CrossMark

Robustness of Populations in Stochastic Environments

Christian Gießen[1] · Timo Kötzing[2]

Abstract We consider stochastic versions of ONEMAX and LEADINGONES and analyze the performance of evolutionary algorithms with and without populations on these problems. It is known that the $(1+1)$ EA on ONEMAX performs well in the presence of very small noise, but poorly for higher noise levels. We extend these results to LEADINGONES and to many different noise models, showing how the application of drift theory can significantly simplify *and generalize* previous analyses. Most surprisingly, even small populations (of size $\Theta(\log n)$) can make evolutionary algorithms perform well for high noise levels, well outside the abilities of the $(1+1)$ EA. Larger population sizes are even more beneficial; we consider both parent and offspring populations. In this sense, populations are *robust* in these stochastic settings.

Keywords Run time analysis · Stochastic fitness function · Evolutionary algorithm · Populations · Robustness

1 Introduction

Evolutionary algorithms (EAs) are general-purpose problem solvers which can be successfully applied to a wide variety of problems with small effort. In particular, EAs are popular where no tailored solutions exist, for example because the structure of the problem is inaccessible (given as a *black box*) or where the structure of the problem

---

✉ Christian Gießen
cgie@dtu.dk

1 Technical University of Denmark, 2800 Kongens Lyngby, Denmark

2 Friedrich-Schiller-Universität Jena, 07743 Jena, Germany

is very complicated. In particular, EAs are popular in settings including uncertainties, such as noisy fitness (quality) evaluations; see [1] for a survey on examples in combinatorial optimization, but also [11] for an excellent survey also discussing different sources of uncertainty.

We are interested in formally analyzing the performance of EAs in settings where the fitness function is probabilistic, i.e., a given search point can have different fitness values for different fitness evaluations. One way to deal with such uncertainty is to replace fitness evaluations with an average of a (large) sample of fitness evaluations and then proceed as if there was no noise. In this paper we are interested in a different approach where we accept the noise and try to analyze how much noise can be overcome by EAs without further modifications (note that this research can also be used to decide *how much* resampling is necessary for successful optimization). This was first done in [5], where a noisy variant of the well-known ONEMAX test function was analyzed for the simplest EA, the $(1 + 1)$ EA. In essence, it was shown that the $(1 + 1)$ EA can deal with small noise levels, but not medium noise levels. Recently, there was a sequence of papers discussing ant colony optimization for path finding problems in the presence of uncertainty [2,6,20], see also [8,9] for early work in this area.

For this paper we are exclusively concerned with optimization problems defined on bit strings of fixed length n. In this domain we have the two well-known (static) test functions ONEMAX and LEADINGONES as follows. For each bit string $x \in \{0, 1\}^n$ we let ONEMAX(x) be the number of 1s in x and LEADINGONES(x) is the number of consecutive 1s counting from the left until the first occurrence of a 0. The performance of various randomized search heuristics on these two *static* problems is known in detail.

We modify these test functions by adding noise. We distinguish between two general noise models: *prior noise* and *posterior noise*. In the first model we assume that the noise comes from not evaluating the search point in question, but a noisy variant; an example of this is the noise model used in [5], where, with probability p, a bit of the search point was flipped *before* evaluation. In the second model the fitness value of a search point obtains noise *after* evaluation. For example, one can add a value drawn from a centered normal distribution (or add a value drawn from any other chosen distribution; we call such noise *additive posterior noise*). Posterior noise is essentially the model used in [2,6,8,20].

In each case we consider the noise to be independent for different elements of the search space and for reevaluations (note that we assume that all algorithms reevaluate each search point under consideration in every iteration).

In this paper we expand on the work done in [5] in three ways. First, we make the results applicable to many different noise models; second, we analyze the LEADINGONES function in noisy settings; third, we show how the use of populations can make the EA much more robust towards noise.

Regarding the generalization, we reprove the results from [5] as a corollary to more general theorems which can be applied in many different settings. The proofs of these more general theorems rely heavily on *drift theory*, a modern tool which facilitates the formal analysis of randomized search heuristics significantly. Note that this tool was not available for [5]. Another tool suitable for the analysis of populations was recently introduced in [14] in the context of *non-elitism*, i.e. just as in the setting with noise, good solutions can get lost.

Regarding the LEADINGONES test function, we give the first formal analysis of this test function in a noisy setting.

Regarding the use of populations, the paper [18] gives a nice overview of several different aspects where populations (and the use of crossover operators) are beneficial for optimization of *static* fitness functions. In contrast to this, we show that populations can also be highly beneficial in the context of *stochastic* fitness functions, as they make much higher noise levels tractable.

1.1 Detailed Contribution

The only algorithm we consider is the $(\mu + \lambda)$ EA, for different values of μ and λ (see Sect. 2 for a detailed description). We consider the $(1 + 1)$ EA as an EA that does not rely on populations; this was the algorithm considered in [5]. Even when we discuss EAs with populations, we only consider cases with $\mu = 1$ or $\lambda = 1$, for simplicity.

We consider optimization successful in the stochastic setting as soon as the algorithm has evaluated the best static solution (in both our cases the all-1s bit string); note that the best static solution is, in all of our models, also the solution with best expected fitness. Whenever we consider the "run time" of an algorithm, this is understood as the *expected* number of *iterations* (or generations) of the EA. In particular, population-based EAs have a higher number of fitness evaluations than iterations (also due to reevaluation of old search points). Note, however, that these numbers differ only by a factor of $\mu + \lambda$. Therefore, it makes sense to take the number of iterations as a measure of the run time and the number of fitness evaluations can be directly inferred. In fact, when fitness evaluations can be done in parallel (due to sufficient hardware), the number of iterations is directly a measure for the run time.

In Sect. 3 we consider ONEMAX. In particular, we give a general theorem for deriving upper bounds in different noise settings (Theorem 5) and a general Theorem for deriving lower bounds (Theorem 6). As a result we completely reprove the theorems from [5] (Corollary 7). These results concern prior noise where, with probability p, a (uniformly chosen) bit is flipped. As a further corollary, we show that the $(1 + 1)$ EA can optimize in the presence of additive posterior noise with variance of $O(\log n/n)$ efficiently, but not, for example, in the presence of additive noise from an exponential distribution with parameter 1 (Corollary 10). We consider additive posterior noise taken from a normally distributed random variable in Corollary 11; here we can give the precise threshold below which the $(1 + 1)$ EA is successful. This list of corollaries can easily be extended, for example to cover the case of prior noise based on mutation (Corollary 8) or the noise model of "partial evaluation" considered in [4] (Corollary 9).

In Sect. 3.2 we show that populations can make an EA much more robust towards noise. For example, a linear population is large enough to allow *arbitrary* values of p in the setting of prior bit-flip noise. Furthermore, for constant p in the setting of prior bit-flip noise (a setting far outside the abilities of the $(1+1)$ EA for efficient optimization), already a *logarithmic population size* suffices for efficient optimization. Similarly, we get robustness of small populations for posterior noise models, for example for exponentially distributed noise with constant parameter.

Table 1 For different algorithms and different target problems, the range of noise parameter p is given for which optimization is successful in polynomial time (in the model of prior noise from [5], with probability p flipping a single bit)

	ONEMAX	LEADINGONES
$(1 + 1)$ EA	$O(\log n/n)$	$O(1/n^2)$
$(\mu + 1)$ EA	$[0, 1], \mu = \Omega(\log n/p)$	$O(1/n^2), \mu \geq 1$
$(1 + \lambda)$ EA	$[0, 1], \lambda = \Omega(n \log n/p)$	$O(1/n), \lambda = \Omega(\log n)$

In Sect. 4, we give our results for LEADINGONES. We show that the $(1 + 1)$ EA optimizes successfully in the presence of small noise, but we also give an example of higher noise levels where optimization is unsuccessful. Here again populations are helpful, even of logarithmic size. As expected, the LEADINGONES function is much more sensitive to prior noise than the ONEMAX function.

For the specific noise model taken from [5], we give an overview of our results in Table 1.

We conclude the paper with a discussion in Sect. 5. Note that this paper is an extended and corrected version of [7]; we added several further corollaries and expanded on many proofs. Since the publication of the conference version of this paper, another work addressing the optimization of stochastic versions of ONEMAX and LEADINGONES was published [4] where the focus was on non-elitist algorithms optimizing with noise models based on certain kinds of unavailability of data (see Corollary 9 for one of the noise models). Furthermore, several papers regarding optimization with stochastic fitness functions were presented at the FOGA'15 conference.

2 Mathematical Preliminaries

In this paper we consider the $(\mu + \lambda)$ EA, an algorithm which bases its progress on mutation (see Algorithm 1 for a detailed description). We consider only the mutation operator which flips each bit independently with probability $1/n$. Ties in the selection of fitter individuals are broken so that individuals from the offspring population are preferred (this allows the $(1 + 1)$ EA to cross plateaus and is consistent with the definition of, for example, [5]); further ties are broken uniformly at random.

Note that all references to the "run time" or the "number of steps" of an algorithm always concern the *expected first hitting time* of the optimum, as mentioned above.

2.1 Drift Theorems

We will use a variety of drift theorems to derive the theorems of this paper. *Drift*, in this context, describes the *expected change* of the best-so-far solution within one iteration with respect to some *potential*. In later proofs we will define potential functions on best-so-far solutions and prove bounds on the drift; these bounds then translate to expected run times with the use of the drift theorems from this section.

Algorithm 1: $(\mu + \lambda)$ EA

1 Let P be a set of μ uniformly chosen bit strings;
2 **repeat**
3     $O \leftarrow \emptyset$;
4   **for** $i = 1$ **to** λ **do**
5   pick x u.a.r. from P;
6         $O \leftarrow O \cup \{\texttt{mutate}(x)\}$;
7   **for** $x \in P \cup O$ **do** evaluate $f(x)$;
8     $P \leftarrow \mu$ f-maximal elements from $P \cup O$;
9 **until** *forever*;

The literature knows a large number of drift theorems; this selection is not representative, but merely contains those theorems needed for this paper.

The simplest drift theorem concerns *additive drift*.

Theorem 1 (Additive Drift [10]) *Let $(D^{(t)})_{t \geq 0}$ be random variables describing a Markov process over a finite state space $S \subseteq \mathbb{R}_0^+$. Let T be the random variable that denotes the earliest point in time $t \geq 0$ such that $D^{(t)} = 0$. If there exist $c > 0$ such that*

$$E\left(D^{(t)} - D^{(t+1)} | T > t\right) \geq c,$$

then

$$E(T | D^{(0)}) \leq \frac{D^{(0)}}{c}.$$

Another useful tool is the *Variable Drift Theorem* given in [13, Theorem 4.6] (independently developed in [15, Section 8]); this drift theorem is applicable when the drift is not uniform across the search space; frequently one can find a uniform lower bound and use the additive drift theorem, but using the variable drift theorem will typically give much better bounds. The version of the Variable Drift Theorem that we use is due to [19], which removes the restriction of h being differentiable.

Theorem 2 (Variable Drift [19]) *Let $(D^{(t)})_{t \geq 0}$ be random variables describing a Markov process over a finite state space $S \subseteq \mathbb{R}_0^+$ and let $x_{\min} := \min\{x \in S \mid x > 0\}$. Furthermore, let T be the random variable that denotes the first point in time $t \in \mathbb{N}$ for which $D^{(t)} = 0$. Suppose that there exists a monotone increasing function $h : \mathbb{R}^+ \to \mathbb{R}^+$ such that $1/h$ is integrable and*

$$E\left(D^{(t)} - D^{(t+1)} \mid D^{(t)}\right) \geq h(D^{(t)})$$

holds for all $t < T$. Then,

$$E(T \mid D^{(0)}) \leq \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{D^{(0)}} \frac{1}{h(x)} dx.$$

Since we will make frequent use of it in the following sections, we will also give the version of the *Multiplicative Drift Theorem* for upper bounds, due to [3], which is implied by the previous Variable Drift Theorem.

Theorem 3 (Multiplicative Drift [3]) *Let $(D^{(t)})_{t \geq 0}$ be random variables describing a Markov process over a finite state space $S \subseteq \mathbb{R}_0^+$ and let $x_{\min} := \min\{x \in S \mid x > 0\}$. Let T be the random variable that denotes the earliest point in time $t \geq 0$ such that $D^{(t)} = 0$. If there exist $\delta > 0$ such that for all $x \in S$ with $P(D^{(t)} = x) > 0$ we have*

$$E\left(D^{(t)} - D^{(t+1)} | D^{(t)} = x\right) \geq \delta x,$$

then for all $x' \in S$ with $P(D^{(0)} = x') > 0$,

$$E\left(T | D^{(0)} = x'\right) \leq \frac{1 + \log\left(\frac{x'}{x_{\min}}\right)}{\delta}.$$

Finally, in order to derive *lower* bounds on the run time of EAs, we use the Negative Drift Theorem.

Theorem 4 (Negative Drift [16,17]) *Let $(D^{(t)})_{t \geq 0}$ be real-valued random variables describing a stochastic process over some state space. Suppose there is an interval $[a, b] \subseteq \mathbb{R}$, two constants $\delta, \varepsilon > 0$ and, possibly depending on $\ell = b - a$, a function $r(\ell)$ satisfying $1 \leq r(\ell) = o(\ell/\log \ell)$ such that, for all $t \geq 0$, the following conditions hold.*

1. *$E(D^{(t+1)} - D^{(t)} \mid a < D^{(t)} < b) \geq \varepsilon$;*
2. *For all $j \geq 0$, $P(|D^{(t+1)} - D^{(t)}| \geq j \mid a < D^{(t)}) \leq \frac{r(\ell)}{(1+\delta)^j}$.*

Then there is a constant c such that, for $T = \min\{t \geq 0 : D^{(t)} \leq a \mid D^{(0)} \geq b\}$, we have

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

3 OneMax

In this section we present our results regarding ONEMAX. We fix a (stochastic) ONEMAX function f according to one of our models. For each of the stochastic models we consider (see Corollaries 7 through 11), it is easy to verify that there is a sequence of independent random variables $(X_k)_{k \leq n}$ such that the following holds.

– For each evaluation of f on a bit string with exactly k 1s, the return value is drawn at random $\sim X_k$ (recall that all evaluations of fitness functions are independent).
– $\forall 0 < j < k < n : P(X_k < X_{k+1}) \leq P(X_j < X_{k+1})$; intuitively, the larger the true difference in ONEMAX-value, the more likely this is reflected in a random ONEMAX evaluation. This simplifies some conditions.

Note that these two properties capture two important properties of the ONEMAX function: symmetry in the positions (only the number of 1s decides on the fitness, not the position), and monotonicity (the more 1s a bit string has, the higher its fitness; we need this comparison-based version, given the comparison-based definition of the $(1 + 1)$ EA).

We start by giving an upper and a lower bound for the $(1 + 1)$ EA in Sect. 3.1. In Sect. 3.2 we give upper bounds for EAs with parent populations and in Sect. 3.3 for EAs with offspring populations, showing that populations are efficient for the stochastic versions of ONEMAX we consider.

3.1 The $(1 + 1)$ EA on ONEMAX

Our first theorem gives an upper bound for the $(1 + 1)$ EA on ONEMAX, generalizing a theorem from [5]. The condition given by Eq. (1) intuitively says that more 1s look better with a probability which is close to 1, getting closer the more 1s there are. This increase in reliability of the evaluation is needed to counterbalance the increased likelihood of finding *worse* search points the closer the best search point is to the optimum.

Theorem 5 *Suppose there is a positive constant $c \leq 1/15$ such that*

$$\forall k < n : P(X_k < X_{k+1}) \geq 1 - c\frac{n-k}{n}. \tag{1}$$

Then the $(1+1)$ EA optimizes f in $\Theta(n \log n)$ steps. Furthermore, if Eq. (1) holds for all $k < n - \ell$ for some ℓ with $2 < \ell \leq n/2$, and we have

$$\forall k < n : P(X_k < X_{k+1}) \geq 1 - \frac{\ell}{n},$$

then $(1 + 1)$ EA optimizes f in expectation in $O(n \log(n)) + n2^{O(\ell)}$ iterations.

Proof Let $p_k = c(n - k)/n$. Let $(D^{(t)})_{t \geq 0}$ be the Markov process describing the number of 1s of the $(1 + 1)$ EA on f. We show that there is a positive multiplicative drift driving the number of 1s up to n, more precisely we will show that there is a constant z such that

$$\forall t : E\left(D^{(t+1)} - D^{(t)} \mid D^{(t)}, D^{(t)} < n\right) \geq \frac{z}{n}\left(n - D^{(t)}\right).$$

Let t be given and let $k := D^{(t)}$ be at most $n - 1$.

Let E_0 be the event that the new search point has at least one more 1 (i.e. $D^{(t+1)} \geq k + 1$) and the comparison of old and new search point indicates correctly that the new search point is better. We have that $P(E_0) \geq (1 - p_k)(n - k)/(en)$, as the probability for flipping exactly one 0-bit is at least $(n - k)/(en)$, and the probability of accepting the new (better) search point is at least $1 - p_k$.

Let E_1 be the event that the new search point has less 1s than the current search point and that this new search point is nonetheless accepted (i.e. $D^{(t+1)} \leq k-1$). We have that $P(E_1) \leq p_{k-1}k/n$, as the probability of creating a worse search point is at most k/n (one of the 1-bits has to flip) and the probability of incorrectly accepting a worse search point is at most p_{k-1}. First, we show that the *expected* number of 1s conditional on E_1 is at least $k-2$. For that notice that this expectation only decreases if we condition additionally on the event E_2 that no 0 flips to a 1. Let A be the set of indices where the current search point has a 1; without loss of generality, assume $0 \in A$. Let, for all $i \in A$, F_i be the event that bit i flipped. Furthermore, let Y be the random variable describing the number of flipping 1s. Note that $E_1 \subseteq \bigcup_{i \in A} F_i$. Now we see

$$E(Y \mid E_1, E_2) \leq \sum_{i \in A} E(Y \mid F_i, E_1, E_2) P(F_i) \leq \frac{|A|}{n} E(Y \mid F_0, E_1, E_2) \leq 2,$$

as Y conditioned on any F_i is 1 (for the ith bit flipping) plus an expected $1/n$ for every other 0-bit. This shows $E(D^{(t+1)} \mid E_1) \geq k-2$ as desired.

Taking all parts together, we have

$$
\begin{aligned}
E(D^{(t+1)} - D^{(t)} \mid D^{(t)}, D^{(t)} < n) &\geq P(E_0) - 2P(E_1) \\
&\geq (1 - p_k)\frac{n-k}{en} - 2p_{k-1}\frac{k}{n} \\
&\geq \frac{n-k}{en} - p_k\left(2 + \frac{n-k}{en}\right) - 2c/n \\
&\geq \frac{n-k}{en} - 3p_k - 2c/n \\
&\geq \frac{n-k}{n}(1/e - 5c).
\end{aligned}
$$

Using $c \leq 1/15$ we see that $1/e - 5c$ is a constant > 0. We apply the Variable Drift Theorem (see Theorem 2) on the process $(n - D^{(t)})_t$ describing the number of 0s at iteration t, $x_{\min} = 1$ and h such that, for all $x > 0$, $h(x) = ax/n$ for a positive constant $a < 1/e - 5c$. As the initial number of 0 is at most n, we get an expected run time of $O(n \log n)$.

Regarding the "furthermore" clause, we do not have sufficient drift when we use the number of 1s as the potential function. Thus, we change our potential function in a way which could also be used to show an $\exp(O(n))$ bound for optimizing the needle function with the $(1+1)$ EA. Intuitively, this drift function takes care of the plateau of the last ℓ ONEMAX values.

We define a helper function a so that, for all $i \geq 1$,

$$
a(i) = \begin{cases} 2\frac{i!}{\ell!}(10e\ell)^{\ell-i}, & \text{if } i \leq \ell; \\ 1, & \text{otherwise.} \end{cases}
$$

We now define the potential in terms of a as follows. A given search point with exactly z 0s has potential

$$g(z) = \sum_{i=1}^{z} a(i).$$

This potential function makes it particularly easy to compute the differences of the potentials of similar search points (it is a sum of successive $a(i)$). We let $(C^{(t)})_t$ be the sequence of random variables describing the number of 0s in the current bit string after t iterations. Let E_0 and E_1 be as above. For $z \geq \ell$ we see that we have a drift of $\Omega(z/n)$ from the computations above. Suppose now $z < \ell$. Later we will need that, for all $i < \ell$,

$$\frac{a(i)}{a(i+1)} = \frac{10e\ell}{i+1}$$

and, thus,

$$a(i) + 2a(i+1) \leq 2a(i). \tag{2}$$

We now estimate how much the potential can fall conditional on sampling and accepting a worse search point. We use $E(C^{(t+1)} \mid C^{(t)} = z, E_1) \leq z + 2$ as before and make the worst-case assumption that we lose the complete potential from $g(z)$ to $g(\ell)$. This gives

$$E\left(g(C^{(t)}) - g(C^{(t+1)}) \mid C^{(t)} = z, E_1\right) \geq g(z) - g(\ell) - 2$$
$$= -2 - \sum_{i=z+1}^{\ell} a(i)$$
$$\geq -2a(z+1).$$

The last step can be seen with a straightforward induction employing Eq. (2).

Conditional on making progress at all (more precisely, conditional on E_0), we gain in potential by at least

$$g(z) - g(z-1) = a(z).$$

The probability of accepting a better individual is, by assumption, at least $1 - \ell/n$. Thus, taking all parts together, we have

$$E\left(g(C^{(t)}) - g(C^{(t+1)}) \mid C^{(t)} = z\right)$$
$$\geq P(E_0)a(z) - P(E_1)2a(z+1)$$
$$\geq \left(1 - \frac{\ell}{n}\right)\frac{z}{en}a(z) - \frac{\ell}{n}\frac{n-z}{n}2a(z+1).$$

Simple arithmetic and the conditions we have on ℓ and z now show that this term is at least b/n for some constant b. It is easy to see that $g(\ell) = \exp(O(\ell))$. We now use the Variable Drift Theorem (see Theorem 2) on the process $(g(C^{(t)}))_t$. We let h be monotone and integrable such that, for all z with $0 < z \leq \ell$ we have $h(g(z)) = b/n$,

and for $z > \ell \ h(g(z)) = az/n$ for a as above. Thus we see that the total expected optimization time is at most

$$O(n \ g(\ell) + n \log(n)),$$

where the first term accounts for potential values up to $g(\ell)$ and the second term accounts for higher values, where we have the higher drift of $O(z/n)$. □

Now we come to our second theorem, a lower bound for the $(1+1)$ EA on ONEMAX.

Theorem 6 *Suppose there is $\ell \leq n/4$ and a constant $c \geq 16$ such that*

$$\forall k, n - \ell \leq k < n : P(X_k < X_{k+1}) \leq 1 - c \frac{n-k}{n}.$$

Then the $(1 + 1)$ EA optimizes f in $2^{\Omega(\ell)}$ iterations with high probability.

Proof We want to show that there is a constant negative drift on the number of ones in the interval between $n - \ell$ and n; however, we will count iterations of the process only when an actual change in the number of 1s occurs (i.e., we condition on the change), as otherwise the drift would be too small (this will only yield a smaller bound than when counting all other steps as well). See also [19] regarding an explicit negative drift theorem in the presence of self-loops. Let $k \geq n - \ell$ be the number of 1s of the current search point.

Let $p_k = c(n - k)/n$. With a probability of at least $p_{k-1}k/(en)$ we decrease the number of 1 by one (this is the chance of flipping exactly one bit, from 1 to 0, and then accepting). This decrease of at least 1 times the probability of doing so is our negative part of the drift.

Countering the negative part is a positive part. Let E_0 be the event that the new search point has at least one more 1 and the comparison of old and new search point indicates correctly that the new search point is better. Clearly, the *expected* number of 1s conditional on E_0 is at most $k + 2$ and $P(E_0) \leq (n - k)/n$. Thus, the expected increase in 1s times the probability of making such an increase is at most $2(n - k)/n$, the positive part of the drift.

Since we do not count steps without change, it suffices to show that the negative part of the drift is at least twice as large in magnitude as the positive part of the drift. This would yield a negative drift of at least 1/3 (this uses that we have a minimal step width of 1). Dividing the lower bound for the negative part of the drift by the upper bound for the positive part of the drift, we get a ratio of

$$\frac{p_{k-1}k}{en} \frac{n}{2(n-k)} \geq \frac{ck}{2en}.$$

From $k \geq 3n/4$ and $c \geq 16$ we get the desired lower bound of 2 on the ratio. As the $(1 + 1)$ EA makes long jumps with sufficiently small probability, an application of the Negative Drift Theorem (Theorem 4) concludes the proof. □

Our two theorems can be used for easy corollaries, showing the optimization time of the $(1+1)$ EA given different noise models. We first consider the noise model given in [5].

Corollary 7 ([5]) *Suppose prior noise which, with probability p, flips a bit chosen uniformly at random. Then we have that the $(1 + 1)$ EA optimizes* ONEMAX *in an expected number of iterations of*

$$
\begin{cases}
\Theta(n \log n), & \text{if } p = O(1/n); \\
\text{polynomial}, & \text{if } p = O(\log(n)/n); \\
\text{superpolynomial}, & \text{if } p = \omega(\log(n)/n) \cap 1 - \omega(\log(n)/n).
\end{cases}
$$

Proof Let c be given and suppose $p \leq c/n$; let $k < n - 2c$. We estimate $P(X_k < X_{k+1})$ by observing that the event $X_k \geq X_{k+1}$ requires the individual with k 1s to be evaluated to $k + 1$ or the other to k. Both have a probability of at most p. By the union bound, we get either of the two events with probability at most $2p \leq 2c/n \leq (n-k)/n$. Thus, we get the desired bound from Theorem 5 using the "furthermore" clause with $\ell = 2c$.

Let now c be given and suppose $p \leq c \log n/n$. We get the bound $P(X_k \geq X_{k+1}) = O(\log n/n)$ just as in the last paragraph; this is sufficient up to a distance of $O(\log n)$ from the optimum, which gives a polynomial bound using the "furthermore" clause from Theorem 5 again.

Suppose $p = \omega(\log n/n) \cap 1 - \omega(\log n/n)$. Then, for all $k \geq 3n/4$, we estimate $P(X_k = X_{k+1})$ as follows. That X_k evaluates to k has a probability of at least $(1 - p)$; that X_{k+1} evaluates to k has a probability of at least $p(k + 1)/n \geq 3p/4$. Since these two probabilities are independent, we get

$$
P(X_k = X_{k+1}) \geq (1 - p)(3p/4).
$$

Using the bounds on p, Theorem 6 now gives a superpolynomial run time.   □

Note that, for the missing case of $p = 1 - O(\log(n)/n)$ in Corollary 7 we can apply neither Theorem 5 nor Theorem 6 as the relevant constant c has a value of roughly 1 (for an upper bound we need $c \leq 1/15$, for a lower bound we need $c \geq 16$).

Just as easily we can make conclusions about different noise models. In the next corollary we consider noise as given by flipping each bit with a certain probability.

Corollary 8 *Suppose prior noise which flips* each *bit independently with probability p. Then we have that the $(1 + 1)$ EA optimizes* ONEMAX *in an expected number of iterations of*

$$
\begin{cases}
\Theta(n \log n), & \text{if } p = O(1/n^2); \\
\text{polynomial}, & \text{if } p = O(\log n/n^2); \\
\text{superpolynomial}, & \text{if } p = \omega(\log n/n^2).
\end{cases}
$$

Proof Suppose first $p \leq c/n^2$ for some constant c and let $k < n - 2c$. We estimate $P(X_k < X_{k+1})$ by observing that that the event $X_k \geq X_{k+1}$ requires the individual with k 1s to be evaluated to at least $k + 1$ or the other to at most k. Both have a probability of at most np by the union bound. Again by the union bound, we get either of the two events with probability at most $2np \leq 2c/n \leq (n-k)/n$. Thus, we get the desired bound from Theorem 5.

Let now c be given and suppose $p \leq c\log n/n^2$. We get the bound $P(X_k \geq X_{k+1}) = O(\log n/n)$ just as in the last paragraph; this is sufficient up to a distance of $O(\log n)$ from the optimum, which gives a polynomial bound using the "furthermore" clause from Theorem 5.

We show the superpolynomial run time bound by distinguishing three cases for the growth rate of p. For all $k \geq 3n/4$, we estimate $P(X_k = X_{k+1})$. We will show a bound of

$$P(X_k \geq X_{k+1}) = \omega(\log(n)/n).$$

so that Theorem 6 gives a superpolynomial run time with $\ell = \omega(\log(n))$. Fix two individuals x and y with k and $k + 1$ 1s respectively.

Suppose first $p = \omega(\log n/n^2) \cap O(1/n)$. That x evaluates to k has a probability of at least $(1 - p)^n = \Omega(1)$ (no bit flips); that y evaluates to at most k has a probability of at least $p(k + 1)/e \geq 3pn/(4e)$. Since these two probabilities are independent, we get

$$P(X_k \geq X_{k+1}) \geq \Omega(1)(3pn/(4e)) = \omega(\log(n)/n).$$

Suppose now $p = \omega(1/n) \cap O(\log(n)/n)$. Fix $\log(n)$ positions for x at which there is exactly one 0; fix $\log(n)$ positions of y where there are only 1s. The probability that, after the noise is applied, x has at most as many 1s as y *on the positions outside of the fixed ones* is at least 1/2 (due to symmetry, they have just as many 1s at these positions). The probability that of all the fixed position exactly one in y flips is at least $\log(n)p(1 - p)^{2\log(n)-1}$, as there are $\log(n)$ choices for the position in y. We can bound this probability, using Bernoulli's Inequality, as

$$\begin{aligned}
\log(n)p(1 - p)^{2\log(n)-1} &\geq \log(n)\omega(1/n)\,(1 - p(2\log(n) - 1)) \\
&\geq \omega(\log(n)/n)\left(1 - O\left(\frac{\log(n)^2}{n}\right)\right) \\
&= \omega(\log(n)/n).
\end{aligned}$$

Together with the probability of 1/2 for x having at most as many 1s on the other bits, this shows the desired bound.

Finally, suppose $p = \omega(\log(n)/n)$. Align the bit positions so that there is only one bit position where x and y differ. We first focus on the bit positions where x and y are the same. After applying the noise, the probability that y has at most as many 1s (within these bit positions) as x, is at least 1/2. The probability that, at the position

where x and y differ, y has its bit flipped, is p. Thus, the probability of y evaluating to at most as much as x is at least $p/2$, which gives again the desired bound. □

Next we consider the noise model "partial evaluation" given in [4]. Note that the case of p constant of the next corollary was shown by [4].

Corollary 9 *Suppose prior noise which sets* each *bit independently with probability p to 0. Then we have that the $(1+1)$ EA optimizes* ONEMAX *in an expected number of iterations of*

$$\begin{cases} \Theta(n \log n), & \text{if } p = O(1/n^2); \\ \text{polynomial}, & \text{if } p = O(\log n/n^2); \\ \text{superpolynomial}, & \text{if } p = \omega(\log n/n^2). \end{cases}$$

We omit the proof, since it is basically the same as for the previous corollary.

Regarding posterior noise we give the following two corollaries.

Corollary 10 *Suppose posterior noise, sampling from some distribution D with variance σ^2. Then we have that the $(1+1)$ EA optimizes* ONEMAX *in an expected polynomial time if $\sigma^2 = O(\log n/n)$. On the other hand, if, for example, D is exponentially distributed with parameter 1, then the $(1+1)$ EA optimizes* ONEMAX *in superpolynomial time only.*

Proof We have $X_k \sim k + D$ and $X_{k+1} \sim k + 1 + D$; let D' be the difference of two independent copies of D. We have

$$P(X_k < X_{k+1}) = P(0 < X_{k+1} - X_k) = P(-1 < D').$$

The variance of the difference of two random variables both with variance σ^2 is $2\sigma^2$. Now we apply Chebyshev's Inequality to see that

$$P(|D'| \geq 1) \leq 2\sigma^2.$$

Thus, for $\sigma^2 = O(\log n/n)$, we get polynomial run time using Theorem 5.

In the case of D an exponential distribution we have a constant chance of $X_k \geq k+2$ and $X_{k+1} \leq k + 2$, which leads to the claimed result using Theorem 6. □

Corollary 11 *Suppose posterior noise, sampling from a Gaussian distribution $D \sim \mathcal{N}(0, \sigma^2)$ with variance σ^2. Then we have that the $(1+1)$ EA optimizes* ONEMAX *in an expected polynomial time if $\sigma^2 \leq 1/(4 \log(n))$. On the other hand, if $\sigma^2 \geq c/(4 \log(n))$ for any $c > 1$, then the $(1+1)$ EA optimizes* ONEMAX *in superpolynomial time only.*

Proof Let D' be the difference of two independent copies of D as in the proof of Corollary 10. We have that $D' \sim \mathcal{N}(0, 2\sigma^2)$; thus

$$P\left(-1 < D'\right) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{1}{2\sigma}\right)\right) = 1 - \frac{1}{2}\text{erfc}\left(\frac{1}{2\sigma}\right).$$

We will use the following standard estimates of the complementary error function [21], somewhat simplified by making the estimate more rough. For all x,

$$\frac{\exp(-x^2)}{2x+2} \leq \mathrm{erfc}(x) \leq \frac{2\exp(-x^2)}{x} \tag{3}$$

We now first consider the positive part, so suppose $\sigma^2 \leq 1/(4\log(n))$. Using the upper bound from Eq. (3) and the fact that erfc is monotonically decreasing, we get

$$\frac{1}{2}\mathrm{erfc}\left(\frac{1}{2\sigma}\right) \leq \frac{1}{2}\mathrm{erfc}\left(\sqrt{\log n}\right) \leq \frac{\exp(-\log n)}{\sqrt{\log n}} \leq \frac{1}{n}.$$

Thus, we get polynomial run time using Theorem 5.

Regarding the negative part, let $c > 1$ and suppose $\sigma^2 \geq c/(4\log(n))$. Using the lower bound from Eq. (3) and again the fact that erfc is monotonically decreasing, we get, asymptotically for any c' with $1/c < c' < 1$,

$$\frac{1}{2}\mathrm{erfc}\left(\frac{1}{2\sigma}\right) \geq \frac{1}{2}\mathrm{erfc}\left(\sqrt{\log n/c}\right) \geq \frac{\exp(-\log n/c)}{2\sqrt{\log n/c}+2} \geq \frac{1}{n^{c'}}.$$

This leads to the claimed result using Theorem 6.   □

3.2 Parent Populations: The $(\mu + 1)$ EA on ONEMAX

In this section we give upper bounds for a EAs using parent populations, i.e. we consider the $(\mu+1)$ EA. From [22] we know that the $(\mu+1)$ EA needs $\Theta(\mu n + n\log n)$ iterations to optimize the *static* version of ONEMAX. We conjecture a similar run time for sufficiently benevolent noisy versions, but, for simplicity also for the requirements on the noise model, we only give the slightly weaker bound of $O(\mu n\log n)$. To bound the negative drift, we also require that the noise has only a very small range.

Theorem 12 *Let μ be given and suppose, for each $k \leq n$, $X_k \in [k-1, k+1]$. For each $k < n$, let A_k be the event that, when drawing μ independent copies of X_k and one copy of X_{k+1} and then sorting with breaking ties uniformly, the value of X_{k+1} does* not *come out least. If there is a positive constant $c \leq 1/15$ such that*

$$\forall k, n/4 < k < n : P(A_k) \geq 1 - c\frac{n-k}{n\mu}, \tag{4}$$

then the $(\mu + 1)$ EA optimizes f in an expected number of $O(\mu n\log n)$ iterations.

Note that the requirement of Eq. (4) might seem to get more restrictive with growing μ; however, it only gets linearly more restrictive in the fraction on the right-hand-side, while the impact of growing μ on the probability of the event A_k is typically much stronger.

Proof Let $p_k = c(n-k)/(n\mu)$.

We show that there is a positive drift on the number of ones in the current best search point. Let k be the number of 1s of the current best search point.

Let E_0 be the event that the new search point has at least one more 1 than the best one and this new search point is not removed in the selection step. We have $P(E_0) \geq (1 - p_k)(n-k)/(e\mu n)$.

Let E_1 be the event that the new search point has less 1s than the current best search point, the best search point is unique, and that this unique search point is discarded (if the best search point is not unique, E_1 is the empty event). From the bound on the range of the noise we get that the best search point with k 1s can only be discarded if all other search point have at least $k-2$ 1s. Thus, the number of 1s conditional on E_1 (if $E_1 \neq \emptyset$) is at least $k-2$. We have $P(E_1) \leq p_{k-1}$. Thus, the expected *increase* in the number of 1s is at least

$$P(E_0) - 2P(E_1) \geq (1 - p_k)\frac{n-k}{e\mu n} - 2p_{k-1}$$
$$= \frac{n-k}{e\mu n} - p_k\left(2 + \frac{n-k}{e\mu n}\right) - 2c/(n\mu)$$
$$\geq \frac{n-k}{e\mu n} - 3p_k - 2c/(n\mu).$$

Using the choice of $c \leq 1/15$, we see that we have sufficient multiplicative drift, and our claim follows from the Multiplicative Drift Theorem (see Theorem 3). □

From these theorems we can again derive many corollaries regarding concrete noise models. These includes corollaries implying an exponential speedups of populations of logarithmic size, when compared with the performance of the $(1+1)$ EA.

Corollary 13 *Suppose prior noise which, with probability p, flips a bit uniformly at random. Let $\mu \geq 12\log(15n)/p$. Then we have that the $(\mu+1)$ EA optimizes* ONEMAX *in an expected number of $O(\mu n \log n)$ iterations. In particular, for $p = 1/2$, we have that a population size of $\mu = 24\log(15n)$ suffices for an expected optimization time of $O(\mu n \log n)$.*

Proof In order to apply Theorem 12 we consider μ independent copies of X_k (called *low individuals*) and one copy of X_{k+1} (called *high individual*).

Suppose first $p \leq 1/(32n)$. Using Chernoff bounds, we can assume that at least $\mu/2$ low individuals will evaluate to k. Now the high individual can only come out last in the sorting described in Theorem 12 if it evaluates down and looses in the tie breaking against at least $\mu/2$ low individuals; this happens with a probability of at most

$$\frac{p}{\mu/2} \leq \frac{1}{16n\mu}.$$

Thus, Theorem 12 concludes the claim for this case.

Suppose now $p > 1/(32n)$. For $k \geq n/4$, the probability that none of the low individuals is evaluated to $k - 1$ is at most

$$\left(1 - \frac{p}{4}\right)^{\mu}$$

(as each individual is evaluated noisily with probability p and, using $k \geq n/4$, with probability at least $1/4$ the noise has a negative effect on the fitness).

For μ large enough (recall that μ increasing with increasing n), we have $p\mu/6 \geq \log(\mu)$, as can be seen by the following case distinction.

$$\mu \leq n^2 : \quad p\mu/6 \geq 2\log(15n) \geq 2\log(n) \geq \log(\mu);$$
$$\mu > n^2 : \quad p\mu/6 \geq \mu/(6 \cdot 32n) \geq \sqrt{\mu}/198 \geq \log(\mu).$$

Let $q = 1 - p/4$. We have

$$
\begin{aligned}
\mu &= 4\left(\mu/12 + \mu/6\right) \\
&\geq 4\left(\frac{\log(15n)}{p} + \frac{\log(\mu)}{p}\right) \\
&= 4\frac{\log(15n\mu)}{p}.
\end{aligned}
$$

We use the inequality $x \leq -\log(1 - x)$ for $x = p/4$ and get

$$\mu \geq \frac{4\log(15n\mu)}{-4\log(q)} = -\frac{\log(15n\mu)}{\log(q)}.$$

Taking all inequalities together we get that the probability that none of μ individuals with k 1s is evaluated to $k - 1$ is at most

$$\left(1 - \frac{p}{4}\right)^{\mu} = q^{\mu} \leq \frac{1}{15n\mu}.$$

Theorem 12 gives the desired result.   □

Corollary 13 requires larger μ for smaller p, while small noise (small values of p) should make optimization easier. Consider the following case as an illustrative example for why small noise might be bad for efficient optimization. Assume we have $\mu = \Theta(\log n)$ and $p = \Theta(1/\sqrt{n})$. We have that, with constant probability, all low individuals are truthfully evaluated. Furthermore, $X_{k+1} = k$ has a probability of $\Theta(1/\sqrt{n})$, assuming a large value of k. This shows that there is a chance of $\Theta(1/\sqrt{n})$ to remove the best individual from the population, making Theorem 12 inapplicable. The assumptions of Theorem 12 are very strong, even if they do not hold there might be efficient optimization, but this will likely require much more refined techniques (probably considering the diversity of the population).

3.3 Offspring Populations: The $(1 + \lambda)$ EA on ONEMAX

In this section we give upper bounds for a EAs using offspring populations, i.e. we consider the $(1 + \lambda)$ EA. From [12, Theorem 4] we know that the $(1 + \lambda)$ EA needs $O(n \log n + n\lambda)$ iterations to optimize the *static* version of ONEMAX.

Theorem 14 *Let $\lambda \geq 24 \log n$ and, for each $k < n$, let Y_k denote the maximum over λ observed values of X_k (belonging to inferior individuals) and let Z_k denote the maximum over at least $\lambda/6$ observed values of X_k (belonging to better individuals). Suppose there is $q < 1$ such that*

$$\forall k < n : P(Y_k < X_{k+1}) \geq q, \tag{5}$$

and

$$\forall k < n : P(Y_{k-1} < Z_k) \geq 1 - \frac{q}{5} \frac{\ell\lambda}{en + \ell\lambda}. \tag{6}$$

Then the $(1 + \lambda)$ EA optimizes f in $O((n \log n/\lambda + n)/q)$ iterations and needs $O((n \log n + n\lambda)/q)$ fitness evaluations.

Proof We show that there is a positive drift on the number of 1bits. Let k be the number of 1-bits of the current search point and ℓ the number of 0-bits, such that $k + \ell = n$. Let $p_k = (q/5)\ell\lambda/(en + \ell\lambda)$.

Let E_0 be the event that at least one offspring is improved by at least 1 and is correctly accepted. Let E_1 be the event that at least one offspring has less 1s than the current search point, and that it is still accepted. We have that $P(E_0)$ can be bounded below by

$$q\left(1 - \left(1 - \frac{\ell}{en}\right)^\lambda\right) \geq q\left(1 - e^{-\frac{\ell\lambda}{en}}\right) \geq q\frac{\ell\lambda}{en + \ell\lambda}.$$

In order to estimate $P(E_1)$ let E_2 be the event that more than $\lambda/6$ copied offspring are created. We have

$$P(E_1) = P(E_1|E_2)P(E_2) + P(E_1|\overline{E_2})P(\overline{E_2}).$$

Equation (6) gives us $P(E_1|E_2) \leq p_k$. By using a Chernoff bound we further get $P(\overline{E_2}) \leq e^{-\lambda/24}$. Bounding the other probabilities by 1 we obtain $P(E_1) \leq p_k + e^{-\lambda/24}$. Conditioning on E_1, the expected number of ones is at least $k-2$. Furthermore, due to our definition of p_k and $\ell = n - k$ we can show the following inequality:

$$p_k = \frac{q}{5} \frac{\ell\lambda}{en + \ell\lambda} = \frac{\frac{q}{5}}{1 + \frac{en}{(n-k)\lambda}} \geq \frac{\frac{q}{5}}{1 + \frac{en}{\lambda}} \geq \frac{\frac{q}{5}}{n\frac{(1+e)}{\lambda}} = \frac{q\lambda}{5n(1 + e)},$$

where the inequalities follow from the trivial bounds $k \leq n-1$ and $n \geq 1$, respectively. Since $\lambda \geq 24 \log n$, the last term is bounded below by $1/n$. Altogether, the expected increase in the number of 1s is

$$P(E_0) - 2P(E_1) \geq q\frac{\ell\lambda}{en + \ell\lambda} - 2\left(p_k + e^{-\frac{\lambda}{24}}\right)$$

$$\geq q\frac{\ell\lambda}{en + \ell\lambda} - 2\left(p_k + \frac{1}{n}\right)$$

$$\geq q\frac{\ell\lambda}{en + \ell\lambda} - 4p_k$$

$$= \left(q - \frac{4q}{5}\right)\frac{\ell\lambda}{en + \ell\lambda},$$

where the third inequality follows from $p_k \geq 1/n$, as shown above. Since the last term is positive, we can apply Theorem 3, which yields an upper bound of $O((n\log n/\lambda + n)/q)$ for the expected number of iterations. Since λ fitness evaluations are performed in each iteration the expected number of fitness evaluations until an optimum is found is therefore $O((n\log n + n\lambda)/q)$.   □

Again, the previous theorem allows us to easily derive corollaries in the context of noise. In the following we present two corollaries showing that offspring populations can handle much higher noise levels than the $(1 + 1)$ EA. We first consider the noise model given in [5].

Corollary 15 *Suppose prior noise which, with probability $p > 0$, flips a bit uniformly at random. Then, for $\lambda \geq \max\{12/p, 24\}n\log n$, the $(1 + \lambda)$ EA optimizes* ONEMAX *in time $O((n^2\log n/\lambda + n^2)/p)$, i.e. the $(1 + \lambda)$ EA needs $O((n^2\log n + n^2\lambda)/p)$ fitness evaluations until an optimum is found.*

Proof Let $c = \max\{12/p, 24\}$. In order to apply Theorem 14 we need to check Eqs. (5) and (6). Let $q = p/n$.

We show Eq. (5) by considering the event that the good individual evaluates better than it is. Note that then, any noise affecting the worse individuals does not harm the comparison since misevaluations can only differ from the true fitness by 1. We get

$$P(Y_k < X_{k+1}) \geq 1 - \left(1 - p + \frac{pk}{n}\right) \geq \frac{p}{n} = q.$$

Reusing our above considerations regarding the probability that at least one of $\lambda/6$ individuals improves, and due to our choice of $\lambda \geq cn\log n$ we have

$$P(Y_{k-1} < Z_k) \geq 1 - \left(1 - p + \frac{pk}{n}\right)^{\lambda/6}$$

$$\geq 1 - \frac{1}{e^{\frac{p\lambda}{6n}}}$$

$$\geq 1 - \frac{1}{n^2}$$

$$\geq 1 - \frac{\frac{q}{5}}{1 + \frac{en}{(n-k)\lambda}},$$

thus fulfilling Eq. (6). The claim now follows from Theorem 14.   □

The last corollary shows that offspring populations permit the optimization process to cope with a much higher level of noise than the $(1 + 1)$ EA (see Corollary 7).

For the posterior noise model, we can show an even greater improvement over the $(1 + 1)$ EA, similar to the result regarding parent populations.

Corollary 16 *Let any non-positive additive posterior noise be given which has a non-zero probability p of evaluating to > -1. Then we have that for $\lambda \geq$ $\max\{10e, -(6\log(n/p)/(\log(1-p)))\}$ (implying $\lambda = \Omega(\log(n/p)/p)$), the $(1+\lambda)$ EA optimizes* ONEMAX *in time $O((n\log n/\lambda + n)/p)$.*

Proof Let D be the posterior noise such that $D \leq 0$ and $P(D > -1) = p$. Furthermore, let

$$\lambda \geq \max\left\{10e, \frac{-6\log\left(\frac{n}{p}\right)}{\log(1-p)}\right\}.$$

Note that since $D \leq 0$ we have that $Y_k \leq k$ which gives us

$$P(Y_k \geq X_{k+1}) \leq P(k \geq X_{k+1}) = P(D \leq -1) = 1 - p,$$

which yields $P(Y_k < X_{k+1}) \geq p$, fulfilling the first condition of Theorem 14 with $q = p$.

In order to show Eq. (6) we use the same bound as before. Note that individuals can not me misevaluated better than they are since we only consider negative, additive posterior noise. Hence, $Y_{k-1} \geq Z_k$ implies that all of at least $\lambda/6$ individuals at fitness k need to evaluate to $\leq k - 1$, i.e. $P(Y_{k-1} \geq Z_k) \leq (1-p)^{\frac{\lambda}{6}}$. We have Furthermore, due to our choice of λ, we have $(1-p)^{\lambda/6} \leq pn^{-1}$. Thus, using $\lambda \geq 10e$ from the above definition, we have

$$P(Y_{k-1} < Z_k) \geq 1 - \frac{p}{n} \geq 1 - \frac{\frac{p}{5}\lambda}{2en} \geq 1 - \frac{\frac{p}{5}}{1 + \frac{en}{(n-k)\lambda}},$$

fulfilling the second condition, and the claim follows from Theorem 14.   □

Similar to Corollary 12 the last corollary applies, for example, to additive posterior noise taken from $-\mathrm{Exp}(1)$ (where $\mathrm{Exp}(\beta)$ denotes the exponential distribution with parameter β).

4 LeadingOnes

We follow up on the section about ONEMAX with results for LEADINGONES. For this purpose we now fix a stochastic LEADINGONES function f according to one of our models. For each k, we let x_k^{opt} be the bit string which has only 1s, except for position $k + 1$; let x_k^{pes} be the bit string with k leading ones and otherwise only 0s. In a sense,

x_k^{opt} is *optimal* for a bit string with a leading ones value of k, while x_k^{pes} is pessimal. We let $(X_k^{\text{opt}})_{k \leq n}$ and $(X_k^{\text{pes}})_{k \leq n}$ be two sequences of independent random variables such that, for all $k \leq n$, $X_k^{\text{opt}} \sim f(x_k^{\text{opt}})$ and $X_k^{\text{pes}} \sim f(x_k^{\text{pes}})$. We will assume the following about f.

- For each evaluation of f on a bit string with the leftmost zero at position $k + 1$, the return value is drawn according to a distribution which is in between X_k^{pes} and X_k^{opt} with respect to stochastic dominance.
- $\forall j \leq k < n : P(X_j^{\text{opt}} < X_{k+1}^{\text{opt}}) \geq P(X_k^{\text{opt}} < X_{k+1}^{\text{opt}})$.
- $\forall j \leq k < n : P(X_j^{\text{pes}} < X_{k+1}^{\text{pes}}) \geq P(X_k^{\text{pes}} < X_{k+1}^{\text{pes}})$.

We show that, despite the more drastic consequences of noise, we still find sufficient conditions for efficient optimization similar to the ones we have already seen in Sect. 3.

We begin by giving upper and lower bounds for the $(1 + 1)$ EA in Sect. 4.1. In Sects. 4.2 and 4.3 we show the effectiveness of parent and offspring populations, respectively, for the stochastic LEADINGONES problem by giving upper bounds.

4.1 The $(1 + 1)$ EA on LEADINGONES

Similar to the analysis of ONEMAX subject to noise, we start by giving an upper bound for the $(1 + 1)$ EA. As we see in the next theorem, the condition for LEADINGONES is more strict than in the case of ONEMAX: the reliability of correct evaluations is required to be $1 - O(1/n^2)$ close to the optimum (see Eq. (7)).

Theorem 17 *Suppose there is a positive constant $c < 1/(6e)$ such that*

$$\forall k < n : P(X_k^{\text{opt}} < X_{k+1}^{\text{pes}}) \geq 1 - \frac{c}{kn} \tag{7}$$

Then the $(1 + 1)$ EA optimizes f in $O(n^2)$ iterations in expectation.

Proof Let $p_k = c/(kn)$. We show that there is a positive drift on the number of leading 1-bits. Let k be the length of the prefix of the current search point consisting of 1s.

We stick to our previous notation and denote by E_0 the event that the new search point has a longer prefix consisting of 1s and the comparison of old and new search point indicates correctly that the new search point is better. Let E_1 be the event that the new search point has a smaller number of leading ones than the current search point, and that it is accepted. Conditioning on E_1 we can trivially bound the expected number of leading ones below by 0. We have $P(E_0) \geq (1 - p_k)/(en)$ and $P(E_1) \leq p_{k-1}(1 - e^{-1}) \leq p_{k-1}$. Therefore, the expected increase in the number of leading 1s is

$$P(E_0) - kP(E_1) \geq \frac{(1 - p_k)}{en} - kp_{k-1}$$
$$= \frac{1 - \frac{c}{kn}}{en} - \frac{ck}{(k-1)n}$$
$$\geq \frac{1}{en} - \frac{c}{ekn^2} - \frac{2c}{n}$$
$$\geq \frac{1}{en} - \frac{c}{n}\left(\frac{1}{e} + 2\right)$$
$$\geq \frac{1}{en} - \frac{3c}{n}.$$

Due to our choice of $c < 1/(6e)$ we have a positive additive drift of $1/(2en)$ leading to an upper bound of $O(n^2)$ for the expected run time of the algorithm by applying the Additive Drift Theorem (see Theorem 1). □

It is only to be expected that noise disrupts the optimization of LEADINGONES immensely. Consequently, our following corollaries to Theorem 17 are rather weak with respect to the noise allowed (basically, the algorithm will only experience constantly many incorrect decisions during optimization, in expectation). First, we consider prior noise where one bit is flipped uniformly at random before evaluation.

Corollary 18 *Suppose prior noise which, with probability p, flips a bit uniformly at random. Then we have that in expectation the $(1 + 1)$ EA optimizes* LEADINGONES *in $O(n^2)$ iterations if $p \leq 1/(6en^2)$.*

Proof Suppose first $p \leq 1/(6en^2)$ and let $k < n$. We estimate $P(X_k^{\text{opt}} < X_{k+1}^{\text{pes}})$ by observing that the event $X_k^{\text{opt}} \geq X_{k+1}^{\text{pes}}$ requires the individual with k 1s to be evaluated to $\geq k + 1$ or the other to $\leq k$. The first option has a probability of at most p/n, the second of pk/n. Hence, $P(X_k^{\text{opt}} < X_{k+1}^{\text{pes}}) \geq 1 - (k+1)/(6en^3) \geq 1 - 1/(6en^2)$ and the desired bound follows from Theorem 17. □

Regarding posterior noise we give the following corollary.

Corollary 19 *Suppose posterior noise, sampling from some distribution D with variance σ^2. Then we have that the $(1+1)$ EA optimizes* LEADINGONES *in $O(n^2)$ iterations in expectation if $\sigma^2 \leq 1/(12en^2)$.*

Proof Note that, in this case, $X_k^{\text{opt}} \sim X_k^{\text{pes}}$, for all $k \leq n$. With the same argument as in Corollary 10 we have $P(X_k^{\text{opt}} < X_{k+1}^{\text{pes}}) \geq 1 - 2\sigma^2$. Thus, for $\sigma^2 = 1/(12en^2)$ the claim follows from Theorem 17. □

Next we give a lower bound for the $(1 + 1)$ EA for the prior noise model. We will not give a general lower bound that holds for both of our models because it is very easy for the $(1 + 1)$ EA to detect an inferior noisy offspring by selection if LEADINGONES is subjected to posterior noise.

Theorem 20 *Suppose prior noise which, with probability 1/2, flips a bit uniformly at random. Then we have that the $(1 + 1)$ EA optimizes* LEADINGONES *in $2^{\Omega(n)}$ iterations in expectation.*

Proof It is sufficient to show that there is a constant negative drift on the number of ones in the interval between $99n/100$ and n. Let $k \geq 99n/100$ be the number of 1s of the current search point. Note that the number of 1s k is a trivial upper bound for the number of leading ones.

Let E_0 be the event that the new search point has at least one more 1 than the current search point and the comparison of old and new search point indicates that the new search point is to be accepted. The expected number of 1s conditional on E_0 is at most $k + 2$ and we have a trivial bound of $P(E_0) \leq (n - k)/n \leq 1/100$.

Let E_1 be the event that the new search point differs from the old by flipping exactly one 1 in the right half of positions, and that it is accepted. We want to estimate $P(E_1)$. There are at least $49n/100$ 1s in the right half of positions, so the probability of flipping exactly one of them and no other is at least $49/(100e)$. In order to estimate the probability of accepting such an offspring, we consider two cases. First, assume that the parent has a leading ones value of at least $n/2$. Then the probability of the noisy evaluation evaluating the parent to a value $<n/2$ is at least $1/4$ (by choosing to flip a bit in the left half in the evaluation), while evaluating the offspring to its true value $\geq n/2$ has a probability of at least $1/2$. In total we have $P(E_1) \geq 49/(800e)$ in this case.

Second, assume that the parent has a leading ones value of $<n/2$. Then both parent and offspring have the same leading ones value; with probability $1/4$ they both evaluate to their true value, which favors the offspring. Thus, in this case, we get $P(E_1) \geq 49/(400e)$. Overall we have now $P(E_1) \geq 49/(800e) > 1/50$

Thus, we have that the total (negative) drift of

$$P(E_1) - 2P(E_0) \geq 49/(800e) - 2/100$$

which gives us a constant negative drift. Since long jumps are sufficiently small (due to our choice of using the number of 1s as potential), we can apply Theorem 4 which yields our result. □

4.2 Parent Populations: The $(\mu + 1)$ EA on LEADINGONES

In this section we give upper bounds for EAs using parent populations, i.e. we consider the $(\mu + 1)$ EA. As for the $(\mu + 1)$ EA on ONEMAX, we again only give the weaker bound of $O(\mu n^2)$ for the sake of simplicity instead of $\Theta(\mu n \log n + n^2)$ which is the bound for the static version of LEADINGONES ([22]). Note that the proofs for the upper bounds for the $(1 + 1)$ EA carry over to the case of the $(\mu + 1)$ EA.

Theorem 21 *Let μ be given and, for each $k < n$, let Y_k denote the minimum over μ observed values of X_k^{opt}. If there is a positive constant $c < 1/(6e)$ such that*

$$\forall k : n/4 < k < n \Rightarrow P(Y_k < X_{k+1}^{\mathrm{pes}}) \geq 1 - \frac{c}{\mu k n}, \qquad (8)$$

then the $(\mu + 1)$ EA optimizes f in $O(\mu n^2)$ iterations in expectation.

Note that, due to the dependence of Y_k on μ, Eq. (8) typically gets less restrictive with growing μ, just as in Theorem 12.

Proof Let $p_k = c/(\mu k n)$. We show that there is a positive drift on the number of leading 1-bits of a current best individual. Let k be the length of the prefix of the current search point consisting of 1s. Let E_0 be the event that a best individual is improved by at least 1 and accepted by mutation. Let E_1 be the event that the new individual has a smaller prefix consisting of 1s and that the unique best individual is dropped from the parent population. We have that $P(E_0) \geq (1 - p_k)/(e\mu n)$, assuming pessimistically that there is only one best individual. On the other hand $P(E_1) \leq p_{k-1}$.

Since, conditioned on E_1, the expected number of leading ones of a best individual is trivially at least 0 we can bound the expected increase in the number of leading 1s of a best individual below by

$$
\begin{aligned}
P(E_0) - k P(E_1) &\geq (1 - p_k)\frac{1}{e\mu n} - k p_{k-1} \\
&= \frac{1}{e\mu n} - \frac{c}{e\mu^2 k n^2} - \frac{2c}{\mu n} \\
&\geq \frac{1}{e\mu n} - \frac{c}{\mu n}\left(\frac{1}{e} + 2\right) \\
&\geq \frac{1}{e\mu n} - \frac{3c}{\mu n}
\end{aligned}
$$

and the last term can be bounded below by $1/(2e\mu n)$ due to our choice of $c < 1/(6e)$. Applying Theorem 1 yields an upper bound of $2e\mu n^2$ for the expected number of iterations until the optimum is found. Taking the cost of initialization into account we have an expected run time of $\mu + 2e\mu n^2$, proving our claim.   □

Theorem 21 is not strong enough to derive a good upper bound for the prior noise model where a bit flip is performed with certain probability. Regarding posterior noise, we can still derive the following corollary.

Corollary 22 *Let any non-negative additive posterior noise be given which has a non-zero constant probability of evaluating to <1. Then there is a constant c such that, for $n/(8e) \geq \mu \geq c \log n$, the $(\mu + 1)$ EA optimizes* LEADINGONES *in time $O(\mu n^2)$.*

Proof Let D be the posterior noise and $p = P(D < 1)$ a non-zero constant. Let $c = -3/\log(1 - p)$ and $\lfloor n/(8e) \rfloor \geq \mu \geq \lceil c \log n \rceil$. We have

$$
P(Y_k < X_{k+1}) = 1 - P(D \geq 1)^\mu \geq 1 - (1 - p)^{-3 \log n / \log(1-p)},
$$

and the last term equals $1 - n^{-3}$. Using $\mu \leq n/(8e)$, we have

$$
1 - \frac{1}{n^3} \geq 1 - \frac{1/(8e)}{\mu n^2} \geq 1 - \frac{1/(8e)}{\mu k n},
$$

for all $1 \leq k \leq n$. Since $1/(8e) < 1/(6e)$ holds, we can apply Theorem 21, which yields the result.   □

4.3 Offspring Populations: The $(1 + \lambda)$ EA on LEADINGONES

In this section we consider the $(1 + \lambda)$ EA. The next theorem gives conditions for efficient optimization of stochastic LEADINGONES problems. Note that Theorem 14 establishes a similar result for ONEMAX. The bound we give is the same as is shown for static LEADINGONES in [12].

Theorem 23 *Let $\lambda \geq 72 \log n$ and, for each $k < n$, let Y_k denote the maximum over λ observed values of X_k^{opt} (belonging to inferior individuals) and let Z_k denote the maximum over at least $\lambda/6$ observed values of X_k^{pes} (belonging to better individuals). Suppose there is $q < 1$ with $q \geq 6e/n$ such that*

$$\forall k < n : P(Y_k < X_{k+1}^{\mathrm{pes}}) \geq q, \tag{9}$$

and

$$\forall k < n : P(Y_{k-1} < Z_k) \geq 1 - \frac{\frac{q}{3}}{k(1 + \frac{en}{\lambda})}. \tag{10}$$

Then the $(1 + \lambda)$ EA optimizes f in expectation in $O((n + n^2/\lambda)/q)$ iterations and needs $O((n^2 + \lambda n)/q)$ fitness evaluations.

The proof of Theorem 23 is essentially identical to the proof of Theorem 14. We therefore omit it here.

We give the following corollary showing that even small offspring populations can lead to an improvement over the $(1 + 1)$ EA.

Corollary 24 *Let $p \leq c/n$ for some constant $c < 1/(12e + 2)$. Suppose prior noise which, with probability p, flips a bit uniformly at random. Then, for $\lambda \geq 72 \log n$ with $\lambda = o(n)$, the $(1 + \lambda)$ EA optimizes LEADINGONES in expectation in $O(\lambda n + n^2)$ iterations.*

Proof In order to apply Theorem 23 we need to show Eqs. (9) and (10). We can show the former considering the event that none of λ inferior individuals is evaluated better than it is and that the good individual is evaluated to its true value. Let $k < n$, then we get

$$P(Y_k < X_{k+1}^{\mathrm{pes}}) \geq \left(1 - \frac{p}{n}\right)^\lambda (1 - p) \geq (1 - p)^2 \geq 1 - 2c.$$

Thus, Eq. (9) holds for $q = 1 - 2c$, since c is a constant.

Regarding Eq. (10) we have that the probability that at least one of at least $\lambda/6$ good individuals is not evaluated worse is $1 - (pk/n)^{\lambda/6}$. Reusing our above considerations regarding the probability that inferior individuals do not improve, we have

$$P(Y_{k-1} < Z_k) \geq \left(1 - \frac{p}{n}\right)^{\lambda} \left(1 - \left(\frac{pk}{n}\right)^{\frac{\lambda}{6}}\right)$$

$$\geq \left(1 - \frac{c\lambda}{n^2}\right) \left(1 - \left(\frac{ck}{n^2}\right)^{\frac{\lambda}{6}}\right)$$

$$\geq \left(1 - \frac{c\lambda}{n^2}\right)^2$$

$$\geq 1 - \frac{2c\lambda}{n^2}.$$

Due to our choice of $c < 1/(12e + 2)$ we have $2c < (1 - 2c)/6e$, i.e. $2c < q/(6e)$. Hence, we can further bound the last term and get

$$1 - \frac{2c\lambda}{n^2} > 1 - \frac{\frac{q}{6e}}{\frac{n^2}{\lambda}} = 1 - \frac{\frac{q}{3}}{2\frac{en^2}{\lambda}} \geq 1 - \frac{\frac{q}{3}}{n + \frac{en^2}{\lambda}} \geq 1 - \frac{\frac{q}{3}}{k + \frac{ekn}{\lambda}}.$$

Thus, Eq. (10) holds and the claim follows from Theorem 23. □

We follow up by showing an even greater improvement over the $(1 + 1)$ EA for the posterior noise model. In this case

Corollary 25 *Let any non-positive, additive posterior noise be given which has a non-zero constant probability p of evaluating to > -1. Then we have that for $en \geq \lambda \geq \max\{6e, -12\log(n/p)/\log(1 - p)\}$ (implying $\lambda = \Omega\left(\log(n/p)/p\right) \cap O(n)$) the $(1 + \lambda)$ EA optimizes* LEADINGONES *in expectation in $O(n + n^2/\lambda)$ iterations.*

Proof Let D be the posterior noise, such that $D \leq 0$ and $P(D > -1) = p$. Furthermore, let

$$en \geq \lambda \geq \max \left\{6e, \frac{-12\log\left(\frac{n}{p}\right)}{\log(1 - p)}\right\}.$$

Note that since $D \leq 0$ we have that $Y_k \leq k$ which gives us

$$P(Y_k \geq X_{k+1}) \leq P(k \geq X_{k+1}) = P(D \leq -1) = 1 - p,$$

which yields $P(Y_k < X_{k+1}) \geq p$, fulfilling the first condition of Theorem 23 with $q = p$.

In order to check Eq. (10) we use the same bound as before. Note that individuals can not me misevaluated better than they are since we only consider negative, additive posterior noise. Hence, $Y_{k-1} \geq Z_k$ implies that all of at least $\lambda/6$ individuals at fitness k need to evaluate to $\leq k - 1$, i.e. $P(Y_{k-1} \geq Z_k) \leq (1 - p)^{\frac{\lambda}{6}}$. Due to our choice of λ, we have $(1 - p)^{\lambda/6} \leq pn^{-2}$. We can estimate

$$P(Y_{k-1} < Z_k) \geq 1 - \frac{p}{n^2} \geq 1 - \frac{\frac{p}{3}\lambda}{2en^2} \geq 1 - \frac{\frac{p}{3}}{n\left(1 + \frac{en}{\lambda}\right)} \geq 1 - \frac{\frac{p}{3}}{k(1 + \frac{en}{\lambda})},$$

where the second inequality follows from $\lambda \geq 6e$ and the third inequality stems from our assumption of $\lambda \leq en$. Thus, the second condition is fulfilled, and the claim follows from Theorem 23.   □

Just as for Corollary 12 the last corollary applies, for example, to additive posterior noise taken from $-\mathrm{Exp}(1)$.

5 Conclusion

In this paper we consider the optimization of noisy versions of ONEMAX and LEADIN-GONES. The summary of the results is that populations are *necessary* for successful optimization for any substantial noise levels. The surprising result is that even very small populations (of size logarithmic in the problem size) already lead to very high robustness to noise (see Corollaries 13 and 24).

From the formal analysis we see the reason for this robustness: In a (parent) population of size μ, while the best individual might *look* bad in a given iteration, there will surely be (objectively) worse individuals which also *look* worse. This holds as long as there are enough individuals in the parent populations to make sure that one of them will evaluate worse than the (objectively) best individual, and while the noise can never be very disruptive. For example, if a non-best individual will evaluate worse than the best individual with constant probability, a logarithmic number of non-best individuals is large enough to get very high confidence that such a bad individual is dropped. This observation probably extends to the analysis of the $(\mu + \lambda)$ EA with $\mu > 1$ *and* $\lambda > 1$; here the same approach of computing the drift could lead to further results.

As for offspring populations, in the $(1 + \lambda)$ EA the current individual is cloned multiple times and thus hedges against bad evaluations (as long as good evaluations are sufficiently likely). This does not extend to the analysis of the $(\mu + \lambda)$ EA with $\mu > 1$ and $\lambda > 1$ in a straightforward way, as clones might be made from sub-optimal individuals.

All these analyses would probably carry over easily to different mutation probabilities: changing them by a constant factor should only change some constants of the theorems. An entirely different question is about what happens if not all old search points are reevaluated, as for example in [20]. For high noise values, this will likely disrupt optimization heavily.

While we get very accurate statements about the capabilities of the $(1 + 1)$ EA in stochastic environments (including upper and lower bounds), there is still a lot to learn about the capabilities of population-based EAs. On the one hand we do not get any lower bounds, on the other hand we believe that the positive results can be extended. We believe that such extensions will require more refined techniques: all our bounds are based on worst-case assumptions about the distribution of the search points (for example in Theorem 12 we assume that we have only one good individual and all other individuals are worse by exactly 1). We think that the results of the $(1 + \lambda)$ EA might be easier to extend than the results for the $(\mu + 1)$ EA, as the former has a much easier to control state from one iteration to the next (only the best individual survives).

Finally, it would be very interesting to see analyses of other algorithms on stochastic fitness functions. Especially interesting would be an analysis of crossover, or any kind of estimation of distribution algorithm, including ant colony optimization algorithms.

References

1. Bianchi, L., Dorigo, M., Gambardella, L.M., Gutjahr, W.J.: A survey on metaheuristics for stochastic combinatorial optimization. Nat. Comput. **8**(2), 239–287 (2009)
2. Doerr, B., Hota, A., Kötzing, T.: Ants easily solve stochastic shortest path problems. In: Genetic and Evolutionary Computation Conference, GECCO '12, Philadelphia, PA, USA, July 7–11, 2012, pp. 17–24 (2012)
3. Doerr, B., Johannsen, D., Winzen, C.: Multiplicative drift analysis. Algorithmica **64**(4), 673–697 (2012)
4. Dang, D.-C., Lehre, P.K.: Evolution under partial information. In: Genetic and Evolutionary Computation Conference, GECCO '14, Vancouver, BC, Canada, July 12–16, 2014, pp. 1359–1366 (2014)
5. Droste, S.: Analysis of the (1+1) EA for a noisy onemax. In: Genetic and Evolutionary Computation—GECCO 2004, Genetic and Evolutionary Computation Conference, Seattle, WA, USA, June 26–30, 2004, Proceedings, Part I, pp. 1088–1099 (2004)
6. Feldmann, M., Kötzing, T.: Optimizing expected path lengths with ant colony optimization using fitness proportional update. In: Foundations of Genetic Algorithms XII, FOGA '13, Adelaide, SA, Australia, January 16–20, 2013, pp. 65–74 (2013)
7. Gießen, C., Kötzing, T.: Robustness of populations in stochastic environments. In: Genetic and Evolutionary Computation Conference, GECCO '14, Vancouver, BC, Canada, July 12–16, 2014, pp. 1383–1390 (2014)
8. Gutjahr, W.J., Pflug, GCh.: Simulated annealing for noisy cost functions. J. Glob. Optim. **8**(1), 1–13 (1996)
9. Gutjahr, W.J.: A converging ACO algorithm for stochastic combinatorial optimization. In: Stochastic Algorithms: Foundations and Applications, Second International Symposium, SAGA 2003, Hatfield, UK, September 22–23, 2003, Proceedings, pp. 10–25 (2003)
10. He, J., Yao, X.: A study of drift analysis for estimating computation time of evolutionary algorithms. Nat. Comput. **3**(1), 21–35 (2004)
11. Jin, Y., Branke, J.: Evolutionary optimization in uncertain environments—a survey. IEEE Trans. Evol. Comput. **9**(3), 303–317 (2005)
12. Jansen, T., De Jong, K.A., Wegener, I.: On the choice of the offspring population size in evolutionary algorithms. Evol. Comput. **13**(4), 413–440 (2005)
13. Johannsen, D.: Random combinatorial structures and randomized search heuristics. PhD thesis, Saarland University (2010)
14. Lehre, P.K.: Fitness-levels for non-elitist populations. In: 13th Annual Genetic and Evolutionary Computation Conference, GECCO 2011, Proceedings, Dublin, Ireland, July 12–16, 2011, pp. 2075–2082 (2011)
15. Mitavskiy, B., Rowe, J.E., Cannings, C.: Preliminary theoretical analysis of a local search algorithm to optimize network communication subject to preserving the total number of links. In: Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2008, June 1–6, 2008, Hong Kong, China, pp. 1484–1491 (2008)
16. Oliveto, P.S., Witt, C.: Simplified drift analysis for proving lower bounds in evolutionary computation. Algorithmica **59**(3), 369–386 (2011). Kindly check whether the references [16] and [17] are correct
17. Oliveto, P.S., Witt, C.: Erratum: Simplified Drift Analysis for Proving Lower Bounds in Evolutionary Computation. arXiv:1211.7184 (2012)

18. Prügel-Bennett, A.: Benefits of a population: five mechanisms that advantage population-based algo-rithms. IEEE Trans. Evol. Comput. **14**(4), 500–517 (2010)
19. Rowe, J.E., Sudholt, D.: The choice of the offspring population size in the $(1, \lambda)$ evolutionary algorithm. Theor. Comput. Sci. **545**, 20–38 (2014)
20. Sudholt, D., Thyssen, C.: A simple ant colony optimizer for stochastic shortest path problems. Algo-rithmica **64**(4), 643–672 (2012)
21. Weisstein, E.W.: Erfc, 2015. From MathWorld—A Wolfram Web Resource. http://mathworld.wolfram.com/Erfc.html
22. Witt, C.: Runtime analysis of the $(\mu + 1)$ EA on simple pseudo-boolean functions. Evol. Comput. **14**(1), 65–86 (2006)