] for more details. As said earlier, we only regard problems defined on bit-strings of length  $n$ , hence all functions are  $\{0, 1\}^n \rightarrow \mathbb{R}$ . In many respects the easiest benchmark problem is the function **OneMax** defined by  $\text{OM}(x) := \text{ONEMAX}(x) := \|x\|_1 = \sum_{i=1}^n x_i$  for all  $x = (x_1, \dots, x_n) \in \{0, 1\}^n$ . This function is a member of the class of **linear functions**, which are all functions  $f$  such that  $f(x) = \sum_{i=1}^n a_i x_i$  for given  $a_1, \dots, a_n \in \mathbb{R}$ . One often assumes that all  $a_i$  are positive, but we do not need this assumption. A function  $f$  is called strictly **monotonic** (or strictly monotonically increasing), where the word strictly often is omitted, when  $f(x) < f(y)$  for all  $x, y \in \{0, 1\}^n$  with  $x \leq y$  (component-wise) and  $x \neq y$ . This is equivalent to saying that whenever one flips a zero of the argument into a one, the fitness strictly increases. Clearly, linear functions with positive weights  $a_i$  are strictly monotonic. ONEMAX, arbitrary linear functions, and monotonic functions can all be solved easily via randomized local search (RLS),

namely in time  $O(n \log n)$ , as follows from a classic coupon collector argument. The  $(1 + 1)$  EA with mutation rate  $\Theta(\frac{1}{n})$  solves all linear functions in time  $O(n \log n)$  [DJW02, DG13, Wit13], but this is less obvious; for the  $(1 + 1)$  EA with fast mutation, an  $O(n \log n)$  runtime bound is only known for ONEMAX [DLMN17]. The situation for strictly monotonic functions is complicated [Jan07, DJS<sup>+</sup>13, CDF14, Len18], but the latest result [LMS19] shows at least that the  $(1 + 1)$  EA with standard mutation rate  $\frac{1}{n}$  solves all strictly monotonic functions in time  $O(n \log^2 n)$ . For the Metropolis algorithm, a temperature of  $T \leq \frac{1}{2 \ln n}$  is easily seen to suffice to optimize ONEMAX in time  $O(n \log n)$ , whereas a temperature of  $T \geq \frac{2}{\ln n}$  leads to a negative drift in a large range around the optimum, and thus to a super-polynomial runtime. The precise boundary between polynomial and super-polynomial, unfortunately hard to interpret, was determined in [KWW09]. We are not aware of such a result for simulated annealing, but it is clear that the cooling schedule needs to lead to temperatures for which the Metropolis algorithm is polynomial for a sufficiently long time to ensure a polynomial runtime.

When noise comes into play, RLS and  $(1 + 1)$  EA can stand low noise levels when optimizing ONEMAX. For the one-bit noise model, any  $p = O(\frac{\log n}{n})$  implies a polynomial runtime, higher values of  $p$  give super-polynomial runtimes [Dro04]. Similar results hold for bit-wise noise and  $(p, q)$  noise [GK16, QBJT19, DNDD<sup>+</sup>18]. For additive posterior noise, the  $(1 + 1)$  EA has a polynomial runtime on ONEMAX if the variance of the noise distribution is  $\sigma^2 = O(\frac{\log n}{n})$ . If the noise has a Gaussian distribution, then the runtime is polynomial if the variance is at most  $\sigma^2 \leq \frac{1}{4 \ln n}$ , and it is super-polynomial, if  $\sigma^2 \geq (1 + \varepsilon) \frac{1}{4 \ln n}$  for a constant  $\varepsilon > 0$  [GK16].

Still unimodal (that is, not having true local optima), but not anymore strictly monotonic is the classic **LeadingOnes** function, which counts the number of ones up to the first zero. Formally,  $\text{LEADINGONES}(x) := \max\{i \in [0..n] \mid \forall j \in [1..i] : x_j = 1\}$ . Both RLS and  $(1 + 1)$  EA optimize LEADINGONES in time  $\Theta(n^2)$ . With one-bit noise  $p = O(\frac{\log n}{n^2})$  gives a polynomial runtime and larger values of  $p$  lead to super-polynomial runtimes of the  $(1 + 1)$  EA; again, similar results hold for bit-wise and  $(p, q)$  noise [QBJT19, Sud20]. For additive posterior noise, the runtime of the  $(1 + 1)$  EA on LEADINGONES remains  $O(n^2)$  if the noise distribution has a variance of  $\sigma \leq \frac{1}{12en^2}$ . Another unimodal, but not strictly monotonic class of functions that has been the subject of a runtime analysis are **monotone polynomials** of degree  $d$ . For these, a (tight) upper bound of  $O(2^{\frac{d}{2}} \log(\frac{n}{d} + 1))$  has been shown for the expected runtime of RLS [WW05]. For the  $(1 + 1)$  EA, only weaker bounds were proven, but the  $O(2^{\frac{d}{2}} \log(\frac{n}{d} + 1))$  was conjectured as well [WW05].

The classic multimodal benchmark is the class of **jump functions**. The jump function with *jump parameter (jump size)*  $k \in [1..n]$  is defined by

$$\text{JUMP}_{nk}(x) = \begin{cases} \|x\|_1 + k & \text{if } \|x\|_1 \in [0..n-k] \cup \{n\}, \\ n - \|x\|_1 & \text{if } \|x\|_1 \in [n-k+1..n-1]. \end{cases}$$

Hence for  $k = 1$ , we have a fitness landscape isomorphic to the one of ONEMAX, but for larger values of  $k$  there is a fitness valley (“gap”)  $G_{nk} := \{x \in \{0,1\}^n \mid n-k < \|x\|_1 < n\}$  consisting of the  $k-1$  highest sub-optimal fitness levels of the ONEMAX function. This valley is impossible to cross for RLS, very hard to cross for the Metropolis algorithm (an expected optimization of  $\exp(\Omega(n))$  was stated without proof in [LOW19]), and hard to cross for the  $(1+1)$  EA. When using standard bit mutation with mutation rate  $\frac{1}{n}$ , the probability to generate the optimum from a parent on the local optimum is only  $p_k := (1 - \frac{1}{n})^{n-k} n^{-k} < n^{-k}$ . For this reason, the runtime of the  $(1+1)$  EA on  $\text{JUMP}_{nk}$  is  $\Theta(n^k)$  [DJW02]. When the fitness in the gap region is not lower than the one of the local optimum, but equal to it (so that there is no local optimum, but all search points in Hamming distance 1 to  $k$  from the optimum have the same fitness  $n$ ), then we have the **plateau function**  $\text{PLATEAU}_{nk}$  introduced in [AD18]. While not multimodal, it is still difficult to optimize and, for  $k$  constant, a runtime of  $\Theta(n^k)$  was shown. The special case  $\text{PLATEAU}_{nn}$  is the famous **needle function** or needle-in-the-haystack problem, where all search points apart from the optimum have the same fitness. Since a black-box optimization algorithm optimizing this function receives no useful feedback during the optimization (except when the optimum is found), no better algorithms than exhaustive search can exist for this problem. We are not aware of any proven results on the optimization of jump or plateau functions subject to noise, but we would conjecture that the runtime is roughly the maximum of the noisy runtime on ONEMAX and the noise-free runtime on the jump or plateau function. In particular for the case of jump functions, this is not obvious since the algorithm could also profit from noise, allowing it to accept a worse search point which is closer to the target. A profit from this effect was mostly ruled out for comma selection [Doe20a] and we expect a similar situation here.

## 2.4 A Technical Tool

Since we shall use it twice, we present here the additive drift theorem for upper bounds of He and Yao [HY01] (see also the recent survey [Len20]), which allows to translate an expected additive progress in a random process (or a lower bound on it) into an upper bound on an expected hitting time.



**Theorem 4.** Let  $S \subseteq \mathbb{R}_{\geq 0}$  be finite and  $0 \in S$ . Let  $X_0, X_1, \dots$  be a random process taking values in  $S$ . Let  $\delta > 0$ . Let  $T = \inf\{t \geq 0 \mid X_t = 0\}$ . Assume that for all  $t \geq 0$  and all  $s \in S \setminus \{0\}$  we have  $E[X_t - X_{t+1} \mid X_t = s] \geq \delta$ . Then  $E[T] \leq \frac{E[X_0]}{\delta}$ .

### 3 Proving Exponential Upper Bounds

We now state our general technical result which in many situations allows one to prove exponential upper bounds without greater difficulties. We formulate our result for single-trajectory algorithms since this is notationally convenient and covers most of our applications. We show an exponential upper bound for an EA with non-trivial parent population in Section 5. The result below is formulated for hitting a general search point  $x^*$ , but the natural application will be for  $x^*$  being the optimum solution. We remind the reader that the key argument of the proof below has already appeared in the conference paper [DJW98], but has, to the best of our knowledge, not been used again since then.

**Theorem 5.** Let  $A$  be a single-trajectory search algorithm for the optimization of pseudo-Boolean functions. Let  $f : \{0, 1\}^n \rightarrow \mathbb{R}$  and let  $x^* \in \{0, 1\}^n$ . Assume that we use  $A$  to optimize  $f$ , possibly in the presence of noise. Assume that this optimization process satisfies the following property.

**(Acc)** There is a number  $0 < c \leq 1$  such that the following is true. Let  $t \geq 1$  and  $x, z \in \{0, 1\}^n$  such that  $x \neq x^*$ ,  $H(x, z) = 1$ , and  $H(z, x^*) = H(x, x^*) - 1$ . Regardless of what happened in the first  $t - 1$  iterations of optimization process, if  $x^{(t-1)} = x$ , then  $\Pr[x^{(t)} = z] \geq \frac{c}{n}$ .

Let  $T = \min\{t \geq 0 \mid x^{(t)} = x^*\}$ . Then  $T$  is stochastically dominated by  $n \text{Geom}((\frac{c}{e})^n)$ . In particular,  $E[T] \leq n(\frac{e}{c})^n$ .

*Proof.* The key argument of this proof is that any interval of  $n$  iterations with probability at least  $(\frac{c}{e})^n$  generates  $x^*$ . More precisely, let  $t \geq 0$ . We condition on  $x^{(t)}$  having any fixed value  $x \in \{0, 1\}^n$  different from  $x^*$ . We also condition on the history of the process up to iteration  $t - 1$ . We now show that conditional on all this, we have  $\Pr[x^* \in \{x^{(t+1)}, \dots, x^{(t+n)}\}] \geq (\frac{c}{e})^n$ .

Let  $d = H(x, x^*)$ . We consider the event  $\mathcal{E}$  that for all  $i \in [1..d]$ , we have  $H(x^{(t+i-1)}, x^{(t+i)}) = 1$  and  $H(x^{(t+i)}, x^*) = d - i$ , that is, that the algorithm  $A$  reduced the distance to  $x^*$  by exactly one in each iteration. Let  $\mathcal{P}$  be the set of all such paths  $x = z^{(0)}, \dots, z^{(d)} = x^*$ , that is, the set of all  $(z^{(0)}, \dots, z^{(d)})$  such that  $z^{(0)} = x$  and for all  $i \in [1..d]$ , we have  $H(z^{(t+i-1)}, z^{(t+i)}) = 1$  and  $H(z^{(t+i)}, x^*) = d - i$ . Each such path can alternatively be described by the

order in which the bits  $x$  and  $x^*$  differ in are flipped. Consequently,  $|\mathcal{P}| = d!$ . By assumption (Acc), the probability that the algorithm follows exactly such a path, that is, that  $x^{(t+i)} = z^{(i)}$  for all  $i \in [0..d]$ , is at least  $(\frac{c}{n})^d$ . We thus have

$$\begin{aligned} \Pr[\mathcal{E}] &= \sum_{P \in \mathcal{P}} \Pr[A \text{ follows the path } P] \\ &\geq |\mathcal{P}| \left(\frac{c}{n}\right)^d = \frac{d!}{\left(\frac{n}{c}\right)^d} \\ &\geq \frac{n!}{\left(\frac{n}{c}\right)^n} \geq \left(\frac{c}{e}\right)^n, \end{aligned} \tag{1}$$

where the last line uses  $c \leq 1$  and the elementary estimate  $n! \geq (\frac{n}{e})^n$ , see, e.g., [Doe20e, (4.13)].

With this, we now know that any interval of  $n$  iterations (“phase”) finds the optimum with probability at least  $p = (\frac{c}{e})^n$ . Consequently, the number of phases until the optimum is found is stochastically dominated by a geometric law with parameter  $p$ , see [Doe19] for more details on this argument. Since each phase by definition lasts exactly  $n$  iterations, the number of iterations until the optimum is found is stochastically dominated by  $n \text{Geom}(p)$  and the expected number of iterations is at most  $E[T] \leq n(\frac{e}{c})^n$ .  $\square$

## 4 Applications of the Main Result

Despite its simplicity, Theorem 5 allows to prove, often without much effort, exponential upper bounds for various different algorithmic settings, as we now show.

### 4.1 Noisy Optimization of Weakly Monotonic Functions

As one such result, we now prove that all  $(1 + 1)$ -type algorithms discussed in Section 2.1 optimize any weakly monotonic function in at most exponential time even in the presence of any noise discussed in Section 2.2 as long as the noise probability is at most  $1 - \varepsilon$ ,  $\varepsilon > 0$  a constant, in the cases of prior or adversarial noise. We recall that the only previous result in this direction [Sud20] shows this claim in the particular case of the  $(1 + 1)$  EA optimizing the LEADINGONES function subject to one-bit or  $(p, q)$  prior noise with noise probability at most  $\frac{1}{2}$ .

We say that a function  $f : \{0, 1\}^n \rightarrow \mathbb{R}$  is *weakly monotonic* (or weakly monotonically increasing) if for all  $x, y \in \{0, 1\}^n$  the condition  $x \leq y$

(component-wise) implies  $f(x) \leq f(y)$ . The class of weakly monotonic functions includes, obviously, all strictly monotonic functions and thus in particular the classic benchmarks ONEMAX and linear functions (with non-negative coefficients). However, this class also contains more difficult functions like LEADINGONES, monotone polynomials, plateau functions, and the needle function.

**Theorem 6.** *Let  $\varepsilon > 0$  be a constant. Let  $A$  be one of the randomized search heuristics RLS, the Metropolis algorithm, simulated annealing, or the  $(1+1)$  EA using standard bit mutation with mutation rate  $\frac{1}{n}$  or using the fast mutation operator with  $\beta > 1$ . Let  $f : \{0,1\}^n \rightarrow \mathbb{R}$  be any weakly monotonic function. Assume that  $A$  optimizes  $f$  under one of the following noise assumptions: one-bit noise or unrestricted adversarial noise with  $p \leq 1-\varepsilon$ , bit-wise noise with  $(1-q)^n \geq \varepsilon$ ,  $(p,q)$ -noise with  $1-p(1-(1-q)^n) \geq \varepsilon$ , or posterior noise with an arbitrary noise distribution.*

*Then there is a constant  $C > 1$ , depending only on  $\varepsilon$  and the choice of  $A$ , such that the time  $T$  to sample the optimum  $(1, \dots, 1)$  of  $f$  is stochastically dominated by  $n \text{Geom}(C^{-n})$ . In particular, the expected optimization time is at most  $E[T] \leq nC^n$ .*

*Proof.* By Theorem 5, it suffices to show that condition (Acc) is satisfied for  $x^* = (1, \dots, 1)$ . To this aim, let  $x, z \in \{0,1\}^n$  such that  $H(x, z) = 1$  and  $H(x, x^*) = H(z, x^*) + 1$ . Assume that for some iteration  $t$  the parent individual satisfies  $x^{(t-1)} = x$ . By Proposition 1, there is a constant  $c_A$  such that the offspring  $y$  generated by  $A$  in this iteration is equal to  $z$  with probability at least  $\frac{c_A}{n}$ . By the weak monotonicity of  $f$ , we have  $f(z) \geq f(x)$ . By Proposition 2 or 3, there is a constant  $c_N = \min\{\frac{1}{2}, \varepsilon^2\}$  depending on the noise model such that the noisy evaluations of both  $x^{(t-1)}$  and  $y = z$  in iteration  $t$  with probability at least  $c_N$  return an at least as good fitness value for  $z$  as for  $x$ . In this case,  $A$  accepts  $z$  with probability one, that is, we have  $x^{(t)} = z$ . In summary, we have shown  $\Pr[x^{(t)} = z] \geq \frac{c_A c_N}{n}$  as desired. Now Theorem 5 immediately gives the claim of Theorem 6 with  $C = \frac{e}{c_A c_N}$ .  $\square$

#### 4.1.1 Discussion: The Base of the Exponential

In this first work on exponential upper bounds, we did not try to optimize the base of the exponential function, that is, the constant  $C$  such that the upper bound is  $\text{poly}(n)C^n$ . As discussed in the introduction, this constant is important to understand the performance of the algorithm and to compare it to other exponential time algorithms such as blind random search and exhaustive search.

Since optimizing implicit constants often is highly non-trivial and technical, we shall not start this endeavor here, but only briefly describe how our current approach compares to the one of [Sud20] and where we see room for improving our constants, possibly below the  $\Theta(2^n)$  runtime of blind random search and exhaustive search.

We first note that the proof of [Sud20], which also is not optimized for giving good constants, shows an upper bound that is at least  $\exp(3en) \geq (3480)^n$ . Under the noise assumptions taken in [Sud20], we have a probability of at least  $c_N \geq \frac{1}{4}$  that parent and offspring are not subject to noise. Regarding the  $(1+1)$  EA, the probability that a particular Hamming neighbor of the parent is generated as offspring is at least  $\frac{1}{en}$ , that is, in the notation of the proof of Theorem 5 we have  $c_A = \frac{1}{e}$ . This would give a runtime bound of at most  $n(\frac{e}{c_A c_N})^n = n(4e^2)^n \leq n(30)^n$ . For other scenarios, the constant is slightly better. For example, for RLS, Metropolis algorithm, and simulated annealing, we have  $c_A = 1$ . For posterior noise,  $c_N$  can be taken as  $\frac{1}{2}$  as shown in Proposition 2. Hence for these combinations, the above runtime could be estimated by  $n(2e)^n \leq n(5.5)^n$ .

We now argue why we are optimistic that with additional arguments, the constants can be improved when taking into account the particular situations, that is, the noise model, the objective function, and the EA. For example, when optimizing any weakly monotonic function subject to 1-bit noise, we accept an offspring strictly dominating the parent (in the proof of Theorem 5) except when the noise flips a zero-bit of the parent or a one-bit of the offspring. This undesired event happens with probability at most  $\frac{1}{2}$ , hence we have  $c_N = \frac{1}{2}$  instead of  $c_N = \frac{1}{4}$  before.

When taking more details of the fitness function into account, we have additional arguments to reduce the impact of the noise. When optimizing ONEMAX, for example, 1-bit noise is detrimental (in our proof) only if both a one-bit of the offspring and a zero-bit of the parent is flipped, increasing  $c_N$  further to  $c_N = (1 - O(\frac{1}{n}))^{\frac{15}{16}}$ .

Taking into account the particular algorithm also allows to fine-tune the analysis. Let us take the  $(1+1)$  EA as an example. If in our main argument we do not wait for the lucky event that in each iteration we approach the target by one Hamming step, but by two steps, then the number of different ways to go from a starting point in (pessimistic) Hamming distance  $n$  to the optimum reduces by a factor of  $2^{n/2}$ , but we need to pay the price of  $\frac{1}{e}$  for flipping exactly one bit only  $n/2$  times instead of  $n$  times. So we reduce the runtime bound by a factor of  $(2/e)^{n/2}$ . Note that in addition now one-bit noise has no chance to prevent us from accepting the desired parent if we optimize ONEMAX.

Finally, a mild understanding of the optimization process can be exploited. In our analysis, we pessimistically pretended that we start a run towards the target from a search point with maximal distance. Unless we are optimizing a very deceptive problem, one should be able to argue that the typical starting point will be rather in distance  $\frac{n}{2}$ , more precisely, that it takes an expected short time to reach such a search point, and then start the run towards the target from there. Note that such an argument does not need a mixing time argument as precise as in [Sud20]. With a little more understanding, we might also be able to argue that the algorithm easily (that is, in expected polynomial time) reaches a search point that is even closer to the optimum. Depending on the strength of the noise and the fitness function, it can be possible to argue that up to a certain fitness level, the influence of the fitness outweighs the negative effect of the noise. Consequently, the resulting drift lets the algorithm quickly approach this fitness level. Such arguments have been used, e.g., in [DNDD<sup>+</sup>18].

We shall not elaborate on these ideas further, but in summary they give us some optimism that further, more detailed and problem-specific studies will be able to show exponential upper bounds which are of order  $C^n$  for a constant  $C$  that is less than 2 (which would prove the algorithm superior to random search).

## 4.2 Noisy Optimization of Jump Functions

To show the versatility of our general approach, we continue with a number of results of varying flavor. We first show that the  $(1+1)$  EA can optimize noisy jump functions with jump size at most  $\frac{n}{\ln n}$  in exponential time.

**Theorem 7.** *The result of Theorem 6 holds also for the  $(1+1)$  EA optimizing  $\text{JUMP}_{nk}$  when  $k \leq \frac{n}{\ln n}$ .*

We omit a formal proof since it is very similar to the proof of Theorem 5 and 6. Let  $x$  be any search point outside the gap region of the jump function and let  $x = z^{(0)}, z^{(1)}, \dots, z^{(d)} = (1, \dots, 1)$  be any shortest path in the Hamming cube from  $x$  to the optimum. Note that  $z^{(d-k)}$  is a local optimum of the jump function  $\text{JUMP}_{nk}$ . We say that the  $(1+1)$  EA follows this path from iteration  $t$  on if  $x^{(t+i)} = z^{(i)}$  for all  $i \in [0..d-k]$  and  $x^{(t+d-k+1)} = (1, \dots, 1)$ . Since the assumptions of this theorem imply condition (Acc) of Theorem 5 (see the proof of Theorem 6), the probability that the  $(1+1)$  EA follows this path is at least  $(\frac{c}{n})^{d-k} n^{-k} (1 - \frac{1}{n})^{n-k}$ . When applying a union bound over all these paths, we have to take care that  $k!$  paths describe the same event.

Consequently, the probability that the  $(1+1)$  EA follows some path from  $x$  to  $(1, \dots, 1)$  is at least  $\frac{d!}{k!} (\frac{c}{n})^d (1 - \frac{1}{n})^{n-k} \geq \frac{1}{k!} (\frac{c}{e})^n \frac{1}{e}$  as in the proof of

Theorem 5. Since  $k \leq \frac{n}{\ln n}$ , we have  $k! \leq k^k \leq (\frac{n}{\ln n})^{\frac{n}{\ln n}} = \exp(\frac{n}{\ln n} \ln(\frac{n}{\ln n})) \leq \exp(n)$ . Hence the probability that the optimum is found in  $d-k+1$  iterations is at least  $(\frac{c}{e^2})^n \frac{1}{e}$ . The remainder of the proof is analogous to the one of Theorem 5.

### 4.3 Optimization of OneMax Under Extreme Bit-Wise Noise

The following result shows that our general method can also exploit particular noise models. Here, for example, we show that ONEMAX can be optimized in exponential time even in the presence of bit-wise noise with constant rate  $q < 1$ . Recall that this means that the search point to be evaluated is disturbed in an expected number of  $qn$  bits!

**Theorem 8.** *Let  $\varepsilon > 0$  be a constant. Let  $A$  be one of the randomized search heuristics RLS, the Metropolis algorithm, simulated annealing, or the  $(1+1)$  EA using standard bit mutation with mutation rate  $\frac{1}{n}$  or using the fast mutation operator with  $\beta > 1$ . Consider optimizing the ONEMAX benchmark function via  $A$  in the presence of bit-wise noise with rate  $q \leq 1 - \varepsilon$ . Then the expected time to find the optimum is at most  $nC^n$ , where  $C$  is a constant depending on  $\varepsilon$  and the algorithm used.*

The theorem above contains as special case  $q = \frac{1}{2}$  the situation that the noisy fitness returns the fitness of a search point uniformly distributed in  $\{0, 1\}^n$ . Here, obviously, the algorithm cannot gain any advantage from the fitness evaluations.

*Proof.* The proof is identical to the one of Theorem 6 except that we cannot invoke Proposition 3 to argue that with constant probability the better of two Hamming neighbors also has the higher noisy fitness. So we complete the proof by showing this missing statement. To this aim, let  $x, y \in \{0, 1\}^n$  such that  $H(x, y) = 1$  and  $\text{ONEMAX}(x) \leq \text{ONEMAX}(y)$ , which implies  $\text{ONEMAX}(y) = \text{ONEMAX}(x) + 1$ . By assumption, there is a unique bit position  $i \in [1..n]$  such that  $x_i \neq y_i$  (“interesting bit”), and for this we have  $x_i = 0$  and  $y_i = 1$ .

Let  $\tilde{f}_x, \tilde{f}_y$  be one-time samples of the noisy fitness values of  $x$  and  $y$ . We have  $\tilde{f}_x = f(x \oplus m^x)$ , where  $\oplus$  denotes addition in  $\mathbb{Z}_2$  (or logical XOR) and  $m^x \in \{0, 1\}^n$  is random such that each bit of  $m^x$  is 1 with probability  $q$ . Likewise, let  $m^y$  be such that  $\tilde{f}_y = f(y \oplus m^y)$ .

With probability  $(1 - q)^2$ , we have  $m_i^x = m_i^y = 0$ , that is, the interesting bit is not affected by the noise. Conditional on this, we have  $\tilde{f}_x = \|(x \oplus m^x)_{[1..n] \setminus \{i\}}\|_1 + 0$  and  $\tilde{f}_y = \|(y \oplus m^y)_{[1..n] \setminus \{i\}}\|_1 + 1$ . The two random variables

$\|(x \oplus m^x)_{|[1..n] \setminus \{i\}}\|_1$  and  $\|(y \oplus m^y)_{|[1..n] \setminus \{i\}}\|_1$  are identically distributed. By symmetry, we have  $\|(x \oplus m^x)_{|[1..n] \setminus \{i\}}\|_1 \leq \|(y \oplus m^y)_{|[1..n] \setminus \{i\}}\|_1$  with probability at least  $\frac{1}{2}$ . Consequently,  $\Pr[\tilde{f}_x < \tilde{f}_y] \geq \frac{1}{2}(1 - q)^2 = \frac{1}{2}\varepsilon^2$ . This shows the missing claim and completes the proof.  $\square$

## 4.4 Fitness Proportionate Selection

Our general method is not restricted to the analysis of noisy optimization. In this subsection, we prove an upper bound matching an exponential lower bound proven in [HJKN08]. The main result of [HJKN08] is that the  $(1 + 1)$  EA needs at least exponential time to optimize any linear function with positive coefficients when the usual elitist selection is replaced by fitness-proportionate selection. Here an offspring  $y$  of the parent  $x$  is accepted with probability  $\frac{f(y)}{f(x) + f(y)}$ . We now show that this result is tight, that is, that an exponential number of iterations suffices to optimize any linear function with this algorithm. This result is true for all weakly monotonic functions.

**Theorem 9.** *Let  $A$  be the  $(1 + 1)$  EA with fitness-proportionate selection. Let  $f : \{0, 1\}^n \rightarrow \mathbb{R}_{>0}$  be any weakly monotonic function. Then the first iteration  $T$  in which the optimum of  $f$  is generated satisfies  $E[T] \leq (2e^2)^n$ .*

*Proof.* We use our main tool Theorem 5. Denote by  $x^* = (1, \dots, 1)$  this particular optimum of  $f$ . The probability that the  $(1 + 1)$  EA generates a particular Hamming neighbor of the parent as offspring is  $\frac{1}{n}(1 - \frac{1}{n})^{n-1} \geq \frac{1}{en}$ . The probability that a Hamming neighbor with better fitness accepted is at least  $\frac{1}{2}$ . This shows condition (Acc) of Theorem 5 with  $c = \frac{1}{2e}$ . By the theorem, the expected runtime is at most  $(2e^2)^n$ .  $\square$

## 4.5 Subexponential Upper Bounds

We now show that our method is not restricted to showing runtime bounds that are exponential in the problem dimension. We recall that the  $(1, \lambda)$  EA is a simple non-elitist algorithm working with a parent population of size one, initialized with a random individual. In each iteration, the algorithm creates independently  $\lambda$  offspring via standard bit mutation (here: with mutation rate  $\frac{1}{n}$ ) and takes a random best offspring as new parent. In their very precise determination of the efficiency threshold of the  $(1, \lambda)$  EA on ONEMAX, Rowe and Sudholt [RS14] showed that the  $(1, \lambda)$  EA has a runtime of at least  $\exp(\Omega(n^{\varepsilon/2}))$  when  $\lambda \leq (1 - \varepsilon) \log_{\frac{e}{e-1}}(n)$ ,  $\varepsilon > 0$  a constant. We now show an upper bound of  $\exp(O(n^\varepsilon))$  for this runtime. We do not know what is the right asymptotic order of the exponent. From the fact that there is a



considerable negative drift when the fitness distance is below  $\frac{n^\varepsilon}{2\lambda}$ , we would rather suspect that also a lower bound of  $\exp(\Omega(\frac{n^\varepsilon}{\lambda}))$  iterations, and hence  $\lambda \exp(\Omega(\frac{n^\varepsilon}{\lambda}))$  fitness evaluations, comes true. Since this is not the main topic of this work, we leave this an open problem.

**Theorem 10.** *Let  $0 < \varepsilon < 1$  be a constant. Then there is a constant  $C_\varepsilon$  such that for all  $\lambda \geq (1 - \varepsilon) \log_{\frac{e}{e-1}}(n)$  the expected runtime of the  $(1, \lambda)$  EA on ONEMAX is at most  $\exp(C_\varepsilon n^\varepsilon)$ .*

*Proof.* Our proof uses some of the arguments given in [RS14], however, since we do not require a precise analysis of the easy part of the process, we simply resort to the additive drift theorem (Theorem 4) for that part. For any search point  $x \in \{0, 1\}^n$ , denote by  $d(x) := n - \text{ONEMAX}(x)$  its fitness distance (which also is its Hamming distance) to the optimum.

For the easy part that the fitness distance is more than  $d_0 := \frac{2e^2 n^\varepsilon}{\lambda}$ , we first show that starting from an arbitrary solution, with probability at least  $\frac{1}{2}$  within  $(1 + o(1)) \frac{2}{e} n^{2-\varepsilon}$  iterations, a solution  $x$  with  $d(x) \leq d_0$  is found. To ease the notation, we show our claim for the initial solution  $x^{(0)}$ . Since the sequence of search points  $x^{(t)}$  generated by the  $(1, \lambda)$  EA is a Markov chain, the claim holds analogously for any starting iteration.

Hence let  $x^{(0)}$  be arbitrary. For  $t = 0, 1, 2, \dots$ , let  $X_t = \max\{0, d(x^{(t)}) - d_0\}$  the distance to our target of having a  $d$ -value of at most  $d_0$ . Let  $t$  be such that  $X_{t-1} > 0$ . By Lemma 7 of [RS14], we have  $E[X_t - X_{t-1} \mid X_t > X_{t-1}] \leq e$ . The event  $X_t > X_{t-1}$  itself means that all  $\lambda$  offspring generated in iteration  $t$  have a fitness inferior to the one of the parent. For an offspring to have an inferior ONEMAX fitness, it is necessary that at least one of the 1-bits of the parent was flipped. This happens with probability at most  $1 - (1 - \frac{1}{n})^{n-d_0} \leq 1 - \frac{1}{e}$ , since  $X_{t-1} > 0$ . Consequently,

$$\Pr[X_t > X_{t-1}] \leq (1 - \frac{1}{e})^\lambda \leq n^{-(1-\varepsilon)}.$$

In summary,

$$\begin{aligned} E[\max\{0, X_t - X_{t-1}\}] &= E[X_t - X_{t-1} \mid X_t > X_{t-1}] \cdot \Pr[X_t > X_{t-1}] \\ &\leq en^{-(1-\varepsilon)}. \end{aligned}$$

For the drift towards the target, we first observe that an offspring  $y$  is better than the parent  $x$  with probability at least

$$d(x) \frac{1}{n} (1 - \frac{1}{n})^{n-1} \geq \frac{d(x)}{en}.$$

Consequently, the probability that at least one offspring is better, is at least  $1 - (1 - \frac{d(x)}{en})^\lambda$ . By the elementary, but very useful Lemma 8 of [RS14], this is at least

$$\begin{aligned} 1 - \frac{1}{\frac{\lambda d(x)}{en} + 1} &\geq 1 - \frac{1}{\frac{\lambda d_0}{en} + 1} \\ &= 1 - \frac{1}{2en^{-(1-\varepsilon)} + 1} \\ &= \frac{2en^{-(1-\varepsilon)}}{2en^{-(1-\varepsilon)} + 1} \geq (1 - o(1))2en^{-(1-\varepsilon)}. \end{aligned}$$

Recalling that a fitness improvement means that the process  $(X_t)$  goes down by at least one, we have,  $E[\max\{0, X_{t-1} - X_t\}] = E[X_{t-1} - X_t \mid X_t > X_{t-1}] \cdot \Pr[X_t > X_{t-1}] \geq 1 \cdot (1 - o(1))2en^{-(1-\varepsilon)}$ .

We have just computed that the drift of the process  $(X_t)$  satisfies

$$\begin{aligned} E[X_{t-1} - X_t \mid X_{t-1} > 0] &\geq E[\max\{0, X_{t-1} - X_t\}] - E[\max\{0, X_t - X_{t-1}\}] \\ &\geq (1 - o(1))2en^{-(1-\varepsilon)} - en^{-(1-\varepsilon)} = (1 - o(1))en^{-(1-\varepsilon)}. \end{aligned}$$

Consequently, by the additive drift theorem (Theorem 4), the first time  $T$  to have  $d(x^{(T)}) \leq d_0$  satisfies

$$E[T] \leq \frac{n}{(1 - o(1))en^{-(1-\varepsilon)}} = (1 + o(1))\frac{1}{e}n^{2-\varepsilon}.$$

By Markov's inequality, with probability at least  $\frac{1}{2}$ ,  $2E[T] \leq (1 + o(1))\frac{2}{e}n^{2-\varepsilon}$  iterations suffice to have a parent individual  $x^{(t)}$  with  $d(x^{(t)}) \leq d_0$ .

We now argue, reusing some arguments from the proof of Theorem 5, that now another  $d_0$  iterations suffice to find the optimum with probability at least  $\exp(-C_en^\varepsilon)$ . In a similar manner as above, we see that if the current parent individual has a fitness distance of  $d \leq d_0$ , then the probability to increase the fitness of the parent in one iteration  $1 - (1 - \frac{d}{en})^\lambda \geq \frac{\lambda d}{en + \lambda d} \geq \frac{\lambda d}{(e + 2e^2)n}$  by our assumption that  $d \leq d_0$  and the blunt estimate  $\lambda d \leq 2e^2n$ .

Consequently, the probability that  $d_0$  iterations suffice to find the optimum is at least

$$\begin{aligned}
\prod_{d=1}^{d_0} \frac{d\lambda}{(e + 2e^2)n} &= \left( \frac{\lambda}{(e + 2e^2)n} \right)^{d_0} (d_0)! \geq \left( \frac{d_0\lambda}{e(e + 2e^2)n} \right)^{d_0} \\
&= \exp \left( -d_0 \ln \left( \frac{e(e + 2e^2)n}{d_0\lambda} \right) \right) \\
&= \exp \left( -\frac{2e^2 n^\varepsilon}{\lambda} \ln \left( \frac{e(e + 2e^2)n^{1-\varepsilon}}{2e^2} \right) \right) \\
&\leq \exp(-C'_\varepsilon n^\varepsilon)
\end{aligned}$$

for a suitable constant  $C'_\varepsilon$ ; note that we used  $\lambda = \Omega(\log n)$  in the last step.

Now as in the proof of Theorem 5, the runtime is stochastically dominated by  $(1 + o(1))\frac{2}{\varepsilon}n^{2-\varepsilon} + d_0$  times a geometric random variable with success rate  $p = \frac{1}{2}\exp(-C'_\varepsilon n^\varepsilon)$ , which gives the claimed expected runtime of at most  $(1 + o(1))(\frac{2}{\varepsilon}n^{2-\varepsilon} + d_0)\frac{1}{p} = \exp(C_\varepsilon n^\varepsilon)$  for a suitable choice of  $C_\varepsilon$ .  $\square$

## 5 An Exponential Upper Bound for the Mutation-based Simple Genetic Algorithm

To indicate that our general analysis method is not restricted to  $(1 + 1)$ -type algorithms, we now prove an exponential upper bound for the mutation-based simple genetic algorithm (simple GA), which uses a non-trivial parent population and which is a generational GA, that is, individuals from the current population are never taken into the next population. We believe that, in principle, this algorithm could be covered via an extension of the framework regarded in Theorem 5, but for the sake of readability we prefer to give an independent analysis even if this requires repeating (in a compact manner) some arguments from the proof of Theorem 5.

The simple GA traditionally is used with crossover [Gol89]. The mutation-only version has been regarded in the runtime analysis community mostly because runtime analyses for crossover-based algorithms are extremely difficult. For example, the first runtime analysis of the crossover-based version is only from 2012 [OW12] and only shows results for relatively small population sizes below  $n^{1/8}$  (the current best result [OW15] covers population sizes up to slightly below  $n^{1/4}$ ), whereas for the mutation-based version already in 2009 a strong result valid for all polynomial population size was shown [NOW09].

We start by making this algorithm precise and then discuss the known results. The mutation-only version of the simple GA with population size  $\mu \in \mathbb{N}$

for the optimization of a non-negative fitness function  $f : \{0, 1\}^n \rightarrow \mathbb{R}_{\geq 0}$  is described in Algorithm 1. The algorithm starts with a population  $P^{(0)}$  of  $\mu$  random individuals from  $\{0, 1\}^n$  (our result will be valid for any initial population, but the usual initialization is random). We view a population  $P$  as an array (tuple)  $P = (P_1, \dots, P_\mu)$  of not necessarily distinct individuals. In each iteration  $t = 1, 2, 3, \dots$ , the simple GA computes from the previous population  $P^{(t-1)}$  a new population  $P^{(t)}$  by  $\mu$  times independently selecting an individual from  $P^{(t-1)}$  via fitness proportionate selection and mutating it via standard bit mutation with mutation rate  $p = \frac{1}{n}$ . Here fitness proportionate selection from  $P^{(t-1)}$  means that the  $i$ -th individual  $P_i^{(t-1)}$  is chosen with probability  $f(P_i^{(t-1)}) / \sum_{j=1}^{\mu} f(P_j^{(t-1)})$ ; except when all individuals have fitness zero, then fitness proportionate selection agrees with uniform selection, that is, the  $i$ -th individual is chosen with probability  $\frac{1}{\mu}$ .

---

**Algorithm 1:** The simple genetic algorithm (simple GA) with population size  $\mu$  and mutation rate  $p = \frac{1}{n}$  to maximize a function  $f : \{0, 1\}^n \rightarrow \mathbb{R}_{\geq 0}$ .

---

```

1 Initialize  $P^{(0)}$  with  $\mu$  individuals chosen independently and uniformly
  at random from  $\{0, 1\}^n$ ;
2 for  $t = 1, 2, \dots$  do
3   for  $i \in [1.. \mu]$  do
4     Select  $x \in P^{(t-1)}$  via fitness proportionate selection;
5     Generate  $P_i^{(t)}$  from  $x$  by flipping each bit independently with
      probability  $p = \frac{1}{n}$ ;

```

---

The existing runtime results for the simple GA are not very encouraging for this algorithm. Apart from results for very small mutation rates  $p = O(n^{-2})$ , see [DL16, Theorem 16] and [DK19, Theorem 22], or results for exponentially transformed fitness functions [NOW09, Theorem 13 and 14], they show that the selection pressure resulting from fitness proportionate selection is not strong enough to give a sufficient progress to the optimum. The precise known results for the performance of Algorithm 1 on the ONEMAX benchmark are the following. [NOW09, Theorem 8] showed that the simple GA with  $\mu \leq \text{poly}(n)$  needs with high probability more than  $2^{n^{1-O(1/\log \log n)}}$  iterations to find the optimum of the ONEMAX function or any search point in Hamming distance at most  $0.003n$  from it. This is only a subexponential lower bound. In [Leh11, Corollary 13], building on the lower bound method from [Leh10], a truly exponential lower bound is shown and this for weaker the task of finding a search point in Hamming distance at most  $0.029n$  from

the optimum, but only for a relatively large population size of  $\mu \geq n^3$  (and again  $\mu \leq \text{poly}(n)$ ). The conditions  $\mu \geq n^3$  and  $\mu \leq \text{poly}(n)$  were removed in [Doe20c]. Lehre [Leh11] also showed an exponential lower bound for the time to reach the optimum for a (mildly) scaled version of fitness proportionate selection and for general  $\Theta(1/n)$  mutation rates. For the crossover-based version of the simple GA the current best results is that with population size  $\mu \leq n^{(1/4)-\varepsilon}$  it takes at least time  $2^{n^{\varepsilon/11}}$  to find a solution that is a small constant factor better than  $\frac{n}{2}$ , which is the expected fitness of a random solution.

For the mutation-based version of the simple GA (Algorithm 1) we now show that the known exponential lower bounds are tight, that is, regardless of the population size (as long as it is not super-exponential) the simple GA finds the optimum of the ONEMAX benchmark in expected time  $\exp(O(n))$ .

**Theorem 11.** *Consider a run of the simple GA with arbitrary population size  $\mu$  and with an arbitrary initial population  $P^{(0)}$  on the ONEMAX benchmark. Let  $T$  be the first iteration in which the optimum of ONEMAX is generated. Then  $T \preceq O(n) \text{Geom}(-\exp(O(n)))$  and  $E[T] = \exp(O(n))$ . In particular, if  $\mu = \exp(O(n))$ , then the optimum is sampled within an exponential number of fitness evaluations.*

*Proof.* We follow the rough outline of the proof of Theorem 5. As there, it suffices to show that for any state of the algorithm the probability that the optimum is found in the next  $O(n)$  iterations is at least  $\exp(-O(n))$ . Then a simple restart argument as in the proof of Theorem 5 shows that  $T \preceq O(n) \text{Geom}(-\exp(O(n)))$  and thus  $E[T] = \exp(O(n))$ .

**Obtaining a fitness of at least  $\frac{n}{3}$ :** To have a better control over the individuals selected via fitness proportionate selection, it helps to have at least one individual with fitness at least  $\frac{n}{3}$  in the population. To this aim, we now show that when starting the simple GA with an arbitrary population, it takes an expected time of at most  $3n$  iterations to have at least one individual in the population that has fitness at least  $\frac{n}{3}$ .

Let  $P^{(0)}$  be an arbitrary initial population and let  $P^{(t)}$ ,  $t = 0, 1, 2, \dots$  be the populations generated in a run of the simple GA started with  $P^{(0)}$ . We regard the following random process  $X_t$ ,  $t = 0, 1, 2, \dots$ . If  $P^{(t)}$  contains at least one individual with fitness  $\frac{n}{3}$  or more, or if  $t \geq 1$  and  $X_{t-1} = n$ , then let  $X_t := n$ . Otherwise, let  $X_t$  be the average fitness of the individuals in  $P^{(t)}$ . This defines a random process in a finite subset of  $[0, n]$ . We estimate  $T_0 := \min\{t \mid X_t = n\}$ . Assume that for some  $t$  we have  $X_t < n$ . We compute the expected fitness of an individual in  $P^{(t+1)}$ . Such an individual  $y$  is generated from a parent  $x$  chosen via fitness proportionate selection from  $P^{(t)}$  by flipping each bit independently with probability  $p = \frac{1}{n}$ . We

compute  $E[\text{OM}(y) \mid x] = \text{OM}(x) + (n - \text{OM}(x))\frac{1}{n} - \text{OM}(x)\frac{1}{n} \geq \text{OM}(x) + \frac{1}{3}$ , using that the  $\text{OM}(x) \leq \frac{n}{3}$ . Since fitness-proportionate selection favors better individuals,  $\text{OM}(x)$  stochastically dominates the fitness of a random individual  $X$  from  $P^{(t)}$ , see [Doe20d, Lemma 12]. Consequently,  $E[\text{OM}(y) - \text{OM}(X)] \geq E[\text{OM}(y) - \text{OM}(x)] \geq \frac{1}{3}$ . Note that  $E[\text{OM}(y)]$  is also the expected average fitness of  $P^{(t+1)}$ . Since  $X_{t+1}$  is this average fitness (which is always at most  $n$ ) or  $n$  (in the case that  $P^{(t+1)}$  contains an individual with fitness at least  $\frac{n}{3}$ ), we have  $E[X_{t+1} - X_t] \geq E[\text{OM}(y) - \text{OM}(x)] \geq \frac{1}{3}$ . Our computation just made was under the assumption that  $X_t < n$ . Hence we have  $E[X_{t+1} - X_t \mid t < T_0] \geq \frac{1}{3}$ . By the additive drift theorem (Theorem 4), we have  $E[T_0] \leq 3n$ . Via a simple Markov bound, we see that with probability at least  $\frac{1}{2}$ , there is a  $t \in [0..6n]$  such that  $P^{(t)}$  contains an individual with fitness at least  $\frac{n}{3}$ .

**From an individual with fitness at least  $\frac{n}{3}$  to the optimal solution:** For any population of the simple GA, an individual  $x$  with fitness  $\text{OM}(x) \geq \frac{n}{3}$  in each selection of a particular parent has a probability of

$$q = \frac{\text{OM}(x)}{\sum_{y \in P} f(y)} \geq \frac{n/3}{\mu n} = \frac{1}{3\mu}$$

of being selected. Hence with probability  $1 - (1 - q)^\mu \geq 1 - (1 - \frac{1}{3\mu})^\mu \geq 1 - \exp(-\frac{1}{3})$ , it is selected at least once as parent (here we used the basic estimate  $1 + r \leq e^r$  valid for all  $r \in \mathbb{R}$ ). For the first offspring generated from  $x$  in this iteration (and for any other, but we want to regard a particular one), the probability that a particular Hamming neighbor of  $x$  results from mutating  $x$  is at least  $\frac{1}{n}(1 - \frac{1}{n})^{n-1} \geq \frac{1}{en}$ . Via an elementary induction we see that if at some time  $t$ , say for simplicity  $t = 0$ , we have an individual  $x^{(0)}$  in the population with  $d := n - \text{OM}(x^{(0)}) \leq \frac{2}{3}n$ , then the probability that at each time  $t \in [1..d]$  there is an  $x^{(t)} \in P^{(t)}$  such that  $\text{OM}(x^{(t)}) = n - (d - t)$  and  $x^{(t)}$  has been generated from  $x^{(t-1)}$  by flipping a single bit, is at least  $(\frac{1}{e} \frac{1}{e} (1 - \exp(-\frac{1}{3})))^n$ , where this number is computed as in (1).

This shows that for any time  $t$  of the run of the simple GA, regardless of the past, the probability that the optimum of ONEMAX is sampled in iterations  $t, \dots, t + 7n - 1$  is at least  $\exp(O(n))$ . This completes the proof of Theorem 11.  $\square$

## 6 Conclusion and Outlook

In this work, we argued for proving exponential runtime guarantees for evolutionary algorithms. Exponential-time algorithms have been a modern sub-field of classic algorithms for around twenty years now for various reasons.

Exponential upper bounds for evolutionary algorithms can not only complement the many exponential lower bounds existing in this field (and by this indicate that the problem is well-understood), but they can also rule out that the algorithm does not suffer from an even worse runtime behavior such as the not uncommon  $n^{\Theta(n)}$  runtime.

With Theorem 5 we provided a simple and general approach towards proving exponential upper bounds. With this method, we easily proved exponential upper bounds for various algorithmic settings.

In this first work on exponential-time evolutionary algorithms, we have surely not developed the full potential of this perspective in evolutionary computation. The clearly most important question for future work is what can be said about the constant  $C$  in the  $\text{poly}(n)C^n$  runtime guarantee. A constant  $C$  less than 2 shows that the algorithm is superior to random or exhaustive search. Taking again the field of classic algorithms as example, another interesting question is if there are EAs with “nice” exponential runtimes such as, e.g., the  $1.0836^n$  runtime of the algorithm of Xiao and Nagamochi [XN13] for finding maximum independent sets in graphs with maximum degree 3.

We mention two directions in which extensions of our method would be desirable. In Section 5, we have analyzed the mutation-based simple GA, but as many previous works, we have shied away from the original simple GA using crossover. Our proof, regarding one lucky lineage, does not apply to settings with crossover. Hence nothing better than the trivial  $n^n$  upper bound is known, which is quite far from the  $2^{n^{\varepsilon/11}}$  lower bound shown in [OW15] for  $\mu \leq n^{(1/4)-\varepsilon}$ .

In Section 4, we have applied our general method, formulated for single-trajectory algorithms, mostly to  $(1+1)$ -type algorithms. The reason is that when regarding more offspring, say in a run of the  $(1+\lambda)$  EA for  $\lambda > 1$ , then each of these could potentially interfere with the survival of the desired offspring. Our methods would remain applicable when also excluding such interferences in the lucky event we are waiting for, but this would lead to upper bounds of order  $\exp(O(n\lambda))$ . We do not believe that larger offspring populations are that detrimental and hope that stronger methods can prove upper bounds which are exponential only in the problem size  $n$ .

## References

- [AD11] Anne Auger and Benjamin Doerr, editors. *Theory of Randomized Search Heuristics*. World Scientific Publishing, 2011.



- [AD18] Denis Antipov and Benjamin Doerr. Precise runtime analysis for plateaus. In *Parallel Problem Solving From Nature, PPSN 2018, Part II*, pages 117–128. Springer, 2018.
- [AMT15] Youhei Akimoto, Sandra Astete Morales, and Olivier Teytaud. Analysis of runtime of optimization algorithms for noisy functions over discrete codomains. *Theoretical Computer Science*, 605:42–50, 2015.
- [BDGG09] Leonora Bianchi, Marco Dorigo, Luca Maria Gambardella, and Walter J. Gutjahr. A survey on metaheuristics for stochastic combinatorial optimization. *Natural Computing*, 8:239–287, 2009.
- [BDN10] Süntje Böttcher, Benjamin Doerr, and Frank Neumann. Optimal fixed and adaptive mutation rates for the LeadingOnes problem. In *Parallel Problem Solving from Nature, PPSN 2010*, pages 1–10. Springer, 2010.
- [BQT18] Chao Bian, Chao Qian, and Ke Tang. Towards a running time analysis of the (1+1)-EA for OneMax and LeadingOnes under general bit-wise noise. In *Parallel Problem Solving from Nature, PPSN 2018, Part II*, pages 165–177. Springer, 2018.
- [CDF14] Sylvain Colin, Benjamin Doerr, and Gaspard Férey. Monotonic functions in EC: anything but monotone! In *Genetic and Evolutionary Computation Conference, GECCO 2014*, pages 753–760. ACM, 2014.
- [DDE15] Benjamin Doerr, Carola Doerr, and Franziska Ebel. From black-box complexity to designing new genetic algorithms. *Theoretical Computer Science*, 567:87–104, 2015.
- [DFW11] Benjamin Doerr, Mahmoud Fouz, and Carsten Witt. Sharp bounds by probability-generating functions and variable drift. In *Genetic and Evolutionary Computation Conference, GECCO 2011*, pages 2083–2090. ACM, 2011.
- [DG13] Benjamin Doerr and Leslie A. Goldberg. Adaptive drift analysis. *Algorithmica*, 65:224–250, 2013.
- [DHK12] Benjamin Doerr, Ashish Ranjan Hota, and Timo Kötzing. Ants easily solve stochastic shortest path problems. In *Genetic and*

- Evolutionary Computation Conference, GECCO 2012*, pages 17–24. ACM, 2012.
- [DJS<sup>+</sup>13] Benjamin Doerr, Thomas Jansen, Dirk Sudholt, Carola Winzen, and Christine Zarges. Mutation rate matters even when optimizing monotone functions. *Evolutionary Computation*, 21:1–21, 2013.
- [DJW98] Stefan Droste, Thomas Jansen, and Ingo Wegener. On the optimization of unimodal functions with the  $(1 + 1)$  evolutionary algorithm. In *Parallel Problem Solving from Nature, PPSN 1998*, pages 13–22. Springer, 1998.
- [DJW02] Stefan Droste, Thomas Jansen, and Ingo Wegener. On the analysis of the  $(1+1)$  evolutionary algorithm. *Theoretical Computer Science*, 276:51–81, 2002.
- [DK19] Benjamin Doerr and Timo Kötzing. Multiplicative up-drift. *CoRR*, abs/1806.01331, 2019.
- [DL15] Duc-Cuong Dang and Per Kristian Lehre. Simplified runtime analysis of estimation of distribution algorithms. In *Genetic and Evolutionary Computation Conference, GECCO 2015*, pages 513–518. ACM, 2015.
- [DL16] Duc-Cuong Dang and Per Kristian Lehre. Runtime analysis of non-elitist populations: from classical optimisation to partial information. *Algorithmica*, 75:428–461, 2016.
- [DLMN17] Benjamin Doerr, Huu Phuoc Le, Régis Makhlara, and Ta Duy Nguyen. Fast genetic algorithms. In *Genetic and Evolutionary Computation Conference, GECCO 2017*, pages 777–784. ACM, 2017.
- [DN20] Benjamin Doerr and Frank Neumann, editors. *Theory of Evolutionary Computation—Recent Developments in Discrete Optimization*. Springer, 2020. Also available at <https://cs.adelaide.edu.au/~frank/papers/TheoryBook2019-selfarchived.pdf>.
- [DNDD<sup>+</sup>18] Raphaël Dang-Nhu, Thibault Dardinier, Benjamin Doerr, Gautier Izacard, and Dorian Nogneng. A new analysis method for evolutionary optimization of dynamic and noisy objective functions. In *Genetic and Evolutionary Computation Conference, GECCO 2018*, pages 1467–1474. ACM, 2018.

- [Doe19] Benjamin Doerr. Analyzing randomized search heuristics via stochastic domination. *Theoretical Computer Science*, 773:115–137, 2019.
- [Doe20a] Benjamin Doerr. Does comma selection help to cope with local optima? In *Genetic and Evolutionary Computation Conference, GECCO 2020*, pages 1304–1313. ACM, 2020.
- [Doe20b] Benjamin Doerr. Exponential upper bounds for the runtime of randomized search heuristics. In *Parallel Problem Solving From Nature, PPSN 2020*. Springer, 2020. To appear.
- [Doe20c] Benjamin Doerr. Lower bounds for non-elitist evolutionary algorithms via negative multiplicative drift. In *Parallel Problem Solving From Nature, PPSN 2020*. Springer, 2020. To appear.
- [Doe20d] Benjamin Doerr. Lower bounds for non-elitist evolutionary algorithms via negative multiplicative drift. *CoRR*, abs/2004.01274, 2020.
- [Doe20e] Benjamin Doerr. Probabilistic tools for the analysis of randomized optimization heuristics. In Benjamin Doerr and Frank Neumann, editors, *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*, pages 1–87. Springer, 2020. Also available at <https://arxiv.org/abs/1801.06733>.
- [Dro04] Stefan Droste. Analysis of the (1+1) EA for a noisy OneMax. In *Genetic and Evolutionary Computation Conference, GECCO 2004*, pages 1088–1099. Springer, 2004.
- [DS19] Benjamin Doerr and Andrew M. Sutton. When resampling to cope with noise, use median, not mean. In *Genetic and Evolutionary Computation Conference, GECCO 2019*, pages 242–248. ACM, 2019.
- [FK10] Fedor V. Fomin and Dieter Kratsch. *Exact Exponential Algorithms*. Springer, 2010.
- [FK13] Fedor V. Fomin and Petteri Kaski. Exact exponential algorithms. *Communications of the ACM*, 56:80–88, 2013.
- [FKKS15] Tobias Friedrich, Timo Kötzing, Martin S. Krejca, and Andrew M. Sutton. Robustness of ant colony optimization to noise. In *Genetic and Evolutionary Computation Conference, GECCO 2015*, pages 17–24. ACM, 2015.

- [FKKS17] Tobias Friedrich, Timo Kötzing, Martin S. Krejca, and Andrew M. Sutton. The compact genetic algorithm is efficient under extreme Gaussian noise. *IEEE Transactions on Evolutionary Computation*, 21:477–490, 2017.
- [GK16] Christian Gießen and Timo Kötzing. Robustness of populations in stochastic environments. *Algorithmica*, 75:462–489, 2016.
- [GKS99] Josselin Garnier, Leila Kallel, and Marc Schoenauer. Rigorous hitting times for binary mutations. *Evolutionary Computation*, 7:173–203, 1999.
- [Gol89] David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., 1989.
- [HJKN08] Edda Happ, Daniel Johannsen, Christian Klein, and Frank Neumann. Rigorous analyses of fitness-proportional selection for optimizing linear functions. In *Genetic and Evolutionary Computation Conference, GECCO 2008*, pages 953–960. ACM, 2008.
- [HY01] Jun He and Xin Yao. Drift analysis and average time complexity of evolutionary algorithms. *Artificial Intelligence*, 127:51–81, 2001.
- [Jan07] Thomas Jansen. On the brittleness of evolutionary algorithms. In *Foundations of Genetic Algorithms, FOGA 2007*, pages 54–69. Springer, 2007.
- [Jan13] Thomas Jansen. *Analyzing Evolutionary Algorithms – The Computer Science Perspective*. Springer, 2013.
- [JB05] Yaochu Jin and Jürgen Branke. Evolutionary optimization in uncertain environments – a survey. *IEEE Transactions on Evolutionary Computation*, 9:303–317, 2005.
- [JW06] Thomas Jansen and Ingo Wegener. On the analysis of a dynamic evolutionary algorithm. *Journal of Discrete Algorithms*, 4:181–199, 2006.
- [KWW09] Lars Kaden, Nicole Weicker, and Karsten Weicker. Metropolis and symmetric functions: a swan song. In *Evolutionary Computation in Combinatorial Optimization, EvoCOP 2009*, pages 204–215. Springer, 2009.

- [Leh10] Per Kristian Lehre. Negative drift in populations. In *Parallel Problem Solving from Nature, PPSN 2010*, pages 244–253. Springer, 2010.
- [Leh11] Per Kristian Lehre. Fitness-levels for non-elitist populations. In *Genetic and Evolutionary Computation Conference, GECCO 2011*, pages 2075–2082. ACM, 2011.
- [Len18] Johannes Lengler. A general dichotomy of evolutionary algorithms on monotone functions. In *Parallel Problem Solving from Nature, PPSN 2018, Part II*, pages 3–15. Springer, 2018.
- [Len20] Johannes Lengler. Drift analysis. In Benjamin Doerr and Frank Neumann, editors, *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*, pages 89–131. Springer, 2020. Also available at <https://arxiv.org/abs/1712.00964>.
- [LMS19] Johannes Lengler, Anders Martinsson, and Angelika Steger. When does hillclimbing fail on monotone functions: an entropy compression argument. In *Analytic Algorithmics and Combinatorics, ANALCO 2019*, pages 94–102. SIAM, 2019.
- [LOW19] Andrei Lissovoi, Pietro S. Oliveto, and John Alasdair Warwicker. On the time complexity of algorithm selection hyperheuristics for multimodal optimisation. In *Conference on Artificial Intelligence, AAI 2019*, pages 2322–2329. AAAI Press, 2019.
- [LZ19] Johannes Lengler and Xun Zou. Exponential slowdown for larger populations: the  $(\mu + 1)$ -EA on monotone functions. In *Foundations of Genetic Algorithms, FOGA 2019*, pages 87–101. ACM, 2019.
- [Müh92] Heinz Mühlenbein. How genetic algorithms really work: mutation and hillclimbing. In *Parallel Problem Solving from Nature, PPSN 1992*, pages 15–26. Elsevier, 1992.
- [NOW09] Frank Neumann, Pietro S. Oliveto, and Carsten Witt. Theoretical analysis of fitness-proportional selection: landscapes and efficiency. In *Genetic and Evolutionary Computation Conference, GECCO 2009*, pages 835–842. ACM, 2009.

- [NPR20] Frank Neumann, Mojgan Pourhassan, and Vahid Roostapour. Analysis of evolutionary algorithms in dynamic and stochastic environments. In Benjamin Doerr and Frank Neumann, editors, *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*, pages 323–357. Springer, 2020. Also available at <https://arxiv.org/abs/1806.08547>.
- [NS20] Frank Neumann and Andrew M. Sutton. Parameterized complexity analysis of randomized search heuristics. In Benjamin Doerr and Frank Neumann, editors, *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*, pages 213–248. Springer, 2020. Also available at <https://arxiv.org/abs/2001.05120>.
- [NW10] Frank Neumann and Carsten Witt. *Bioinspired Computation in Combinatorial Optimization – Algorithms and Their Computational Complexity*. Springer, 2010.
- [OW12] Pietro S. Oliveto and Carsten Witt. On the analysis of the simple genetic algorithm. In *Genetic and Evolutionary Computation Conference, GECCO 2012*, pages 1341–1348. ACM, 2012.
- [OW15] Pietro S. Oliveto and Carsten Witt. Improved time complexity analysis of the simple genetic algorithm. *Theoretical Computer Science*, 605:21–41, 2015.
- [PHST17] Tiago Paixão, Jorge Pérez Heredia, Dirk Sudholt, and Barbora Trubenová. Towards a runtime comparison of natural and artificial evolution. *Algorithmica*, 78:681–713, 2017.
- [QBJT19] Chao Qian, Chao Bian, Wu Jiang, and Ke Tang. Running time analysis of the  $(1 + 1)$ -EA for OneMax and LeadingOnes under bit-wise noise. *Algorithmica*, 81:749–795, 2019.
- [QYZ18] Chao Qian, Yang Yu, and Zhi-Hua Zhou. Analyzing evolutionary optimization in noisy environments. *Evolutionary Computation*, 26:1–41, 2018.
- [RS14] Jonathan E. Rowe and Dirk Sudholt. The choice of the offspring population size in the  $(1, \lambda)$  evolutionary algorithm. *Theoretical Computer Science*, 545:20–38, 2014.

- [RW20] Amirhossein Rajabi and Carsten Witt. Self-adjusting evolutionary algorithms for multimodal optimization. In *Genetic and Evolutionary Computation Conference, GECCO 2020*, pages 1314–1322. ACM, 2020.
- [ST12] Dirk Sudholt and Christian Thyssen. A simple ant colony optimizer for stochastic shortest path problems. *Algorithmica*, 64:643–672, 2012.
- [Sud13] Dirk Sudholt. A new method for lower bounds on the running time of evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 17:418–435, 2013.
- [Sud20] Dirk Sudholt. Analysing the robustness of evolutionary algorithms to noise: refined runtime bounds and an example where noise is beneficial. *Algorithmica*, 2020. To appear.
- [Wit05] Carsten Witt. Worst-case and average-case approximations by simple randomized search heuristics. In *Symposium on Theoretical Aspects of Computer Science, STACS 2005*, pages 44–56. Springer, 2005.
- [Wit13] Carsten Witt. Tight bounds on the optimization time of a randomized search heuristic on linear functions. *Combinatorics, Probability & Computing*, 22:294–318, 2013.
- [WW05] Ingo Wegener and Carsten Witt. On the optimization of monotone polynomials by simple randomized search heuristics. *Combinatorics, Probability & Computing*, 14:225–247, 2005.
- [XN13] Mingyu Xiao and Hiroshi Nagamochi. Confining sets and avoiding bottleneck cases: a simple maximum independent set algorithm in degree-3 graphs. *Theoretical Computer Science*, 469:92–104, 2013.