

# Running Time Analysis of the (1+1)-EA for OneMax and LeadingOnes under Bit-wise Noise

Chao Qian<sup>1</sup> · Chao Bian<sup>1</sup> · Wu Jiang<sup>1</sup> · Ke Tang<sup>2\*</sup>

Received: date / Accepted: date

**Abstract** In many real-world optimization problems, the objective function evaluation is subject to noise, and we cannot obtain the exact objective value. Evolutionary algorithms (EAs), a type of general-purpose randomized optimization algorithm, have been shown to be able to solve noisy optimization problems well. However, previous theoretical analyses of EAs mainly focused on noise-free optimization, which makes the theoretical understanding largely insufficient for the noisy case. Meanwhile, the few existing theoretical studies under noise often considered the one-bit noise model, which flips a randomly chosen bit of a solution before evaluation; while in many realistic applications, several bits of a solution can be changed simultaneously. In this paper, we study a natural extension of one-bit noise, the bit-wise noise model, which independently flips each bit of a solution with some probability. We analyze the running time of the (1+1)-EA solving OneMax and LeadingOnes under bit-wise noise for the first time, and derive the ranges of the noise level for polynomial and super-polynomial running time bounds. The analysis on LeadingOnes under bit-wise noise can be easily transferred to one-bit noise, and improves the previously known results. Since our analysis discloses that the (1+1)-EA can be efficient only under low noise levels, we also study whether the sampling strategy can bring robustness to noise. We prove that using sampling can significantly increase the largest noise level allowing a polynomial running time, that is, sampling is robust to noise.

---

A preliminary version of this paper has appeared at GECCO'17 [27].

<sup>1</sup> Anhui Province Key Lab of Big Data Analysis and Application, School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China

<sup>2</sup> Shenzhen Key Lab of Computational Intelligence, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

\* Corresponding author, tangk3@sustc.edu.cn

**Keywords** Noisy optimization · evolutionary algorithms · sampling · running time analysis · computational complexity

## 1 Introduction

In real-world optimization tasks, the exact objective (i.e., fitness) function evaluation of candidate solutions is often impossible, instead we can obtain only a noisy one due to a wide range of uncertainties [21]. For example, in machine learning, a prediction model is evaluated only on a limited amount of data, which makes the estimated performance deviate from the true performance; in product design, the design variables can be subject to perturbations due to manufacturing tolerances, which brings noise.

In the presence of noise, the difficulty of solving an optimization problem may increase. Evolutionary algorithms (EAs) [5], inspired by natural phenomena, are a type of randomized metaheuristic optimization algorithm. They are likely to be able to handle noise, since the corresponding natural phenomena have been well processed in noisy natural environments. In fact, EAs have been successfully applied to solve many noisy optimization problems [7, 22].

Compared with the application, the theoretical analysis of EAs is far behind. But in the last two decades, much effort has been devoted to the running time analysis (an essential theoretical aspect) of EAs. Numerous analytical results for EAs solving synthetic problems as well as combinatorial problems have been derived, e.g., [4, 23]. Meanwhile, a few general approaches for running time analysis have been proposed, e.g., drift analysis [10, 12, 20], fitness-level methods [8, 32], and switch analysis [34].

However, previous running time analyses of EAs mainly focused on noise-free optimization, where the fitness evaluation is exact. Only a few pieces of work on noisy evolutionary optimization have been reported, which mainly considered two kinds of noise models, prior and posterior. The prior noise comes from the variation on a solution, e.g., one-bit noise [13] flips a random bit of a binary solution before evaluation with probability  $p$ . The posterior noise comes from the variation on the fitness of a solution, e.g., additive noise [19] adds a value randomly drawn from some distribution. Droste [13] first analyzed the  $(1+1)$ -EA on the OneMax problem in the presence of one-bit noise and showed that the tight range of the noise probability  $p$  allowing a polynomial running time is  $O(\log n/n)$ , where  $n$  is the problem size. Gießen and Kötzing [19] recently studied the LeadingOnes problem, and proved that the expected running time is polynomial if  $p \leq 1/(6en^2)$  and exponential if  $p = 1/2$ . They also considered additive noise with variance  $\sigma^2$ , and proved that the  $(1+1)$ -EA can solve OneMax and LeadingOnes in polynomial time when  $\sigma^2 = O(\log n/n)$  and  $\sigma^2 \leq 1/(12en^2)$ , respectively.

For inefficient optimization of the  $(1+1)$ -EA under high noise levels, some implicit mechanisms of EAs were proved to be robust to noise. In [19], it was shown that the  $(\mu+1)$ -EA with a small population of size  $\Theta(\log n)$  can solve

**Table 1** The noise models mainly considered in running time analyses on noisy evolutionary optimization.

References	Noise models
Droste [13], Qian et al. [29]	one-bit noise
Akimoto et al. [2]	additive noise
Prugel-Bennett et al. [26], Friedrich et al. [16, 17, 18]	additive Gaussian noise
Dang and Lehre [9], Gießen and Kötzing [19]	one-bit noise, additive noise
Qian et al. [28]	one-bit noise, additive Gaussian noise
Doerr et al. [11], Feldmann and Kötzing [15], Sudholt and Thyssen [33]	the single-destination shortest path problem with stochastic edge weights
This paper	bit-wise noise

OneMax in polynomial time even if the probability of one-bit noise reaches 1. The robustness of populations to noise was also proved in the setting of non-elitist EAs [9, 26]. However, Friedrich et al. [17] showed the limitation of populations by proving that the  $(\mu+1)$ -EA needs super-polynomial time for solving OneMax under additive Gaussian noise  $\mathcal{N}(0, \sigma^2)$  with  $\sigma^2 \geq n^3$ . This difficulty can be overcome by the compact genetic algorithm (cGA) [17] and a simple Ant Colony Optimization (ACO) algorithm [16], both of which find the optimal solution in polynomial time with a high probability. ACO was also shown to be able to efficiently find solutions with reasonable approximations on some instances of the single-destination shortest path problem with edge weights disturbed by noise [11, 15, 33].

The ability of explicit noise handling strategies was also theoretically studied. Qian et al. [29] proved that the threshold selection strategy is robust to noise: the expected running time of the (1+1)-EA using threshold selection on OneMax under one-bit noise is always polynomial regardless of the noise probability  $p$ . For the (1+1)-EA solving OneMax under one-bit noise with  $p = 1$  or additive Gaussian noise  $\mathcal{N}(0, \sigma^2)$  with  $\sigma^2 \geq 1$ , the sampling strategy was shown to be able to reduce the running time from exponential to polynomial [28]. The robustness of sampling to noise was also proved for the (1+1)-EA solving LeadingOnes under one-bit noise with  $p = 1/2$  or additive Gaussian noise with  $\sigma^2 \geq n^2$ . Akimoto et al. [2] proved that sampling with a large sample size can make optimization under additive unbiased noise behave as optimization in a noise-free environment. The interplay between sampling and implicit noise-handling mechanisms (e.g., crossover) has been statistically studied in [18].

The noise models considered in the studies mentioned above are summarized in Table 1. We can observe that for the prior noise model, one-bit noise was mainly considered, which flips a random bit of a solution before evaluation with probability  $p$ . However, the noise model, which can change several bits of a solution simultaneously, may be more realistic and needs to be studied, as mentioned in the first noisy theoretical work [13].

In this paper, we study the bit-wise noise model, which is characterized by a pair  $(p, q)$  of parameters. It happens with probability  $p$ , and indepen-

**Table 2** For the running time of the (1+1)-EA on OneMax and LeadingOnes under prior noise models, the ranges of noise parameters for a polynomial upper bound and a super-polynomial lower bound are shown below.

(1+1)-EA	OneMax	LeadingOnes
bit-wise noise $(p, \frac{1}{n})$	$O(\log n/n), \omega(\log n/n)$	$O(\log n/n^2), \omega(\log n/n)$
bit-wise noise $(1, q)$	$O(\log n/n^2), \omega(\log n/n^2)$ [19]	$O(\log n/n^3), \omega(\log n/n^2)$
one-bit noise	$O(\log n/n), \omega(\log n/n)$ [13]	$[0, 1/(6en^2)], 1/2$ [19]; $O(\log n/n^2), \omega(\log n/n)$

**Table 3** For the running time of the (1+1)-EA using sampling on OneMax and LeadingOnes under prior noise models, the ranges of noise parameters for a polynomial upper bound and a super-polynomial lower bound are shown below.

(1+1)-EA using sampling	OneMax	LeadingOnes
bit-wise noise $(p, \frac{1}{n})$	$[0, 1], \emptyset$	$[0, 1], \emptyset$
bit-wise noise $(1, q)$	$1/2 - 1/n^{O(1)}, 1/2 - 1/n^{\omega(1)} \cup [1/2, 1]$	$O(\log n/n), \omega(\log n/n)$
one-bit noise	$[0, 1], \emptyset$	$[0, 1], \emptyset$

dently flips each bit of a solution with probability  $q$  before evaluation. We analyze the running time of the (1+1)-EA solving OneMax and LeadingOnes under bit-wise noise with two specific parameter settings  $(p, \frac{1}{n})$  and  $(1, q)$ . The ranges of  $p$  and  $q$  for a polynomial upper bound and a super-polynomial lower bound are derived, as shown in the middle row of Table 2. For the (1+1)-EA on LeadingOnes, we also transfer the running time bounds from bit-wise noise  $(p, \frac{1}{n})$  to one-bit noise by using the same proof procedure. As shown in the bottom right of Table 2, our results improve the previously known ones [19].

Note that for the (1+1)-EA on LeadingOnes, the current analysis (as shown in the last column of Table 2) does not cover all the ranges of  $p$  and  $q$ . We thus conduct experiments to estimate the expected running time for the uncovered values of  $p$  and  $q$ . The empirical results show that the currently derived ranges of  $p$  and  $q$  allowing a polynomial running time are possibly tight.

From the results in Table 2, we find that the (1+1)-EA is efficient only under low noise levels. For example, for the (1+1)-EA solving OneMax under bit-wise noise  $(p, \frac{1}{n})$ , the expected running time is polynomial only when  $p = O(\log n/n)$ . We then study whether the sampling strategy can bring robustness to noise. Sampling is a popular way to cope with noise in fitness evaluation [3], which, instead of evaluating the fitness of one solution only once, evaluates the fitness multiple ( $m$ ) times and then uses the average to approximate the true fitness. We analyze the running time of the (1+1)-EA using sampling under both bit-wise noise and one-bit noise. The ranges of  $p$  and  $q$  for a polynomial upper bound and a super-polynomial lower bound are shown in Table 3. Our analysis covers all the ranges of  $p$  and  $q$ . Note that for proving a polynomial upper bound, it is sufficient to show that using sampling with a specific sample size  $m$  can guarantee a polynomial running time, while for proving a super-polynomial lower bound, we need to prove that using sampling with any polynomially bounded  $m$  fails to guarantee

a polynomial running time. Compared with the results in Table 2, we find that using sampling significantly improves the noise-tolerance ability. For example, by using sampling with  $m = 4n^3$ , the (1+1)-EA now can always solve OneMax under bit-wise noise  $(p, \frac{1}{n})$  in polynomial time.

From the analysis, we also find the reason why sampling is effective or not. Let  $f(x)$  and  $f^n(x)$  denote the true and noisy fitness of a solution, respectively. For two solutions  $x$  and  $y$  with  $f(x) > f(y)$ , when the noise level is high (i.e., the values of  $p$  and  $q$  are large), the probability  $P(f^n(x) \leq f^n(y))$  (i.e., the true worse solution  $y$  appears to be better) becomes large, which will mislead the search direction and then lead to a super-polynomial running time. In such a situation, if the expected gap between  $f^n(x)$  and  $f^n(y)$  is positive, sampling will increase this trend and make  $P(f^n(x) \leq f^n(y))$  sufficiently small; if it is negative (e.g., on OneMax under bit-wise noise  $(1, q)$  with  $q \geq 1/2$ ), sampling will continue to increase  $P(f^n(x) \leq f^n(y))$ , and obviously will not work. We also note that if the positive gap between  $f^n(x)$  and  $f^n(y)$  is too small (e.g., on OneMax under bit-wise noise  $(1, q)$  with  $q = 1/2 - 1/n^{\omega(1)}$ ), a polynomial sample size will be not sufficient and sampling also fails to guarantee a polynomial running time.

This paper extends our preliminary work [27]. Since the theoretical analysis on the LeadingOnes problem is not complete, we add experiments to complement the theoretical results (i.e., Section 4.4). We also add the robustness analysis of sampling to noise (i.e., Section 5). Note that the robustness of sampling to one-bit noise has been studied in our previous work [28]. It was shown that sampling can reduce the running time of the (1+1)-EA from exponential to polynomial on OneMax when the noise probability  $p = 1$  as well as on LeadingOnes when  $p = 1/2$ . Therefore, our results here are more general. We prove that sampling is effective for any value of  $p$ , as shown in the last row of Table 3. Furthermore, we analyze the robustness of sampling to bit-wise noise for the first time.

The rest of this paper is organized as follows. Section 2 introduces some preliminaries. The running time analysis of the (1+1)-EA on OneMax and LeadingOnes under noise is presented in Sections 3 and 4, respectively. Section 5 analyzes the (1+1)-EA using sampling. Section 6 concludes the paper.

## 2 Preliminaries

In this section, we first introduce the optimization problems, noise models and evolutionary algorithms studied in this paper, respectively, then introduce the sampling strategy, and finally present the analysis tools that we use throughout this paper.

### 2.1 OneMax and LeadingOnes

In this paper, we use two well-known pseudo-Boolean functions OneMax and LeadingOnes. The OneMax problem as presented in Definition 1 aims

to maximize the number of 1-bits of a solution. The LeadingOnes problem as presented in Definition 2 aims to maximize the number of consecutive 1-bits counting from the left of a solution. Their optimal solution is  $11 \dots 1$  (briefly denoted as  $1^n$ ). It has been shown that the expected running time of the (1+1)-EA on OneMax and LeadingOnes is  $\Theta(n \log n)$  and  $\Theta(n^2)$ , respectively [14]. In the following analysis, we will use  $LO(x)$  to denote the number of leading 1-bits of a solution  $x$ .

**Definition 1 (OneMax)** The OneMax Problem of size  $n$  is to find an  $n$  bits binary string  $x^*$  such that

$$x^* = \arg \max_{x \in \{0,1\}^n} \left( f(x) = \sum_{i=1}^n x_i \right).$$

**Definition 2 (LeadingOnes)** The LeadingOnes Problem of size  $n$  is to find an  $n$  bits binary string  $x^*$  such that

$$x^* = \arg \max_{x \in \{0,1\}^n} \left( f(x) = \sum_{i=1}^n \prod_{j=1}^i x_j \right).$$

## 2.2 Bit-wise Noise

There are mainly two kinds of noise models: prior and posterior [19, 21]. Let  $f^n(x)$  and  $f(x)$  denote the noisy and true fitness of a solution  $x$ , respectively. The prior noise comes from the variation on a solution, i.e.,  $f^n(x) = f(x')$ , where  $x'$  is generated from  $x$  by random perturbations. The posterior noise comes from the variation on the fitness of a solution, e.g., additive noise  $f^n(x) = f(x) + \delta$  and multiplicative noise  $f^n(x) = f(x) \cdot \delta$ , where  $\delta$  is randomly drawn from some distribution. Previous theoretical analyses involving prior noise [9, 13, 19, 28, 29] often focused on a specific model, one-bit noise. As presented in Definition 3, it flips a random bit of a solution before evaluation with probability  $p$ . However, in many realistic applications, noise can change several bits of a solution simultaneously rather than only one bit. We thus consider the bit-wise noise model. As presented in Definition 4, it happens with probability  $p$ , and independently flips each bit of a solution with probability  $q$  before evaluation. We use bit-wise noise  $(p, q)$  to denote the bit-wise noise model with a scenario of  $(p, q)$ .

**Definition 3 (One-bit Noise)** Given a parameter  $p \in [0, 1]$ , let  $f^n(x)$  and  $f(x)$  denote the noisy and true fitness of a solution  $x \in \{0, 1\}^n$ , respectively, then

$$f^n(x) = \begin{cases} f(x) & \text{with probability } 1 - p, \\ f(x') & \text{with probability } p, \end{cases}$$

where  $x'$  is generated by flipping a uniformly randomly chosen bit of  $x$ .

**Definition 4 (Bit-wise Noise)** Given parameters  $p, q \in [0, 1]$ , let  $f^n(x)$  and  $f(x)$  denote the noisy and true fitness of a solution  $x \in \{0, 1\}^n$ , respectively, then

$$f^n(x) = \begin{cases} f(x) & \text{with probability } 1 - p, \\ f(x') & \text{with probability } p, \end{cases}$$

where  $x'$  is generated by independently flipping each bit of  $x$  with probability  $q$ .

To the best of our knowledge, only bit-wise noise  $(1, q)$  has been recently studied. Gießen and Kötzing [19] proved that for the (1+1)-EA solving OneMax, the expected running time is polynomial if  $q = O(\log n/n^2)$  and super-polynomial if  $q = \omega(\log n/n^2)$ . Besides bit-wise noise  $(1, q)$ , we also study another specific model bit-wise noise  $(p, \frac{1}{n})$  in this paper. Note that bit-wise noise  $(p, \frac{1}{n})$  is a natural extension of one-bit noise; their random behaviors of perturbing a solution correspond to the two common mutation operators, bit-wise mutation and one-bit mutation, respectively.

To investigate whether the performance of the (1+1)-EA for bit-wise noise with two scenarios of  $(p, q)$  and  $(p', q')$  where  $p \cdot q = p' \cdot q'$  can be significantly different, we consider the OneMax problem under bit-wise noise  $(1, \frac{\log n}{30n})$  and  $(\frac{\log n}{30n}, 1)$ . The comparison gives a positive answer. For bit-wise noise  $(1, \frac{\log n}{30n})$ , we know that the (1+1)-EA needs super-polynomial time to solve OneMax [19], while for bit-wise noise  $(\frac{\log n}{30n}, 1)$ , we will prove in Theorem 7 that the (1+1)-EA can solve OneMax in polynomial time. Thus, the analysis on general bit-wise noise without fixing  $p$  or  $q$  would be complicated, and  $p \cdot q$  may not be the only deciding factor. We leave it as a future work.

### 2.3 (1+1) Evolutionary Algorithm

The (1+1)-EA as described in Algorithm 1 is studied in this paper. For noisy optimization, only a noisy fitness value  $f^n(x)$  instead of the exact one  $f(x)$  can be accessed, and thus line 4 of Algorithm 1 is “if  $f^n(x') \geq f^n(x)$ ” instead of “if  $f(x') \geq f(x)$ ”. Note that the reevaluation strategy is used as in [11, 13, 19]. That is, besides evaluating  $f^n(x')$ ,  $f^n(x)$  will be reevaluated in each iteration of the (1+1)-EA. The running time is usually defined as the number of fitness evaluations needed to find an optimal solution w.r.t. the true fitness function  $f$  for the first time [2, 13, 19].

**Algorithm 1 ((1+1)-EA)** *Given a function  $f$  over  $\{0, 1\}^n$  to be maximized, it consists of the following steps:*

1.  $x :=$  uniformly randomly selected from  $\{0, 1\}^n$ .
2. Repeat until the termination condition is met
3.  $x' :=$  flip each bit of  $x$  independently with prob.  $1/n$ .
4. if  $f^n(x') \geq f^n(x)$
5.  $x := x'$ .

## 2.4 Sampling

In noisy evolutionary optimization, sampling has often been used to reduce the negative effect of noise [1, 6]. As presented in Definition 5, it approximates the true fitness  $f(x)$  using the average of multiple ( $m$ ) independent random evaluations. For the (1+1)-EA using sampling, line 4 of Algorithm 1 changes to be “if  $\hat{f}(x') \geq \hat{f}(x)$ ”. Its pseudo-code is described in Algorithm 2. Note that the sample size  $m = 1$  is equivalent to that sampling is not used. The effectiveness of sampling was not theoretically analyzed until recently. Qian et al. [28] proved that sampling is robust to one-bit noise and additive Gaussian noise. Particularly, under one-bit noise, it was shown that sampling can reduce the running time exponentially for the (1+1)-EA solving OneMax when the noise probability  $p = 1$  and LeadingOnes when  $p = 1/2$ .

**Definition 5 (Sampling)** Sampling first evaluates the fitness of a solution  $m$  times independently and obtains the noisy fitness values  $f_1^n(x), \dots, f_m^n(x)$ , and then outputs their average, i.e.,

$$\hat{f}(x) = \frac{1}{m} \sum_{i=1}^m f_i^n(x).$$

**Algorithm 2 ((1+1)-EA with sampling)** *Given a function  $f$  over  $\{0, 1\}^n$  to be maximized, it consists of the following steps:*

1.  $x :=$  uniformly randomly selected from  $\{0, 1\}^n$ .
2. Repeat until the termination condition is met
3.  $x' :=$  flip each bit of  $x$  independently with prob.  $1/n$ .
4. if  $\hat{f}(x') \geq \hat{f}(x)$
5.  $x := x'$ .

## 2.5 Analysis Tools

The process of the (1+1)-EA solving any pseudo-Boolean function with one unique global optimum can be directly modeled as a Markov chain  $\{\xi_t\}_{t=0}^{+\infty}$ . We only need to take the solution space  $\{0, 1\}^n$  as the chain's state space (i.e.,  $\xi_t \in \mathcal{X} = \{0, 1\}^n$ ), and take the optimal solution  $1^n$  as the chain's optimal state (i.e.,  $\mathcal{X}^* = \{1^n\}$ ). Note that we can assume without loss of generality that the optimal solution is  $1^n$ , because unbiased EAs treat the bits 0 and 1 symmetrically, and thus the 0 bits in an optimal solution can be interpreted as 1 bits without affecting the behavior of EAs. Given a Markov chain  $\{\xi_t\}_{t=0}^{+\infty}$  and  $\xi_t = x$ , we define its *first hitting time* (FHT) as  $\tau = \min\{t \mid \xi_{t+t} \in \mathcal{X}^*, t \geq 0\}$ . The mathematical expectation of  $\tau$ ,  $E(\tau \mid \xi_t = x) = \sum_{i=0}^{+\infty} i \cdot P(\tau = i \mid \xi_t = x)$ , is called the *expected first hitting time* (EFHT) starting from  $\xi_t = x$ . If  $\xi_0$  is drawn from a distribution  $\pi_0$ ,  $E(\tau \mid \xi_0 \sim \pi_0) = \sum_{x \in \mathcal{X}} \pi_0(x) E(\tau \mid \xi_0 = x)$  is called the EFHT of the Markov chain over the initial distribution  $\pi_0$ . Thus, the expected running time of the (1+1)-EA starting from  $\xi_0 \sim \pi_0$  is equal to  $1 + 2 \cdot E(\tau \mid \xi_0 \sim \pi_0)$ , where the term 1 corresponds to evaluating the initial



solution, and the factor 2 corresponds to evaluating the offspring solution  $x'$  and reevaluating the parent solution  $x$  in each iteration. If using sampling, the expected running time of the (1+1)-EA is  $m + 2m \cdot \mathbb{E}(\tau \mid \xi_0 \sim \pi_0)$ , since estimating the fitness of a solution needs  $m$  number of independent fitness evaluations. Note that we consider the expected running time of the (1+1)-EA starting from a uniform initial distribution in this paper.

In the following, we give three drift theorems that will be used to analyze the EFHT of Markov chains in the paper. The additive drift theorem [20] as presented in Theorem 1 is used to derive upper bounds on the EFHT of Markov chains. To use it, a function  $V(x)$  has to be constructed to measure the distance of a state  $x$  to the optimal state space  $\mathcal{X}^*$ . Then, we need to investigate the progress on the distance to  $\mathcal{X}^*$  in each step, i.e.,  $\mathbb{E}(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t)$ . An upper bound on the EFHT can be derived through dividing the initial distance by a lower bound on the progress.

**Theorem 1 (Additive Drift [20])** *Given a Markov chain  $\{\xi_t\}_{t=0}^{+\infty}$  and a distance function  $V(x)$ , if for any  $t \geq 0$  and any  $\xi_t$  with  $V(\xi_t) > 0$ , there exists a real number  $c > 0$  such that*

$$\mathbb{E}(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t) \geq c,$$

*then the EFHT satisfies that  $\mathbb{E}(\tau \mid \xi_0) \leq V(\xi_0)/c$ .*

The negative drift theorem [24, 25] as presented in Theorem 2 was proposed to prove exponential lower bounds on the EFHT of Markov chains, where  $X_t$  is usually represented by a mapping of  $\xi_t$ . It requires two conditions: a constant negative drift and exponentially decaying probabilities of jumping towards or away from the goal state. To relax the requirement of a constant negative drift, the negative drift theorem with self-loops [30] as presented in Theorem 3 has been proposed, which takes into account large self-loop probabilities.

**Theorem 2 (Negative Drift [24, 25])** *Let  $X_t, t \geq 0$ , be real-valued random variables describing a stochastic process. Suppose there exists an interval  $[a, b] \subseteq \mathbb{R}$ , two constants  $\delta, \epsilon > 0$  and, possibly depending on  $l := b - a$ , a function  $r(l)$  satisfying  $1 \leq r(l) = o(l/\log(l))$  such that for all  $t \geq 0$  the following two conditions hold:*

1.  $\mathbb{E}(X_t - X_{t+1} \mid a < X_t < b) \leq -\epsilon,$
2.  $\mathbb{P}(|X_{t+1} - X_t| \geq j \mid X_t > a) \leq \frac{r(l)}{(1 + \delta)^j} \text{ for } j \in \mathbb{N}_0.$

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

**Theorem 3 (Negative Drift with Self-loops [30])** *Let  $X_t, t \geq 0$ , be real-valued random variables describing a stochastic process. Suppose there exists an interval  $[a, b] \subseteq \mathbb{R}$ , two constants  $\delta, \epsilon > 0$  and, possibly depending on  $l := b - a$ ,*

a function  $r(l)$  satisfying  $1 \leq r(l) = o(l/\log(l))$  such that for all  $t \geq 0$  the following two conditions hold:

1.  $\forall a < i < b : E(X_t - X_{t+1} \mid X_t = i) \leq -\epsilon \cdot P(X_{t+1} \neq i \mid X_t = i),$
2.  $\forall i > a, j \in \mathbb{N}_0 : P(|X_{t+1} - X_t| \geq j \mid X_t = i) \leq \frac{r(l)}{(1+\delta)^j} \cdot P(X_{t+1} \neq i \mid X_t = i).$

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

### 3 The OneMax problem

In this section, we analyze the running time of the (1+1)-EA on OneMax under bit-wise noise. Note that for bit-wise noise  $(1, q)$ , it has been proved that the expected running time is polynomial if and only if  $q = O(\log n/n^2)$ , as shown in Theorem 4.

**Theorem 4** [19] *For the (1+1)-EA on OneMax under bit-wise noise  $(1, q)$ , the expected running time is polynomial if  $q = O(\log n/n^2)$  and super-polynomial if  $q = \omega(\log n/n^2)$ .*

For bit-wise noise  $(p, \frac{1}{n})$ , we prove in Theorems 5 and 6 that the tight range of  $p$  allowing a polynomial running time is  $O(\log n/n)$ . Instead of using the original drift theorems, we apply the upper and lower bounds of the (1+1)-EA on noisy OneMax in [19]. Let  $x^k$  denote any solution with  $k$  number of 1-bits, and  $f^n(x^k)$  denote its noisy objective value, which is a random variable. Lemma 1 intuitively means that if the probability of recognizing the true better solution by noisy evaluation is large (i.e., Eq. (1)), the running time can be upper bounded. Particularly, if Eq. (1) holds with  $l = O(\log n)$ , the running time can be polynomially upper bounded. On the contrary, Lemma 2 shows that if the probability of making a right comparison is small (i.e., Eq. (2)), the running time can be lower bounded. Particularly, if Eq. (2) holds with  $l = \Omega(n)$ , the running time can be exponentially lower bounded. Both Lemmas 1 and 2 are proved by applying standard drift theorems, and can be used to simplify our analysis. Note that in the original upper bound of the (1+1)-EA on noisy OneMax (i.e., Theorem 5 in [19]), it requires that Eq. (1) holds with only  $j = k$ , but the proof actually also requires that noisy OneMax satisfies the monotonicity property, i.e., for all  $j < k < n$ ,  $P(f^n(x^k) < f^n(x^{k+1})) \leq P(f^n(x^j) < f^n(x^{k+1}))$ . We have combined these two conditions in Lemma 1 by requiring Eq. (1) to hold with any  $j \leq k$  instead of only  $j = k$ .

**Lemma 1** [19] *Suppose there is a positive constant  $c \leq 1/15$  and some  $2 < l \leq n/2$  such that*

$$\begin{aligned} \forall j \leq k < n : P(f^n(x^j) < f^n(x^{k+1})) &\geq 1 - \frac{l}{n}; \\ \forall j \leq k < n - l : P(f^n(x^j) < f^n(x^{k+1})) &\geq 1 - c \frac{n-k}{n}, \end{aligned} \tag{1}$$

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

**Lemma 2** [19] Suppose there is some  $l \leq n/4$  and a constant  $c \geq 16$  such that

$$\forall n-l \leq k < n : P(f^n(x^k) < f^n(x^{k+1})) \leq 1 - c \frac{n-k}{n}, \quad (2)$$

then the (1+1)-EA optimizes  $f$  in  $2^{\Omega(l)}$  iterations with a high probability.

First, we apply Lemma 1 to show that the expected running time is polynomially upper bounded for bit-wise noise  $(p, \frac{1}{n})$  with  $p = O(\log n/n)$ .

**Theorem 5** For the (1+1)-EA on OneMax under bit-wise noise  $(p, \frac{1}{n})$ , the expected running time is polynomial if  $p = O(\log n/n)$ .

*Proof* We prove it by using Lemma 1. For some positive constant  $b$ , suppose that  $p \leq b \log n/n$ . We set the two parameters in Lemma 1 as  $c = \min\{\frac{1}{15}, b\}$  and  $l = \frac{2b \log n}{c} \in (2, \frac{n}{2}]$ .

For any  $j \leq k < n$ ,  $f^n(x^j) \geq f^n(x^{k+1})$  implies that  $f^n(x^j) \geq k+1$  or  $f^n(x^{k+1}) \leq k$ , either of which happens with probability at most  $p$ . By the union bound, we get  $\forall j \leq k < n$ ,

$$P(f^n(x^j) \geq f^n(x^{k+1})) \leq 2p \leq \frac{2b \log n}{n} = \frac{lc}{n} \leq \frac{l}{n}.$$

For any  $j \leq k < n-l$ , we easily get

$$P(f^n(x^j) \geq f^n(x^{k+1})) \leq \frac{lc}{n} < c \frac{n-k}{n}.$$

By Lemma 1, we know that the expected running time is  $O(n \log n) + n2^{O(2b \log n/c)}$ , i.e., polynomial.  $\square$

Next we apply Lemma 2 to show that the expected running time is super-polynomial for bit-wise noise  $(p, \frac{1}{n})$  with  $p = \omega(\log n/n)$ . Note that for  $p = 1 - O(\log n/n)$ , we actually give a stronger result that the expected running time is exponential.

**Theorem 6** For the (1+1)-EA on OneMax under bit-wise noise  $(p, \frac{1}{n})$ , the expected running time is super-polynomial if  $p = \omega(\log n/n) \cap 1 - \omega(\log n/n)$  and exponential if  $p = 1 - O(\log n/n)$ .

*Proof* We use Lemma 2 to prove it. Let  $c = 16$ . The case  $p = \omega(\log n/n) \cap 1 - \omega(\log n/n)$  is first analyzed. For any positive constant  $b$ , let  $l = b \log n$ . For any  $k \geq n-l$ , we get

$$P(f^n(x^k) \geq f^n(x^{k+1})) \geq P(f^n(x^k) = k) \cdot P(f^n(x^{k+1}) \leq k).$$

To make  $f^n(x^k) = k$ , it is sufficient that the noise does not happen, i.e.,  $P(f^n(x^k) = k) \geq 1 - p$ . To make  $f^n(x^{k+1}) \leq k$ , it is sufficient to flip one 1-bit

and keep other bits unchanged by noise, i.e.,  $P(f^n(x^{k+1}) \leq k) \geq p \cdot \frac{k+1}{n} (1 - \frac{1}{n})^{n-1}$ . Thus,

$$P(f^n(x^k) \geq f^n(x^{k+1})) \geq (1-p) \cdot p \frac{k+1}{en} = \omega(\log n/n).$$

Since  $c \frac{n-k}{n} \leq c \frac{l}{n} = \frac{cb \log n}{n}$ , the condition of Lemma 2 holds. Thus, the expected running time is  $2^{\Omega(b \log n)}$  (where  $b$  is any constant), i.e., super-polynomial.

For the case  $p = 1 - O(\log n/n)$ , let  $l = \sqrt{n}$ . We use another lower bound  $p(1 - \frac{1}{n})^n$  for  $P(f^n(x^k) = k)$ , since it is sufficient that no bit flips by noise. Thus, we have

$$P(f^n(x^k) \geq f^n(x^{k+1})) \geq p \left(1 - \frac{1}{n}\right)^n \cdot p \frac{k+1}{en} = \Omega(1).$$

Since  $c \frac{n-k}{n} \leq \frac{c\sqrt{n}}{n}$ , the condition of Lemma 2 holds. Thus, the expected running time is  $2^{\Omega(\sqrt{n})}$ , i.e., exponential.  $\square$

To show that the performance of the (1+1)-EA for bit-wise noise with two scenarios  $(p, q)$  and  $(p', q')$  where  $p \cdot q = p' \cdot q'$  can be significantly different, we compare the expected running time of the (1+1)-EA for bit-wise noise  $(1, \frac{\log n}{30n})$  and  $(\frac{\log n}{30n}, 1)$ . For the former case, we know from Theorem 4 that the expected running time is super-polynomial, while for the latter case, we prove in the following theorem that the expected running time can be polynomially upper bounded.

**Theorem 7** *For the (1+1)-EA on OneMax under bit-wise noise  $(\frac{\log n}{30n}, 1)$ , the expected running time is polynomial.*

*Proof* We use Lemma 1 to prove it. For any  $j \leq k < n$ ,  $f^n(x^j) \geq f^n(x^{k+1})$  implies that the fitness evaluation of  $x^j$  or  $x^{k+1}$  is affected by noise, whose probability is at most  $2 \cdot \frac{\log n}{30n} = \frac{\log n}{15n}$ . Thus, we have  $P(f^n(x^j) < f^n(x^{k+1})) \geq 1 - \frac{\log n}{15n}$ . It is then easy to verify that the condition of Lemma 1 holds with  $c = \frac{1}{15}$  and  $l = \log n$ . Thus, the expected running time is polynomial.  $\square$

#### 4 The LeadingOnes problem

In this section, we first analyze the running time of the (1+1)-EA on the LeadingOnes problem under bit-wise noise  $(p, \frac{1}{n})$  and bit-wise noise  $(1, q)$ , respectively. Then, we transfer the analysis from bit-wise noise  $(p, \frac{1}{n})$  to one-bit noise; the results are complementary to the known ones recently derived in [19]. However, our analysis does not cover all the ranges of  $p$  and  $q$ . For those values of  $p$  and  $q$  where no theoretical results are known, we conduct experiments to empirically investigate the running time.

#### 4.1 Bit-wise Noise $(p, \frac{1}{n})$

For bit-wise noise  $(p, \frac{1}{n})$ , we first apply the additive drift theorem (i.e., Theorem 1) to prove that the expected running time is polynomial when  $p = O(\log n/n^2)$ .

**Theorem 8** *For the (1+1)-EA on LeadingOnes under bit-wise noise  $(p, \frac{1}{n})$ , the expected running time is polynomial if  $p = O(\log n/n^2)$ .*

*Proof* We use Theorem 1 to prove it. For some positive constant  $b$ , suppose that  $p \leq b \log n/n^2$ . Let  $\theta \in (0, 1)$  be some constant close to 0. We first construct a distance function  $V(x)$  as, for any  $x$  with  $\text{LO}(x) = i$ ,

$$V(x) = \left(1 + \frac{c}{n}\right)^n - \left(1 + \frac{c}{n}\right)^i, \quad (3)$$

where  $c = \frac{2b \log n}{1-\theta} + 1$ . It is easy to verify that  $V(x \in \mathcal{X}^* = \{1^n\}) = 0$  and  $V(x \notin \mathcal{X}^*) > 0$ .

Then, we investigate  $E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x)$  for any  $x$  with  $\text{LO}(x) < n$  (i.e.,  $x \notin \mathcal{X}^*$ ). Assume that currently  $\text{LO}(x) = i$ , where  $0 \leq i \leq n-1$ . Let  $P_{\text{mut}}(x, x')$  denote the probability of generating  $x'$  by mutation on  $x$ . We divide the drift into two parts: positive  $E^+$  and negative  $E^-$ . That is,

$$E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x) = E^+ - E^-,$$

where

$$E^+ = \sum_{x': \text{LO}(x') > i} P_{\text{mut}}(x, x') \cdot P(f^n(x') \geq f^n(x)) \cdot (V(x) - V(x')), \quad (4)$$

$$E^- = \sum_{x': \text{LO}(x') < i} P_{\text{mut}}(x, x') \cdot P(f^n(x') \geq f^n(x)) \cdot (V(x') - V(x)). \quad (5)$$

For the positive drift  $E^+$ , we need to consider that the number of leading 1-bits is increased. By mutation, we have

$$\sum_{x': \text{LO}(x') > i} P_{\text{mut}}(x, x') = P(\text{LO}(x') \geq i+1) = \left(1 - \frac{1}{n}\right)^i \frac{1}{n}, \quad (6)$$

since it needs to flip the  $(i+1)$ -th bit (which must be 0) of  $x$  and keep the  $i$  leading 1-bits unchanged. For any  $x'$  with  $\text{LO}(x') \geq i+1$ ,  $f^n(x') < f^n(x)$  implies that  $f^n(x') \leq i-1$  or  $f^n(x') \geq i+1$ . Note that,

$$P(f^n(x') \leq i-1) = p \left(1 - \left(1 - \frac{1}{n}\right)^i\right), \quad (7)$$

since at least one of the first  $i$  leading 1-bits of  $x'$  needs to be flipped by noise;

$$P(f^n(x') \geq i+1) = p \left(1 - \frac{1}{n}\right)^i \frac{1}{n}, \quad (8)$$

since it needs to flip the first 0-bit of  $x$  and keep the leading 1-bits unchanged by noise. By the union bound, we get

$$\begin{aligned} P(f^n(x') \geq f^n(x)) &= 1 - P(f^n(x') < f^n(x)) \\ &\geq 1 - p \left( 1 - \left( 1 - \frac{1}{n} \right)^{i+1} \right) \geq 1 - p \frac{i+1}{n} \geq 1 - p \geq 1 - \theta, \end{aligned} \quad (9)$$

where the last inequality holds with sufficiently large  $n$ , since  $p = O(\log n/n^2)$  and  $\theta \in (0, 1)$  is some constant close to 0. Furthermore, for any  $x'$  with  $V(x') \geq i + 1$ ,

$$V(x) - V(x') \geq \left( 1 + \frac{c}{n} \right)^{i+1} - \left( 1 + \frac{c}{n} \right)^i = \frac{c}{n} \left( 1 + \frac{c}{n} \right)^i. \quad (10)$$

By combining Eqs. (6), (9) and (10), we have

$$E^+ \geq \left( 1 - \frac{1}{n} \right)^i \frac{1}{n} \cdot (1 - \theta) \cdot \frac{c}{n} \left( 1 + \frac{c}{n} \right)^i \geq \frac{(1 - \theta)c}{3n^2} \left( 1 + \frac{c}{n} \right)^i,$$

where the last inequality is by  $(1 - \frac{1}{n})^i \geq (1 - \frac{1}{n})^{n-1} \geq \frac{1}{e} \geq \frac{1}{3}$ .

For the negative drift  $E^-$ , we need to consider that the number of leading 1-bits is decreased. By mutation, we have

$$\sum_{x': \text{LO}(x') < i} P_{mut}(x, x') = P(\text{LO}(x') \leq i - 1) = 1 - \left( 1 - \frac{1}{n} \right)^i, \quad (11)$$

since it needs to flip at least one leading 1-bit of  $x$ . For any  $x'$  with  $\text{LO}(x') \leq i - 1$  (where  $i \geq 1$ ),  $f^n(x') \geq f^n(x)$  implies that  $f^n(x') \geq i$  or  $f^n(x) \leq i - 1$ . Note that,

$$P(f^n(x') \geq i) \leq p \left( 1 - \frac{1}{n} \right)^{i-1} \frac{1}{n}, \quad (12)$$

since for the first  $i$  bits of  $x'$ , it needs to flip the 0-bits (whose number is at least 1) and keep the 1-bits unchanged by noise;

$$P(f^n(x) \leq i - 1) = p \left( 1 - \left( 1 - \frac{1}{n} \right)^i \right), \quad (13)$$

since at least one leading 1-bit of  $x$  needs to be flipped by noise. By the union bound, we get

$$P(f^n(x') \geq f^n(x)) \leq p - p \left( 1 - \frac{2}{n} \right) \left( 1 - \frac{1}{n} \right)^{i-1} \leq p \frac{i+1}{n}. \quad (14)$$

Furthermore, according to the definition of the distance function (i.e., Eq. (3)), we have for any  $x'$  with  $0 \leq \text{LO}(x') \leq i - 1$ ,

$$V(x') - V(x) = \left( 1 + \frac{c}{n} \right)^i - \left( 1 + \frac{c}{n} \right)^{\text{LO}(x')} \leq \left( 1 + \frac{c}{n} \right)^i - 1. \quad (15)$$

By combining Eqs. (11), (14) and (15), we have

$$\begin{aligned} E^- &\leq \left(1 - \left(1 - \frac{1}{n}\right)^i\right) \cdot p \frac{i+1}{n} \cdot \left(\left(1 + \frac{c}{n}\right)^i - 1\right) \\ &\leq \left(1 - \frac{1}{e}\right) \cdot p \cdot \left(1 + \frac{c}{n}\right)^i \leq \frac{2p}{3} \left(1 + \frac{c}{n}\right)^i. \end{aligned}$$

Thus, by subtracting  $E^-$  from  $E^+$ , we have

$$\begin{aligned} E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x) &\geq \left(1 + \frac{c}{n}\right)^i \left(\frac{(1-\theta)c}{3n^2} - \frac{2p}{3}\right) \\ &\geq \left(1 + \frac{c}{n}\right)^i \left(\frac{2b \log n + 1 - \theta}{3n^2} - \frac{2b \log n}{3n^2}\right) \geq \frac{1-\theta}{3n^2}, \end{aligned} \quad (16)$$

where the second inequality is by  $c = \frac{2b \log n}{1-\theta} + 1$  and  $p \leq b \log n / n^2$ . Note that  $V(x) \leq (1 + \frac{c}{n})^n \leq e^c = e^{\frac{2b \log n}{1-\theta} + 1} = en^{\frac{2b}{1-\theta}}$ . By Theorem 1, we get

$$E(\tau \mid \xi_0) \leq \frac{3n^2}{1-\theta} \cdot en^{\frac{2b}{1-\theta}} = O\left(n^{\frac{2b}{1-\theta} + 2}\right),$$

i.e., the expected running time is polynomial.  $\square$

Next we use the negative drift with self-loops theorem (i.e., Theorem 3) to show that the expected running time is super-polynomial for bit-wise noise  $(p, \frac{1}{n})$  with  $p = \omega(\log n / n) \cap o(1)$ .

**Theorem 9** *For the (1+1)-EA on LeadingOnes under bit-wise noise  $(p, \frac{1}{n})$ , if  $p = \omega(\log n / n) \cap o(1)$ , the expected running time is super-polynomial.*

*Proof* We use Theorem 3 to prove it. Let  $X_t = |x|_0$  be the number of 0-bits of the solution  $x$  after  $t$  iterations of the (1+1)-EA. Let  $c$  be any positive constant. We consider the interval  $[0, c \log n]$ , i.e., the parameters  $a = 0$  (i.e., the global optimum) and  $b = c \log n$  in Theorem 3.

Then, we analyze the drift  $E(X_t - X_{t+1} \mid X_t = i)$  for  $1 \leq i < c \log n$ . As in the proof of Theorem 8, we divide the drift into two parts: positive  $E^+$  and negative  $E^-$ . That is,

$$E(X_t - X_{t+1} \mid X_t = i) = E^+ - E^-,$$

where

$$\begin{aligned} E^+ &= \sum_{x': |x'|_0 < i} P_{mut}(x, x') \cdot P(f^n(x') \geq f^n(x)) \cdot (i - |x'|_0), \\ E^- &= \sum_{x': |x'|_0 > i} P_{mut}(x, x') \cdot P(f^n(x') \geq f^n(x)) \cdot (|x'|_0 - i). \end{aligned}$$

Note that the drift here depends on the number of 0-bits due to the definition of  $X_t$ . It is different from that in the proof of Theorem 8, which depends

on the number of leading 1-bits due to the definition of the distance function (i.e., Eq. (3)).

For the positive drift  $E^+$ , we need to consider that the number of 0-bits is decreased. For mutation on  $x$  (where  $|x|_0 = i$ ), let  $X$  and  $Y$  denote the number of flipped 0-bits and 1-bits, respectively. Then,  $X \sim B(i, \frac{1}{n})$  and  $Y \sim B(n - i, \frac{1}{n})$ , where  $B(\cdot, \cdot)$  is the binomial distribution. To estimate an upper bound on  $E^+$ , we assume that the offspring solution  $x'$  with  $|x'|_0 < i$  is always accepted. Thus, we have

$$\begin{aligned} E^+ &\leq \sum_{x': |x'|_0 < i} P_{mut}(x, x')(i - |x'|_0) = \sum_{k=1}^i k \cdot P(X - Y = k) \quad (17) \\ &= \sum_{k=1}^i k \cdot \sum_{j=k}^i P(X = j) \cdot P(Y = j - k) \\ &= \sum_{j=1}^i \sum_{k=1}^j k \cdot P(X = j) \cdot P(Y = j - k) \\ &\leq \sum_{j=1}^i j \cdot P(X = j) = \frac{i}{n}. \end{aligned}$$

For the negative drift  $E^-$ , we need to consider that the number of 0-bits is increased. We analyze the  $n - i$  cases where only one 1-bit is flipped (i.e.,  $|x'|_0 = i + 1$ ), which happens with probability  $\frac{1}{n}(1 - \frac{1}{n})^{n-1}$ . Assume that  $LO(x) = k \leq n - i$ . If the  $j$ -th (where  $1 \leq j \leq k$ ) leading 1-bit is flipped, the offspring solution  $x'$  will be accepted (i.e.,  $f^n(x') \geq f^n(x)$ ) if  $f^n(x') \geq j - 1$  and  $f^n(x) \leq j - 1$ . Note that,

$$P(f^n(x') \geq j - 1) = 1 - p + p \left(1 - \frac{1}{n}\right)^{j-1} \geq 1 - p \frac{j-1}{n} \geq \frac{1}{2}, \quad (18)$$

where the equality is since it needs to keep the  $j - 1$  leading 1-bits of  $x'$  unchanged, and the last inequality is by  $p = o(1)$ ;

$$\begin{aligned} P(f^n(x) \leq j - 1) &= p \left(1 - \left(1 - \frac{1}{n}\right)^j\right) \quad (19) \\ &= p \left(1 - \frac{1}{n}\right)^j \left(\left(1 + \frac{1}{n-1}\right)^j - 1\right) \geq \frac{p}{e} \cdot \frac{j}{n-1} \geq \frac{pj}{3n}, \end{aligned}$$

where the equality is since at least one of the first  $j$  leading 1-bits of  $x$  needs to be flipped by noise. Thus, we get

$$P(f^n(x') \geq f^n(x)) \geq \frac{pj}{6n}. \quad (20)$$

If one of the  $n - i - k$  non-leading 1-bits is flipped,  $LO(x') = LO(x) = k$ . We can use the same analysis procedure as Eq. (9) in the proof of Theorem 8 to derive that

$$P(f^n(x') \geq f^n(x)) \geq 1 - p \frac{k+1}{n} \geq \frac{1}{2}, \quad (21)$$



where the last inequality is by  $p = o(1)$ . Combining all the  $n - i$  cases, we get

$$\begin{aligned} E^- &\geq \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1} \cdot \left(\sum_{j=1}^k \frac{pj}{6n} + \frac{n-i-k}{2}\right) \cdot (i+1-i) \\ &\geq \frac{1}{en} \left(\frac{pk(k+1)}{12n} + \frac{n-i-k}{2}\right) \geq \frac{pk^2}{36n^2} + \frac{n-i-k}{6n}. \end{aligned} \quad (22)$$

By subtracting  $E^-$  from  $E^+$ , we get

$$E(X_t - X_{t+1} \mid X_t = i) \leq \frac{i}{n} - \frac{pk^2}{36n^2} - \frac{n-i-k}{6n}.$$

To investigate condition (1) of Theorem 3, we also need to analyze the probability  $P(X_{t+1} \neq i \mid X_t = i)$ . For  $X_{t+1} \neq i$ , it is necessary that at least one bit of  $x$  is flipped and the offspring  $x'$  is accepted. We consider two cases: (1) at least one of the  $k$  leading 1-bits of  $x$  is flipped; (2) the  $k$  leading 1-bits of  $x$  are not flipped and at least one of the last  $n - k$  bits is flipped. For case (1), the mutation probability is  $1 - (1 - \frac{1}{n})^k$  and the acceptance probability is at most  $p \frac{k+1}{n}$  by Eq. (14). For case (2), the mutation probability is  $(1 - \frac{1}{n})^k (1 - (1 - \frac{1}{n})^{n-k}) \leq \frac{n-k}{n}$  and the acceptance probability is at most 1. Thus, we have

$$P(X_{t+1} \neq i \mid X_t = i) \leq p + \frac{n-k}{n}. \quad (23)$$

When  $k < n - np$ , we have

$$\begin{aligned} E(X_t - X_{t+1} \mid X_t = i) &\leq \frac{i}{n} - \frac{n-i-k}{6n} \\ &\leq -\frac{n-k}{12n} - \frac{np/2 - 7c \log n}{6n} \leq -\frac{n-k}{12n} \leq -\frac{1}{24} \left(p + \frac{n-k}{n}\right), \end{aligned} \quad (24)$$

where the second inequality is by  $n - k > np$  and  $i < c \log n$ , the third inequality is by  $p = \omega(\log n/n)$  and the last is by  $n - k > np$ . When  $k \geq n - np$ , we have

$$\begin{aligned} E(X_t - X_{t+1} \mid X_t = i) &\leq \frac{i}{n} - \frac{pk^2}{36n^2} \\ &\leq \frac{c \log n}{n} - \frac{p}{144} \leq -\frac{p}{288} \leq -\frac{1}{576} \left(p + \frac{n-k}{n}\right), \end{aligned} \quad (25)$$

where the second inequality is by  $p = o(1)$  and  $i < c \log n$ , the third is by  $p = \omega(\log n/n)$  and the last is by  $n - k \leq np$ . Combining Eqs. (23), (24) and (25), we get that condition (1) of Theorem 3 holds with  $\epsilon = \frac{1}{576}$ .

For condition (2) of Theorem 3, we need to show  $P(|X_{t+1} - X_t| \geq j \mid X_t = i) \leq \frac{r(l)}{(1+\delta)^j} \cdot P(X_{t+1} \neq i \mid X_t = i)$  for  $i \geq 1$ . For  $P(X_{t+1} \neq i \mid X_t = i)$ , we analyze the  $n$  cases where only one bit is flipped. Using the similar

analysis procedure as  $E^-$ , except that flipping any bit rather than only 1-bit is considered here, we easily get

$$P(X_{t+1} \neq i \mid X_t = i) \geq \frac{pk(k+1)}{36n^2} + \frac{n-k}{6n}. \quad (26)$$

For  $|X_{t+1} - X_t| \geq j$ , it is necessary that at least  $j$  bits of  $x$  are flipped and the offspring solution  $x'$  is accepted. We consider two cases: (1) at least one of the  $k$  leading 1-bits is flipped; (2) the  $k$  leading 1-bits are not flipped. For case (1), the mutation probability is at most  $\frac{k}{n} \binom{n-1}{j-1} \frac{1}{n^{j-1}}$  and the acceptance probability is at most  $p \frac{k+1}{n}$  by Eq. (14). For case (2), the mutation probability is at most  $(1 - \frac{1}{n})^k \binom{n-k}{j} \frac{1}{n^j}$  and the acceptance probability is at most 1. Thus, we have

$$\begin{aligned} P(|X_{t+1} - X_t| \geq j \mid X_t = i) & \leq \frac{k}{n} \binom{n-1}{j-1} \frac{1}{n^{j-1}} \cdot p \frac{k+1}{n} + \left(1 - \frac{1}{n}\right)^k \binom{n-k}{j} \frac{1}{n^j} \\ & \leq \frac{pk(k+1)}{n^2} \cdot \frac{4}{2^j} + \frac{n-k}{n} \cdot \frac{2}{2^j} \leq \left( \frac{pk(k+1)}{36n^2} + \frac{n-k}{6n} \right) \cdot \frac{144}{2^j}. \end{aligned} \quad (27)$$

By combining Eq. (26) with Eq. (27), we get that condition (2) of Theorem 3 holds with  $\delta = 1$  and  $r(l) = 144$ .

Note that  $l = b - a = c \log n$ . By Theorem 3, the expected running time is  $2^{\Omega(c \log n)}$ , where  $c$  is any positive constant. Thus, the expected running time is super-polynomial.  $\square$

For  $p = \Omega(1)$ , we can use the negative drift theorem (i.e., Theorem 2) to derive a stronger result that the expected running time is exponentially lower bounded.

**Theorem 10** *For the (1+1)-EA on LeadingOnes under bit-wise noise  $(p, \frac{1}{n})$ , the expected running time is exponential if  $p = \Omega(1)$ .*

*Proof* We use Theorem 2 to prove it. Let  $X_t = i$  be the number of 0-bits of the solution  $x$  after  $t$  iterations of the (1+1)-EA. We consider the interval  $i \in [0, n^{1/2}]$ . To analyze the drift  $E(X_t - X_{t+1} \mid X_t = i) = E^+ - E^-$ , we use the same analysis procedure as Theorem 9. For the positive drift, we have  $E^+ \leq \frac{i}{n} = o(1)$ . For the negative drift, we re-analyze Eqs. (20) and (21). From Eqs. (18) and (19), we get that  $P(f^n(x') \geq j-1) \geq p(1 - \frac{j-1}{n})$  and  $P(f^n(x) \leq j-1) \geq \frac{pj}{3n}$ . Thus, Eq. (20) becomes

$$P(f^n(x') \geq f^n(x)) \geq \frac{p^2 j}{3n} \left(1 - \frac{j-1}{n}\right). \quad (28)$$

For Eq. (21), we need to analyze the acceptance probability for  $LO(x') = LO(x) = k$ . Since it is sufficient to keep the first  $(k+1)$  bits of  $x$  and  $x'$  unchanged in noise, Eq. (21) becomes

$$P(f^n(x') \geq f^n(x)) \geq p^2 \left(1 - \frac{1}{n}\right)^{2(k+1)} \geq p^2 \left(1 - \frac{k+1}{n}\right)^2. \quad (29)$$

By applying the above two inequalities to Eq. (22), we have

$$E^- \geq \frac{p^2}{en} \left( \sum_{j=1}^k \frac{j(n-j+1)}{3n^2} + \frac{(n-i-k)(n-1-k)^2}{n^2} \right) = \Omega(1),$$

where the equality is by  $p = \Omega(1)$ . Thus,  $E^+ - E^- = -\Omega(1)$ . That is, condition (1) of Theorem 2 holds.

Since it is necessary to flip at least  $j$  bits of  $x$ , we have

$$P(|X_{t+1} - X_t| \geq j \mid X_t \geq 1) \leq \binom{n}{j} \frac{1}{n^j} \leq \frac{1}{j!} \leq 2 \cdot \frac{1}{2^j},$$

which implies that condition (2) of Theorem 2 holds with  $\delta = 1$  and  $r(l) = 2$ . Note that  $l = n^{1/2}$ . Thus, by Theorem 2, the expected running time is exponential.  $\square$

#### 4.2 Bit-wise Noise $(1, q)$

For bit-wise noise  $(1, q)$ , the proof idea is similar to that for bit-wise noise  $(p, \frac{1}{n})$ . The main difference led by the change of noise is the probability of accepting the offspring solution, i.e.,  $P(f^n(x') \geq f^n(x))$ . We first prove that the expected running time is polynomial when  $q = O(\log n/n^3)$ .

**Theorem 11** *For the (1+1)-EA on LeadingOnes under bit-wise noise  $(1, q)$ , the expected running time is polynomial if  $q = O(\log n/n^3)$ .*

*Proof* The proof is very similar to that of Theorem 8. The change of noise only affects the probability of accepting the offspring solution in the analysis. For some positive constant  $b$ , suppose that  $q \leq b \log n/n^3$ .

For the positive drift  $E^+$ , we need to re-analyze  $P(f^n(x') \geq f^n(x))$  (i.e., Eq. (9) in the proof of Theorem 8) for the parent  $x$  with  $LO(x) = i$  and the offspring  $x'$  with  $LO(x') \geq i + 1$ . By bit-wise noise  $(1, q)$ , Eqs. (7) and (8) change to

$$P(f^n(x') \leq i - 1) = 1 - (1 - q)^i; \quad P(f^n(x) \geq i + 1) = (1 - q)^i q.$$

Thus, by the union bound, Eq. (9) becomes

$$\begin{aligned} P(f^n(x') \geq f^n(x)) &\geq 1 - (1 - (1 - q)^i + (1 - q)^i q) \\ &= (1 - q)^{i+1} \geq 1 - q(i + 1) \geq 1 - \theta, \end{aligned} \tag{30}$$

where the last inequality holds with sufficiently large  $n$ , since  $q = O(\log n/n^3)$  and  $\theta \in (0, 1)$  is some constant close to 0.

For the negative drift  $E^-$ , we need to re-analyze  $P(f^n(x') \geq f^n(x))$  (i.e., Eq. (14) in the proof of Theorem 8) for the parent  $x$  with  $LO(x) = i$  (where

$i \geq 1$ ) and the offspring  $x'$  with  $\text{LO}(x') \leq i - 1$ . By bit-wise noise  $(1, q)$ , Eqs. (12) and (13) change to

$$P(f^n(x') \geq i) \leq q(1 - q)^{i-1}, \quad P(f^n(x) \leq i - 1) = 1 - (1 - q)^i.$$

Thus, by the union bound, Eq. (14) becomes

$$\begin{aligned} P(f^n(x') \geq f^n(x)) &\leq q(1 - q)^{i-1} + 1 - (1 - q)^i \\ &= 1 - (1 - q)^{i-1}(1 - 2q) \leq 1 - (1 - (i - 1)q)(1 - 2q) \leq (i + 1)q, \end{aligned} \quad (31)$$

where the second inequality is by  $(1 - q)^{i-1} \geq 1 - (i - 1)q$  and  $1 - 2q > 0$  for  $q = O(\log n/n^3)$ .

By applying Eq. (30) and Eq. (31) to  $E^+$  and  $E^-$ , respectively, Eq. (16) changes to

$$\begin{aligned} E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x) &\geq \left(1 + \frac{c}{n}\right)^i \left(\frac{(1 - \theta)c}{3n^2} - \frac{2q(i + 1)}{3}\right) \\ &\geq \left(1 + \frac{c}{n}\right)^i \left(\frac{2b \log n + 1 - \theta}{3n^2} - \frac{2bn \log n}{3n^3}\right) \geq \frac{1 - \theta}{3n^2}. \end{aligned}$$

That is, the condition of Theorem 1 still holds with  $\frac{1 - \theta}{3n^2}$ . Thus, the expected running time is polynomial.  $\square$

Next we prove that the expected running time is super-polynomial when  $q$  is in the range of  $\omega(\log n/n^2) \cap o(1/n)$ .

**Theorem 12** *For the (1+1)-EA on LeadingOnes under bit-wise noise  $(1, q)$ , if  $q = \omega(\log n/n^2) \cap o(1/n)$ , the expected running time is super-polynomial.*

*Proof* We use the same analysis procedure as Theorem 9. The only difference is the probability of accepting the offspring solution  $x'$  due to the change of noise. For the positive drift, we still have  $E^+ \leq \frac{i}{n}$ , since we optimistically assume that  $x'$  is always accepted in the proof of Theorem 9.

For the negative drift, we need to re-analyze  $P(f^n(x') \geq f^n(x))$  for the parent solution  $x$  with  $\text{LO}(x) = k$  and the offspring solution  $x'$  with  $\text{LO}(x') = j - 1$  (where  $1 \leq j \leq k + 1$ ). For  $j \leq k$ , to derive a lower bound on  $P(f^n(x') \geq f^n(x))$ , we consider the  $j$  cases where  $f^n(x) = l$  and  $f^n(x') \geq l$  for  $0 \leq l \leq j - 1$ . Since  $P(f^n(x) = l) = (1 - q)^l q$  and  $P(f^n(x') \geq l) = (1 - q)^l$ , Eq. (20) changes to

$$\begin{aligned} P(f^n(x') \geq f^n(x)) &\geq \sum_{l=0}^{j-1} (1 - q)^l q \cdot (1 - q)^l \geq \frac{1 - (1 - q)^{2j}}{2} \\ &= \frac{1}{2}(1 - q)^{2j} \left( \left(1 + \frac{q}{1 - q}\right)^{2j} - 1 \right) \geq (1 - q)^{2j} \frac{qj}{1 - q} \geq \frac{qj}{2}, \end{aligned} \quad (32)$$

where the last inequality is by  $(1 - q)^{2j} \geq 1 - 2qj \geq 1/2$  since  $q = o(1/n)$ . For  $j = k + 1$  (i.e.,  $\text{LO}(x') = \text{LO}(x) = k$ ), we can use the same analysis as

Eq. (30) to derive a lower bound  $1 - q(k+1) \geq 1/2$ , where the inequality is by  $q = o(1/n)$ . Thus, Eq. (21) also holds here, i.e.,

$$P(f^n(x') \geq f^n(x)) \geq \frac{1}{2}. \quad (33)$$

By applying Eqs. (32) and (33) to  $E^-$ , Eq. (22) changes to

$$E^- \geq \frac{qk^2}{12n} + \frac{n-i-k}{6n}.$$

Thus, we have

$$E(X_t - X_{t+1} \mid X_t = i) = E^+ - E^- \leq \frac{i}{n} - \frac{qk^2}{12n} - \frac{n-i-k}{6n}.$$

For the upper bound analysis of  $P(X_{t+1} \neq i \mid X_t = i)$  in the proof of Theorem 9, we only need to replace the acceptance probability  $p^{\frac{k+1}{n}}$  in the case of  $LO(x') < LO(x)$  with  $(k+1)q$  (i.e., Eq. (31)). Thus, Eq. (23) changes to

$$P(X_{t+1} \neq i \mid X_t = i) \leq (k+1)q + \frac{n-k}{n} \leq nq + \frac{n-k}{n}.$$

To compare  $E(X_t - X_{t+1} \mid X_t = i)$  with  $P(X_{t+1} \neq i \mid X_t = i)$ , we consider two cases:  $k < n - n^2q$  and  $k \geq n - n^2q$ . By using  $q = \omega(\log n/n^2)$  and applying the same analysis procedure as Eqs. (24) and (25), we can derive that condition (1) of Theorem 3 holds with  $\epsilon = \frac{1}{192}$ .

For the lower bound analysis of  $P(X_{t+1} \neq i \mid X_t = i)$ , by applying Eqs. (32) and (33), Eq. (26) changes to

$$P(X_{t+1} \neq i \mid X_t = i) \geq \frac{qk(k+1)}{12n} + \frac{n-k}{6n}.$$

For the analysis of  $|X_{t+1} - X_t| \geq j$ , by replacing the acceptance probability  $p^{\frac{k+1}{n}}$  in the case of  $LO(x') < LO(x)$  with  $(k+1)q$ , Eq. (27) changes to

$$\begin{aligned} P(|X_{t+1} - X_t| \geq j \mid X_t = i) &\leq \frac{qk(k+1)}{n} \cdot \frac{4}{2^j} + \frac{n-k}{n} \cdot \frac{2}{2^j} \\ &\leq \left( \frac{qk(k+1)}{12n} + \frac{n-k}{6n} \right) \cdot \frac{48}{2^j}. \end{aligned}$$

That is, condition (2) of Theorem 3 holds with  $\delta = 1, r(l) = 48$ . Thus, the expected running time is super-polynomial.  $\square$

For  $q = \Omega(1/n)$ , we prove a stronger result that the expected running time is exponentially lower bounded.

**Theorem 13** *For the (1+1)-EA on LeadingOnes under bit-wise noise  $(1, q)$ , the expected running time is exponential if  $q = \Omega(1/n)$ .*

*Proof* We use Theorem 2 to prove it. Let  $X_t = i$  be the number of 0-bits of the solution  $x$  after  $t$  iterations of the (1+1)-EA. We consider the interval  $i \in [0, n^{1/2}]$ . To analyze the drift  $E(X_t - X_{t+1} \mid X_t = i)$ , we use the same analysis procedure as the proof of Theorem 9.

We first consider  $q = \Omega(1/n) \cap o(1)$ . We need to analyze the probability  $P(f^n(x') \geq f^n(x))$ , where the offspring solution  $x'$  is generated by flipping only one 1-bit of  $x$ . Let  $\text{LO}(x) = k$ . For the case where the  $j$ -th (where  $1 \leq j \leq k$ ) leading 1-bit is flipped, as the analysis of Eq. (32), we get

$$P(f^n(x') \geq f^n(x)) \geq \frac{1 - (1-q)^{2j}}{2} \geq (1-q)^{2j} \frac{qj}{1-q}.$$

If  $(1-q)^{2j} < \frac{1}{2}$ ,  $\frac{1-(1-q)^{2j}}{2} \geq \frac{1}{4}$ ; otherwise,  $(1-q)^{2j} \frac{qj}{1-q} \geq \frac{qj}{2}$ . Thus, we have

$$P(f^n(x') \geq f^n(x)) \geq \min\{1/4, qj/2\}.$$

For the case that flips one non-leading 1-bit (i.e.,  $\text{LO}(x') = \text{LO}(x) = k$ ), to derive a lower bound on  $P(f^n(x') \geq f^n(x))$ , we consider  $f^n(x) = l$  and  $f^n(x') \geq l$  for  $0 \leq l \leq k$ . Thus,

$$\begin{aligned} P(f^n(x') \geq f^n(x)) &\geq \sum_{l=0}^{k-1} (1-q)^l q \cdot (1-q)^l + (1-q)^{k+1} \cdot (1-q)^k \\ &\geq \frac{1 - (1-q)^{2k}}{2} + (1-q)^{2k+1} = \frac{1}{2} + (1-q)^{2k} \left( \frac{1}{2} - q \right) \geq \frac{1}{2}, \end{aligned}$$

where the last inequality is by  $q = o(1)$ . By applying the above two inequalities to Eq. (22), we get

$$E^- \geq \frac{1}{en} \left( \sum_{j=1}^k \min \left\{ \frac{1}{4}, \frac{qj}{2} \right\} + \frac{n-i-k}{2} \right).$$

If  $k \geq \frac{n}{2}$ ,  $\sum_{j=1}^k \min \left\{ \frac{1}{4}, \frac{qj}{2} \right\} = \Omega(n)$  since  $q = \Omega(1/n)$ . If  $k < \frac{n}{2}$ ,  $\frac{n-i-k}{2} = \Omega(n)$  since  $i \leq \sqrt{n}$ . Thus,  $E^- = \Omega(1)$ .

For  $q = \Omega(1)$ , we use the trivial lower bound  $q$  for the probability of accepting the offspring solution  $x'$ , since it is sufficient to flip the first leading 1-bit of  $x$  by noise. Then,

$$E^- \geq \frac{1}{en} (kq + (n-i-k)q) = \frac{(n-i)q}{en} = \Omega(1).$$

Thus, for  $q = \Omega(1/n)$ , we have

$$E(X_t - X_{t+1} \mid X_t = i) = E^+ - E^- \leq \frac{i}{n} - \Omega(1) = -\Omega(1).$$

That is, condition (1) of Theorem 2 holds. Its condition (2) trivially holds with  $\delta = 1$  and  $r(l) = 2$ . Thus, the expected running time is exponential.  $\square$

### 4.3 One-bit Noise

For the (1+1)-EA on LeadingOnes under one-bit noise, it has been known that the expected running time is polynomial if  $p \leq 1/(6en^2)$  and exponential if  $p = 1/2$  [19]. We extend this result by proving in Theorem 14 that the expected running time is polynomial if  $p = O(\log n/n^2)$  and super-polynomial if  $p = \omega(\log n/n)$ . The proof can be accomplished in the same way as that of Theorems 8, 9 and 10 for bit-wise noise  $(p, \frac{1}{n})$ . This is because although the probabilities  $P(f^n(x') \geq f^n(x))$  of accepting the offspring solution are different, their bounds used in the proofs for bit-wise noise  $(p, \frac{1}{n})$  still hold for one-bit noise.

**Theorem 14** *For the (1+1)-EA on LeadingOnes under one-bit noise, the expected running time is polynomial if  $p = O(\log n/n^2)$ , super-polynomial if  $p = \omega(\log n/n) \cap o(1)$  and exponential if  $p = \Omega(1)$ .*

*Proof* We re-analyze  $P(f^n(x') \geq f^n(x))$  for one-bit noise, and show that the bounds on  $P(f^n(x') \geq f^n(x))$  used in the proofs for bit-wise noise  $(p, \frac{1}{n})$  still hold for one-bit noise.

For the proof of Theorem 8, Eqs. (7) and (8) change to

$$P(f^n(x') \leq i - 1) = p \frac{i}{n}, \quad P(f^n(x) \geq i + 1) = p \frac{1}{n},$$

and thus Eq. (9) still holds; Eqs. (12) and (13) change to

$$P(f^n(x') \geq i) \leq p \frac{1}{n}, \quad P(f^n(x) \leq i - 1) = p \frac{i}{n},$$

and thus Eq. (14) still holds.

For the proof of Theorem 9, Eqs. (18) and (19) change to

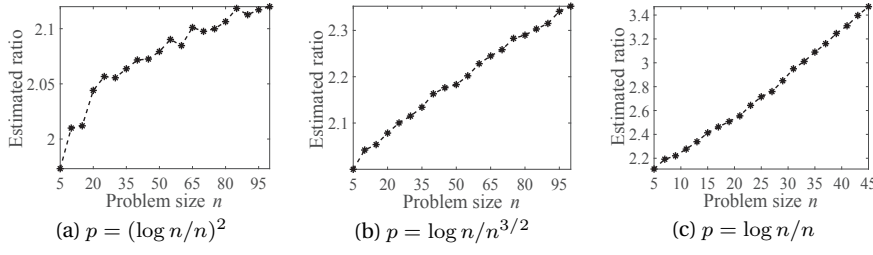
$$P(f^n(x') \geq j - 1) = 1 - p \frac{j - 1}{n}, \quad P(f^n(x) \leq j - 1) = p \frac{j}{n},$$

and thus Eq. (20) still holds.

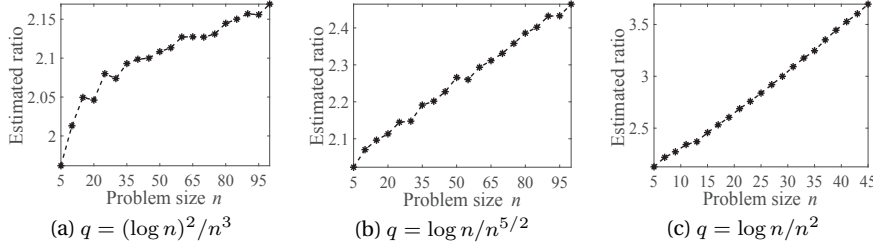
For the proof of Theorem 10, Eq. (28) still holds by the above two equalities; Eq. (29) still holds since the probability of keeping the first  $(k + 1)$  bits of a solution unchanged in one-bit noise is  $1 - p \frac{k+1}{n} \geq p(1 - \frac{k+1}{n})$ .  $\square$

### 4.4 Experiments

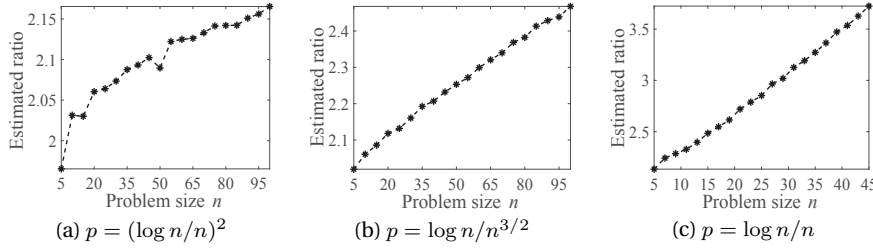
In the previous three subsections, we have proved that for the (1+1)-EA solving the LeadingOnes problem, if under bit-wise noise  $(p, \frac{1}{n})$ , the expected running time is polynomial when  $p = O(\log n/n^2)$  and super-polynomial when  $p = \omega(\log n/n)$ ; if under bit-wise noise  $(1, q)$ , the expected running time is polynomial when  $q = O(\log n/n^3)$  and super-polynomial when  $q = \omega(\log n/n^2)$ ; if under one-bit noise, the expected running time is polynomial when  $p = O(\log n/n^2)$  and super-polynomial when  $p = \omega(\log n/n)$ . However,



**Figure 1** Estimated expected running time for the (1+1)-EA on LeadingOnes under bit-wise noise  $(p, \frac{1}{n})$ , where the  $y$ -axis is (the logarithm of estimated expected running time) divided by  $\log n$ .



**Figure 2** Estimated expected running time for the (1+1)-EA on LeadingOnes under bit-wise noise  $(1, q)$ , where the  $y$ -axis is (the logarithm of estimated expected running time) divided by  $\log n$ .



**Figure 3** Estimated expected running time for the (1+1)-EA on LeadingOnes under one-bit noise, where the  $y$ -axis is (the logarithm of estimated expected running time) divided by  $\log n$ .

the current analysis does not cover all the ranges of  $p$  and  $q$ . We thus have conducted experiments to complement the theoretical results.

For bit-wise noise  $(p, \frac{1}{n})$ , we do not know whether the running time is polynomial or super-polynomial when  $p = \omega(\log n/n^2) \cap O(\log n/n)$ . We empirically estimate the expected running time for  $p = (\log n/n)^2$ ,  $\log n/n^{3/2}$  and  $\log n/n$ . On each problem size  $n$ , we run the (1+1)-EA 1000 times independently. In each run, we record the number of fitness evaluations until an optimal solution w.r.t. the true fitness function is found for the first time. Then the total number of evaluations of the 1000 runs are averaged as the estimation of the expected running time. To show the relationship between the estimated expected running time and the problem size  $n$  clearly, we plot the curve of  $\log(\text{estimated expected running time})/\log n$ , as shown in Figure 1. Note that in subfigures (a) and (b), the problem size  $n$  is in the range



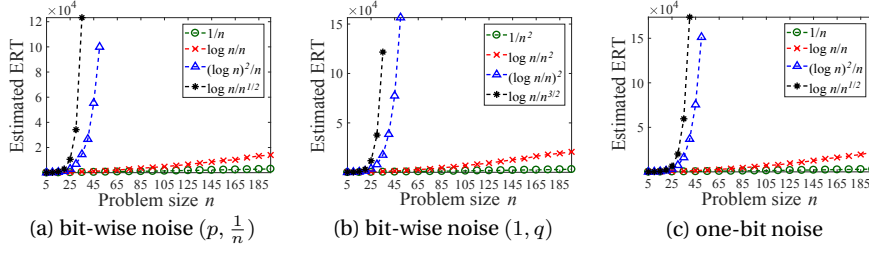
from 5 to 100, while in subfigure (c),  $n$  is from 5 to 45. This is because the expected running time with  $n > 45$  in subfigure (c) is too large to be estimated. We can observe that all the curves continue to rise as  $n$  increases, which suggests that the expected running time for the three tested  $p$  values is all  $n^{\omega(1)}$ , i.e., super-polynomial. For bit-wise noise  $(1, q)$  and one-bit noise, we also empirically estimate the expected running time for the values of  $q$  and  $p$ , which are uncovered by our theoretical analysis. The results are plotted in Figures 2 and 3, respectively, which are similar to that observed for bit-wise noise  $(p, \frac{1}{n})$ .

Therefore, these empirical results suggest that the expected running time is super-polynomial for the uncovered ranges of  $p$  and  $q$  in theoretical analysis, and thus the currently derived ranges of  $p$  and  $q$  allowing a polynomial running time might be tight. The rigorous analysis is not easy. We may need to analyze transition probabilities between fitness levels more precisely, and design an ingenious distance function or use more advanced analysis tools. We leave it as a future work.

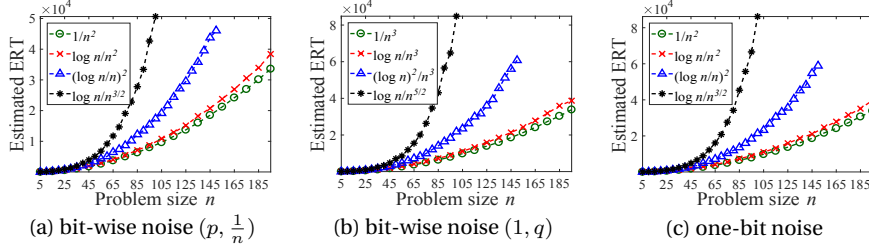
Since the theoretical results are all asymptotic, we also empirically compare the expected running time to see when the asymptotic behaviors can be clearly distinguished. For each problem and each kind of noise, we estimate the expected running time for the largest noise level (denoted by  $a$ ) allowing a polynomial running time derived in our theoretical analysis, and then compare it with the estimated expected running time for one relatively smaller noise level  $a/\log n$ , and two relatively larger noise levels  $a \cdot \log n$  and  $a \cdot \sqrt{n}$ . For example, for the OneMax problem under bit-wise noise  $(p, \frac{1}{n})$ , the largest noise level allowing a polynomial running time is  $O(\log n/n)$ ; thus we compare the estimated expected running time for  $p = 1/n$ ,  $\log n/n$ ,  $(\log n)^2/n$  and  $\log n/n^{1/2}$ . The results are plotted in Figures 4 and 5. We can observe that for the OneMax and LeadingOnes problems, the asymptotic polynomial behaviors (i.e., ‘green  $\circ$ ’ and ‘red  $\times$ ’) can be distinguished when the problem size  $n$  reaches nearly 200; while the asymptotic behaviors of polynomial (i.e., ‘green  $\circ$ ’ and ‘red  $\times$ ’) and super-polynomial (i.e., ‘blue  $\triangle$ ’ and ‘black  $*$ ’) can be clearly distinguished when  $n$  reaches 50 and 100, respectively.

## 5 The Robustness of Sampling to Noise

From the derived results in the above two sections, we can observe that the (1+1)-EA is efficient for solving OneMax and LeadingOnes only under low noise levels. For example, for the (1+1)-EA solving OneMax under bit-wise noise  $(p, \frac{1}{n})$ , the optimal solution can be found in polynomial time only when  $p = O(\log n/n)$ . In this section, we analyze the robustness of the sampling strategy to noise. Sampling as presented in Definition 5 evaluates the fitness of a solution multiple ( $m$ ) times independently and then uses the average to approximate the true fitness. We show that using sampling can significantly increase the largest noise level allowing a polynomial running



**Figure 4** Estimated expected running time (ERT) for the (1+1)-EA on OneMax under noise. Note that for the three studied noise models, the largest noise levels allowing a polynomial running time derived in theoretical analysis are  $O(\log n/n)$ ,  $O(\log n/n^2)$  and  $O(\log n/n)$ , respectively.



**Figure 5** Estimated expected running time (ERT) for the (1+1)-EA on LeadingOnes under noise. Note that for the three studied noise models, the largest noise levels allowing a polynomial running time derived in theoretical analysis are  $O(\log n/n^2)$ ,  $O(\log n/n^3)$  and  $O(\log n/n^2)$ , respectively.

time. For example, if using sampling with  $m = 4n^3$ , the (1+1)-EA can always solve OneMax under bit-wise noise  $(p, \frac{1}{n})$  in polynomial time, regardless of the value of  $p$ .

### 5.1 The OneMax Problem

We prove in Theorems 15 and 18 that under bit-wise noise  $(p, \frac{1}{n})$  or one-bit noise, the (1+1)-EA can always solve OneMax in polynomial time by using sampling. For bit-wise noise  $(1, q)$ , the tight range of  $q$  allowing a polynomial running time is  $1/2 - 1/n^{O(1)}$ , as shown in Theorems 16 and 17. Let  $x^k$  denote any solution with  $k$  number of 1-bits, and  $f^n(x^k)$  denote its noisy objective value. For proving polynomial upper bounds, we use Lemma 1, which gives a sufficient condition based on the probability  $P(f^n(x^j) < f^n(x^{k+1}))$  for  $j \leq k$ . But for the (1+1)-EA using sampling, the probability changes to be  $P(\hat{f}(x^j) < \hat{f}(x^{k+1}))$ , where  $\hat{f}(x^j) = \frac{1}{m} \sum_{i=1}^m f_i^n(x^j)$  as shown in Definition 5. Lemma 1 requires a lower bound on  $P(\hat{f}(x^j) < \hat{f}(x^{k+1}))$ . Our proof idea as presented in Lemma 3 is to derive a lower bound on the expectation of  $f^n(x^{k+1}) - f^n(x^j)$  and then apply Chebyshev's inequality. We will directly use Lemma 3 in the following proofs. For proving super-polynomial lower bounds, we use Lemma 2 by replacing  $P(f^n(x^k) < f^n(x^{k+1}))$  with  $P(\hat{f}(x^k) < \hat{f}(x^{k+1}))$ . Let  $poly(n)$  indicate any polynomial of  $n$ .

**Lemma 3** *Suppose there exists a real number  $\delta > 0$  such that*

$$\forall j \leq k < n : \mathbb{E}(f^n(x^{k+1}) - f^n(x^j)) \geq \delta,$$

*then the (1+1)-EA using sampling with  $m = n^3/\delta^2$  needs polynomial number of iterations in expectation for solving noisy OneMax.*

*Proof* We use Lemma 1 to prove it. For any  $j \leq k < n$ , let  $Y_{k,j} = f^n(x^{k+1}) - f^n(x^j)$  and  $\hat{Y}_{k,j} = \hat{f}(x^{k+1}) - \hat{f}(x^j)$ . We then need to analyze the probability  $\mathbb{P}(\hat{f}(x^j) < \hat{f}(x^{k+1})) = \mathbb{P}(\hat{Y}_{k,j} > 0)$ .

Denote the expectation  $\mathbb{E}(Y_{k,j})$  as  $\mu_{k,j}$  and the variance  $\text{Var}(Y_{k,j})$  as  $\sigma_{k,j}^2$ . It is easy to verify that  $\mathbb{E}(\hat{Y}_{k,j}) = \mu_{k,j}$  and  $\text{Var}(\hat{Y}_{k,j}) = \sigma_{k,j}^2/m$ . By Chebyshev's inequality, we have

$$\mathbb{P}(\hat{Y}_{k,j} \leq 0) \leq \mathbb{P}(|\hat{Y}_{k,j} - \mu_{k,j}| \geq \mu_{k,j}/2) \leq 4\sigma_{k,j}^2/(m\mu_{k,j}^2).$$

Since  $\mu_{k,j} \geq \delta > 0$ ,  $\sigma_{k,j}^2 = \mathbb{E}(Y_{k,j}^2) - \mu_{k,j}^2 \leq n^2$  and  $m = n^3/\delta^2$ , we have

$$\mathbb{P}(\hat{Y}_{k,j} \leq 0) \leq 4/n \leq \log n/(15n),$$

where the last inequality holds with sufficiently large  $n$ . Let  $l = \log n$ . Then,  $\mathbb{P}(\hat{Y}_{k,j} > 0) \geq 1 - \frac{\log n}{15n} > 1 - \frac{l}{n}$ . Let  $c = \frac{1}{15}$ . For  $k < n - l$ ,  $\mathbb{P}(\hat{Y}_{k,j} > 0) \geq 1 - c \frac{n-k}{n}$ . Thus, the condition of Lemma 1 (i.e., Eq. (1)) holds. We then get that the expected number of iterations is  $O(n \log n) + n2^{O(\log n)} = n^{O(1)}$ , i.e., polynomial.  $\square$

For bit-wise noise  $(p, \frac{1}{n})$ , we apply Lemma 3 to prove that the (1+1)-EA using sampling with  $m = 4n^3$  can always solve OneMax in polynomial time, regardless of the value of  $p$ .

**Theorem 15** *For the (1+1)-EA on OneMax under bit-wise noise  $(p, \frac{1}{n})$ , if using sampling with  $m = 4n^3$ , the expected running time is polynomial.*

*Proof* We use Lemma 3 to prove it. Since  $\mathbb{E}(f^n(x^j)) = j(1 - \frac{p}{n}) + (n-j)\frac{p}{n} = (1 - \frac{2p}{n})j + p$ , we have, for any  $j \leq k < n$ ,

$$\mathbb{E}(f^n(x^{k+1}) - f^n(x^j)) = \left(1 - \frac{2p}{n}\right)(k+1-j) \geq 1 - \frac{2p}{n} \geq 1/2,$$

where the last inequality holds with  $n \geq 4$ . Thus, by Lemma 3, we get that the expected number of iterations of the (1+1)-EA using sampling with  $m = 4n^3$  is polynomial. Since each iteration takes  $2m = 8n^3$  number of fitness evaluations, the expected running time is also polynomial.  $\square$

For bit-wise noise  $(1, q)$ , we prove in the following two theorems that by using sampling, the tight range of  $q$  allowing a polynomial running time is  $1/2 - 1/n^{O(1)}$ .

**Theorem 16** *For the (1+1)-EA on OneMax under bit-wise noise  $(1, q)$  with  $q = 1/2 - 1/n^{O(1)}$ , if using sampling, there exists some  $m = O(\text{poly}(n))$  such that the expected running time is polynomial.*

*Proof* We use Lemma 3 to prove it. Since  $q = 1/2 - 1/n^{O(1)}$ , there exists a positive constant  $c$  such that  $q \leq 1/2 - 1/n^c$ . It is easy to verify that  $E(f^n(x^j)) = j(1 - q) + (n - j)q = (1 - 2q)j + nq$ . Thus, for any  $j \leq k < n$ ,

$$E(f^n(x^{k+1}) - f^n(x^j)) = (1 - 2q)(k + 1 - j) \geq 1 - 2q \geq 2/n^c.$$

By Lemma 3, we get that if using sampling with  $m = n^{3+2c}/4$ , the expected number of iterations is polynomial, and then the expected running time is polynomial. Thus, the theorem holds.  $\square$

**Theorem 17** *For the (1+1)-EA on OneMax under bit-wise noise  $(1, q)$  with  $q = 1/2 - 1/n^{\omega(1)}$  or  $q \geq 1/2$ , if using sampling with any  $m = O(\text{poly}(n))$ , the expected running time is exponential.*

*Proof* We use Lemma 2 to prove it. Note that for the (1+1)-EA using sampling, we have to analyze  $P(\hat{f}(x^k) < \hat{f}(x^{k+1}))$  instead of  $P(f^n(x^k) < f^n(x^{k+1}))$ .

Let  $Z$  denote a random variable which satisfies that  $P(Z = 0) = q$  and  $P(Z = 1) = 1 - q$ . In the following proof, each  $Z_i$  is an independent random variable, which has the same distribution as  $Z$ . We have  $f^n(x^k) = \sum_{i=1}^k Z_i + \sum_{i=k+1}^n (1 - Z_i)$ , and then,

$$\begin{aligned} f^n(x^{k+1}) - f^n(x^k) &= \sum_{i=1}^{k+1} Z_i + \sum_{i=k+2}^n (1 - Z_i) - \sum_{i=1}^k Z_i - \sum_{i=k+1}^n (1 - Z_i) \\ &= \sum_{i=1}^{n+1} Z_i - \sum_{i=n+2}^{2n} Z_i - 1. \end{aligned}$$

Since  $\hat{f}(x^k) = \frac{1}{m} \sum_{i=1}^m f_i^n(x^k)$ , which is the average of  $m$  independent evaluations, we have

$$\begin{aligned} m(\hat{f}(x^{k+1}) - \hat{f}(x^k)) &= \sum_{j=0}^{m-1} \sum_{i=2nj+1}^{2nj+n+1} Z_i - \sum_{j=0}^{m-1} \sum_{i=2nj+n+2}^{2n(j+1)} Z_i - m \\ &= \sum_{j=0}^{m-1} \sum_{i=2nj+1}^{2nj+2} Z_i + \left( \sum_{j=0}^{m-1} \sum_{i=2nj+3}^{2nj+n+1} Z_i - \sum_{j=0}^{m-1} \sum_{i=2nj+n+2}^{2n(j+1)} Z_i \right) - m \\ &= \sum_{j=0}^{m-1} \sum_{i=2nj+1}^{2nj+2} Z_i + Z^* - m, \end{aligned}$$

where  $Z^* = \sum_{j=0}^{m-1} \sum_{i=2nj+3}^{2nj+n+1} Z_i - \sum_{j=0}^{m-1} \sum_{i=2nj+n+2}^{2nj+1} Z_i$ . To make  $\hat{f}(x^k) \geq \hat{f}(x^{k+1})$ , it is sufficient that  $Z^* \leq 0$  and  $\sum_{j=0}^{m-1} \sum_{i=2nj+1}^{2nj+2} Z_i \leq m$ . That is,

$$P(\hat{f}(x^k) \geq \hat{f}(x^{k+1})) \geq P(Z^* \leq 0) \cdot P\left(\sum_{j=0}^{m-1} \sum_{i=2nj+1}^{2nj+2} Z_i \leq m\right). \quad (34)$$

Since  $Z^*$  is the difference between the sum of the same number of  $Z_i$ ,  $Z^*$  has the same distribution as  $-Z^*$ . Thus,  $P(Z^* \leq 0) + P(Z^* \geq 0) = P(Z^* \leq 0) + P(-Z^* \leq 0) = 2P(Z^* \leq 0) \geq 1$ , which implies that

$$P(Z^* \leq 0) \geq 1/2. \quad (35)$$

We then investigate  $P(\sum_{j=0}^{m-1} \sum_{i=2nj+1}^{2nj+2} Z_i \leq m)$ . Since  $\sum_{j=0}^{m-1} \sum_{i=2nj+1}^{2nj+2} Z_i$  is the sum of  $2m$  independent random variables which have the same distribution as  $Z$ , we have

$$P\left(\sum_{j=0}^{m-1} \sum_{i=2nj+1}^{2nj+2} Z_i \leq m\right) = \sum_{t=0}^m \binom{2m}{t} (1-q)^t q^{2m-t},$$

$$P\left(\sum_{j=0}^{m-1} \sum_{i=2nj+1}^{2nj+2} Z_i > m\right) = \sum_{t=m+1}^{2m} \binom{2m}{t} (1-q)^t q^{2m-t} = \sum_{t=0}^{m-1} \binom{2m}{t} (1-q)^{2m-t} q^t.$$

For any  $t < m$ , let  $r = \frac{(1-q)^t q^{2m-t}}{(1-q)^{2m-t} q^t} = \left(\frac{q}{1-q}\right)^{2m-2t}$ . If  $q \geq 1/2$ , we have  $r \geq 1$ . If  $q = 1/2 - 1/n^{\omega(1)}$ , we have

$$r \geq \left(\frac{q}{1-q}\right)^{2m} = \left(1 - \frac{1-2q}{1-q}\right)^{2m}$$

$$\geq \left(1 - \frac{4}{n^{\omega(1)}}\right)^{2m} \geq \left(\frac{1}{e}\right)^{2m/(n^{\omega(1)}/4-1)} \geq \frac{1}{e},$$

where the first inequality is by  $q \leq 1/2$ , the second inequality is by  $1 - 2q = 2/n^{\omega(1)}$  and  $1 - q \geq 1/2$ , and the last is by  $m = O(\text{poly}(n))$ . Thus,  $P(\sum_{j=0}^{m-1} \sum_{i=2nj+1}^{2nj+2} Z_i \leq m) > 1/3 \cdot P(\sum_{j=0}^{m-1} \sum_{i=2nj+1}^{2nj+2} Z_i > m)$ , which implies that

$$P\left(\sum_{j=0}^{m-1} \sum_{i=2nj+1}^{2nj+2} Z_i \leq m\right) > 1/4. \quad (36)$$

By applying Eqs. (35) and (36) to Eq. (34), we get

$$P(\hat{f}(x^k) \geq \hat{f}(x^{k+1})) \geq 1/8.$$

Let  $c = 16$  and  $l = n/128$ . For any  $n - l \leq k < n$ ,

$$P(\hat{f}(x^k) < \hat{f}(x^{k+1})) = 1 - P(\hat{f}(x^k) \geq \hat{f}(x^{k+1})) \leq 1 - \frac{cl}{n} \leq 1 - \frac{c(n-k)}{n},$$

i.e., the condition of Lemma 2 holds. Thus, the expected number of iterations is  $2^{\Omega(n/128)}$ , and the expected running time is exponential.  $\square$

For one-bit noise, we show that using sampling with  $m = 4n^3$  is sufficient to make the (1+1)-EA solve OneMax in polynomial time.

**Theorem 18** *For the (1+1)-EA on OneMax under one-bit noise, if using sampling with  $m = 4n^3$ , the expected running time is polynomial.*

*Proof* It is easy to verify that the expectation of  $f^n(x^j)$  (i.e.,  $E(f^n(x^j))$ ) under one-bit noise is the same as that under bit-wise noise  $(p, \frac{1}{n})$ . Thus, the proof can be finished in the same way as that of Theorem 15.  $\square$

From the above analysis, we can intuitively explain why sampling is always effective for bit-wise noise  $(p, \frac{1}{n})$  and one-bit noise, while it fails for bit-wise noise  $(1, q)$  when  $q = 1/2 - 1/n^{\omega(1)}$  or  $q \geq 1/2$ . For two solutions  $x$  and  $y$  with  $f(x) > f(y)$ , if under bit-wise noise  $(p, \frac{1}{n})$  and one-bit noise, the noisy fitness  $f^n(x)$  is larger than  $f^n(y)$  in expectation, and using sampling will increase this trend and make the probability of accepting the true worse solution  $y$  sufficiently small. If under bit-wise noise  $(1, q)$ , when  $q = 1/2 - 1/n^{\omega(1)}$ , although the noisy fitness  $f^n(x)$  is still larger in expectation, the gap is very small (in the order of  $1/n^{\omega(1)}$ ) and a polynomial sample size is not sufficient to make the probability of accepting the true worse solution  $y$  small enough; when  $q \geq 1/2$ , the noisy fitness  $f^n(x)$  is smaller in expectation, and using sampling will increase this trend and it obviously does not work.

## 5.2 The LeadingOnes Problem

The bit-wise noise  $(p, \frac{1}{n})$  model is first considered. We prove in Theorem 19 that the (1+1)-EA using sampling can solve the LeadingOnes problem in polynomial time, regardless of the value of  $p$ . The proof idea is similar to that of Theorem 8. The main difference is the probability of accepting the offspring solution  $x'$ , which is changed from  $P(f^n(x') \geq f^n(x))$  to  $P(\hat{f}(x') \geq \hat{f}(x))$  due to sampling. Lemma 4 gives some bounds on this probability, which will be used in the proof of Theorem 19.

**Lemma 4** *For the LeadingOnes problem under bit-wise noise  $(p, \frac{1}{n})$ , if using sampling with  $m = 144n^6$ , it holds that*

- (1) *for any  $x$  with  $\text{LO}(x) = i < n$  and  $y$  with  $\text{LO}(y) \leq i - 2$  or  $\text{LO}(y) = i - 1 \wedge y_{i+1} = 0$ ,  $P(\hat{f}(x) \leq \hat{f}(y)) \leq 1/n^2$ .*
- (2) *for any  $y$  with  $\text{LO}(y) < n$ ,  $P(\hat{f}(1^n) \leq \hat{f}(y)) \leq 1/(4n^4)$ .*

*Proof* The proof is finished by deriving a lower bound on the expectation of  $f^n(x) - f^n(y)$  (which is equal to the expectation of  $\hat{f}(x) - \hat{f}(y)$ ) and then

applying Chebyshev's inequality. We first consider case (1). For any  $x$  with  $\text{LO}(x) = i < n$ ,

$$\begin{aligned} \mathbb{E}(f^n(x)) &\geq (1-p) \cdot i + \sum_{j=1}^i p \left(1 - \frac{1}{n}\right)^{j-1} \frac{1}{n} \cdot (j-1) \\ &\quad + p \left(1 - \frac{1}{n}\right)^i \frac{1}{n} \cdot (i+1) + p \left(1 - \frac{1}{n}\right)^{i+1} \cdot i, \\ \mathbb{E}(f^n(x)) &\leq (1-p) \cdot i + \sum_{j=1}^i p \left(1 - \frac{1}{n}\right)^{j-1} \frac{1}{n} \cdot (j-1) \\ &\quad + p \left(1 - \frac{1}{n}\right)^i \frac{1}{n} \cdot n + p \left(1 - \frac{1}{n}\right)^{i+1} \cdot i. \end{aligned} \quad (37)$$

Note that when flipping the first 0-bit of  $x$  and keeping the  $i$  leading 1-bits unchanged, the fitness is at least  $i+1$  and at most  $n$ . Then for any  $1 \leq i < n$ , we have

$$\begin{aligned} &\mathbb{E}(f^n(x) - f^n(y) \mid \text{LO}(x) = i \wedge \text{LO}(y) = i-1) \\ &\geq 1-p + p \left(1 - \frac{1}{n}\right)^{i-1} \frac{1}{n} \cdot (i-1) + \frac{p}{n} \left(1 - \frac{1}{n}\right)^{i-1} \cdot \left(\left(1 - \frac{1}{n}\right)(i+1) - n\right) \\ &\quad + p \left(1 - \frac{1}{n}\right)^i \left(\left(1 - \frac{1}{n}\right)i - (i-1)\right) \\ &= 1-p + p \left(1 - \frac{1}{n}\right)^{i-1} \left(\frac{i-1}{n} - \frac{1}{n^2}\right) \geq \begin{cases} -1/n^2, & \text{if } i=1 \\ 1/(6n), & \text{if } i \geq 2 \end{cases}. \end{aligned} \quad (38)$$

Thus, for any  $x$  with  $\text{LO}(x) = i$  and  $y$  with  $\text{LO}(y) \leq i-2$  (where  $2 \leq i < n$ ), letting  $z$  be any solution with  $\text{LO}(z) = \text{LO}(y) + 1$ , we have

$$\begin{aligned} \mathbb{E}(f^n(x) - f^n(y)) &= \mathbb{E}(f^n(x) - f^n(z)) + \mathbb{E}(f^n(z) - f^n(y)) \\ &\geq \frac{1}{6n} \cdot (i - \text{LO}(y) - 1) - \frac{1}{n^2} \geq \frac{1}{6n} - \frac{1}{n^2} \geq \frac{1}{12n}, \end{aligned} \quad (39)$$

where the first inequality is by repeatedly applying Eq. (38), the second inequality is by  $\text{LO}(y) \leq i-2$ , and the last holds with  $n \geq 12$ .

When  $\text{LO}(y) = i-1 \wedge y_{i+1} = 0$ , we have to re-analyze Eq. (38) to derive a tighter lower bound, because applying Eq. (38) directly will lead to a negative lower bound for  $i=1$ . We first derive a tighter upper bound on  $\mathbb{E}(f^n(y))$ . When flipping the  $i$ -th bit of  $y$  and keeping the  $i-1$  leading 1-bits unchanged, we can now further consider the flipping of the  $(i+1)$ -th bit since we know that  $y_{i+1} = 0$ , rather than directly using a trivial upper bound  $n$  on the noisy fitness. If  $y_{i+1}$  is not flipped,  $f^n(y) = i$ ; otherwise,  $f^n(y) \leq n$ . Thus, we get

$$\mathbb{E}(f^n(y) \mid \text{LO}(y) = i-1 \wedge y_{i+1} = 0)$$

$$\begin{aligned}
&\leq (1-p) \cdot (i-1) + \sum_{j=1}^{i-1} p \left(1 - \frac{1}{n}\right)^{j-1} \frac{1}{n} \cdot (j-1) \\
&\quad + p \left(1 - \frac{1}{n}\right)^{i-1} \frac{1}{n} \cdot \left(\frac{1}{n}n + \left(1 - \frac{1}{n}\right)i\right) + p \left(1 - \frac{1}{n}\right)^i \cdot (i-1).
\end{aligned}$$

By combining this inequality and the lower bound in Eq. (37), we have that, for any  $x$  with  $\text{LO}(x) = i$  and  $y$  with  $\text{LO}(y) = i-1 \wedge y_{i+1} = 0$  (where  $1 \leq i < n$ ),

$$\begin{aligned}
&\mathbb{E}(f^n(x) - f^n(y)) \tag{40} \\
&\geq 1-p + p \left(1 - \frac{1}{n}\right)^{i-1} \frac{1}{n} \cdot (i-1) + p \left(1 - \frac{1}{n}\right)^i \left(\left(1 - \frac{1}{n}\right)i - (i-1)\right) \\
&\quad + \frac{p}{n} \left(1 - \frac{1}{n}\right)^{i-1} \cdot \left(\left(1 - \frac{1}{n}\right)(i+1) - \left(i+1 - \frac{i}{n}\right)\right) \\
&= 1-p + p \left(1 - \frac{1}{n}\right)^{i-1} \left(1 - \frac{2}{n} + \frac{i-1}{n^2}\right) \\
&\geq 1-p + p \cdot \frac{1}{e} \cdot \frac{1}{2} \geq 1 - \frac{5}{6}p \geq \frac{1}{6},
\end{aligned}$$

where the second inequality holds with  $n \geq 4$ .

According to Eqs. (39) and (40), we have a unified lower bound  $1/(12n)$  on  $\mathbb{E}(f^n(x) - f^n(y))$ . Denote  $\mathbb{E}(f^n(x) - f^n(y))$  as  $\mu$  and  $\text{Var}(f^n(x) - f^n(y))$  as  $\sigma^2$ . We have  $\mu \geq 1/(12n)$  and  $\sigma^2 \leq n^2$  since  $|f^n(x) - f^n(y)| \leq n$ . As  $\hat{f}(x)$  is the average of  $m = 144n^6$  independent evaluations, it is easy to verify that  $\mathbb{E}(\hat{f}(x) - \hat{f}(y)) = \mu$  and  $\text{Var}(\hat{f}(x) - \hat{f}(y)) = \sigma^2/m$ . By Chebyshev's inequality,

$$P(\hat{f}(x) \leq \hat{f}(y)) \leq P(|\hat{f}(x) - \hat{f}(y) - \mu| \geq \mu) \leq \sigma^2/(m\mu^2) \leq 1/n^2. \tag{41}$$

Thus, case (1) holds.

For case (2), we first analyze  $\mathbb{E}(f^n(1^n) - f^n(1^{n-1}0))$ . The expectation on  $f^n(1^n)$  can be easily calculated as follows:

$$\mathbb{E}(f^n(1^n)) = (1-p) \cdot n + \sum_{j=1}^n p \left(1 - \frac{1}{n}\right)^{j-1} \frac{1}{n} \cdot (j-1) + p \left(1 - \frac{1}{n}\right)^n \cdot n.$$

Combining this equality with the upper bound in Eq. (37), we get

$$\begin{aligned}
&\mathbb{E}(f^n(1^n) - f^n(1^{n-1}0)) \tag{42} \\
&\geq 1-p + p \left(1 - \frac{1}{n}\right)^{n-1} \frac{1}{n} \cdot (n-1) + p \left(1 - \frac{1}{n}\right)^n - p \left(1 - \frac{1}{n}\right)^{n-1} \\
&= 1-p + p \left(1 - \frac{1}{n}\right)^{n-1} \left(1 - \frac{2}{n}\right) \geq \frac{1}{6}.
\end{aligned}$$

Then, for any  $y$  with  $\text{LO}(y) < n$ , we have

$$\mathbb{E}(f^n(1^n) - f^n(y)) = \mathbb{E}(f^n(1^n) - f^n(1^{n-1}0)) + \sum_{k=\text{LO}(y)+1}^{n-1} \mathbb{E}(f^n(z^k) - f^n(z^{k-1})) \geq \frac{1}{6},$$



where  $z^{n-1} = 1^{n-1}0$ ,  $z^k$  ( $\text{LO}(y) < k < n-1$ ) denotes one solution  $z$  with  $\text{LO}(z) = k$ ,  $z^{\text{LO}(y)} = y$ , and the inequality is by applying Eqs. (42) and (38). As the analysis for  $P(\hat{f}(x) \leq \hat{f}(y))$  in case (1) (i.e., Eq. (41)), we can similarly use Chebyshev's inequality to derive that, noting that  $\mu \geq 1/6$  here,

$$P(\hat{f}(1^n) \leq \hat{f}(y)) \leq \sigma^2 / (m\mu^2) \leq 1/(4n^4).$$

Thus, case (2) holds.  $\square$

The following theorem shows that for bit-wise noise  $(p, \frac{1}{n})$ , using sampling with  $m = 144n^6$  is sufficient to make the (1+1)-EA solve LeadingOnes in polynomial time.

**Theorem 19** *For the (1+1)-EA on LeadingOnes under bit-wise noise  $(p, \frac{1}{n})$ , if using sampling with  $m = 144n^6$ , the expected running time is polynomial.*

*Proof* We use Theorem 1 to prove it. We first construct a distance function  $V(x)$  as, for any  $x$  with  $\text{LO}(x) = i$ ,

$$V(x) = \left(1 + \frac{c}{n}\right)^n - \left(1 + \frac{c}{n}\right)^i,$$

where  $c = 13$ . Then, we investigate  $E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x)$  for any  $x$  with  $\text{LO}(x) < n$ . Assume that currently  $\text{LO}(x) = i$ , where  $0 \leq i \leq n-1$ . We divide the drift into two parts: positive  $E^+$  and negative  $E^-$ . That is,

$$E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x) = E^+ - E^-.$$

The positive drift  $E^+$  can be expressed as Eq. (4), except that  $P(f^n(x') \geq f^n(x))$  changes to  $P(\hat{f}(x') \geq \hat{f}(x))$  due to sampling. To derive a lower bound on  $E^+$ , we only consider that the  $(i+1)$ -th bit of  $x$  is flipped and the other bits keep unchanged, the probability of which is  $\frac{1}{n}(1 - \frac{1}{n})^{n-1}$ . The only difference between  $x'$  and  $x$  is the  $(i+1)$ -th bit and  $\text{LO}(x') > \text{LO}(x) = i$ . If  $\text{LO}(x') = n$ ,  $P(\hat{f}(x') \leq \hat{f}(x)) \leq 1/(4n^4)$  by case (2) of Lemma 4, and then  $P(\hat{f}(x') \geq \hat{f}(x)) \geq 1 - 1/(4n^4)$ . If  $\text{LO}(x') < n$ , it must hold that  $\text{LO}(x) \leq \text{LO}(x') - 1 \wedge x_{\text{LO}(x)+1} = x'_{\text{LO}(x')+1} = 0$ . By case (1) of Lemma 4,  $P(\hat{f}(x') \leq \hat{f}(x)) \leq 1/n^2$ , and then  $P(\hat{f}(x') \geq \hat{f}(x)) \geq 1 - 1/n^2$ . Thus, the probability of accepting the offspring solution  $x'$  is at least  $1/2$ . Since  $\text{LO}(x') > i$ ,  $V(x) - V(x') \geq (1 + \frac{c}{n})^{i+1} - (1 + \frac{c}{n})^i = \frac{c}{n}(1 + \frac{c}{n})^i$ . Then,  $E^+$  can be lower bounded as follows:

$$E^+ \geq \left(1 - \frac{1}{n}\right)^{n-1} \frac{1}{n} \cdot \frac{1}{2} \cdot \frac{c}{n} \left(1 + \frac{c}{n}\right)^i \geq \frac{c}{6n^2} \left(1 + \frac{c}{n}\right)^i.$$

For the negative drift  $E^-$ , we need to consider  $\text{LO}(x') < \text{LO}(x) = i$ . Since  $V(x') - V(x) \leq V(0^n) - V(x) = (1 + \frac{c}{n})^i - 1$ , Eq. (5) becomes

$$E^- \leq \left( \left(1 + \frac{c}{n}\right)^i - 1 \right) \sum_{x': \text{LO}(x') < i} P_{\text{mut}}(x, x') \cdot P(\hat{f}(x') \geq \hat{f}(x)).$$

We further divide  $\text{LO}(x') < i$  into two cases. If  $\text{LO}(x') \leq i - 2$  or  $\text{LO}(x') = i - 1 \wedge x'_{i+1} = 0$ , then  $P(\hat{f}(x') \geq \hat{f}(x)) \leq 1/n^2$  by case (1) of Lemma 4. If  $\text{LO}(x') = i - 1 \wedge x'_{i+1} = 1$ , then  $\sum_{x': \text{LO}(x')=i-1 \wedge x'_{i+1}=1} P_{\text{mut}}(x, x') \leq 1/n^2$  since it is necessary to flip the  $i$ -th and the  $(i + 1)$ -th bits of  $x$  in mutation. Then, we get

$$E^- \leq \left( \left(1 + \frac{c}{n}\right)^i - 1 \right) \cdot \left( 1 \cdot \frac{1}{n^2} + \frac{1}{n^2} \cdot 1 \right) \leq \frac{2}{n^2} \left(1 + \frac{c}{n}\right)^i.$$

By subtracting  $E^-$  from  $E^+$ , we have, noting that  $c = 13$ ,

$$E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x) \geq \left(1 + \frac{c}{n}\right)^i \cdot \left(\frac{c}{6n^2} - \frac{2}{n^2}\right) \geq \frac{1}{6n^2}.$$

Since  $V(x) \leq (1 + \frac{13}{n})^n \leq e^{13} = O(1)$ , we have  $E(\tau \mid \xi_0) = O(n^2)$  by Theorem 1. Each iteration of the (1+1)-EA using sampling takes  $2m = 288n^6$  number of fitness evaluations, thus the expected running time is polynomial.  $\square$

For bit-wise noise  $(1, q)$ , we prove in Theorems 20 and 21 that the expected running time is polynomial if and only if  $q = O(\log n/n)$ . The proof of Theorem 20 is similar to that of Theorem 19, which considers bit-wise noise  $(p, \frac{1}{n})$ . The main difference is the probability of accepting the offspring solution  $x'$  (i.e.,  $P(\hat{f}(x') \geq \hat{f}(x))$ ), due to the change of noise. Lemma 5 gives some bounds on this probability, which will be used in the proof of Theorem 20.

**Lemma 5** *For the LeadingOnes problem under bit-wise noise  $(1, q)$  with  $q \leq c_0 \log n/n$  (where  $c_0$  is a positive constant), if using sampling with  $m = 36n^{2c_0+4}$ , it holds that*

- (1) *for any  $x$  with  $\text{LO}(x) = i < n$  and  $y$  with  $\text{LO}(y) \leq i - 5c_0 \log n$  or  $i - 5c_0 \log n < \text{LO}(y) \leq i - 1 \wedge y_{i+1} = 0$ ,  $P(\hat{f}(x) \leq \hat{f}(y)) \leq 1/n^2$ .*
- (2) *for any  $y$  with  $\text{LO}(y) < n$ ,  $P(\hat{f}(1^n) \leq \hat{f}(y)) \leq 1/n^2$ .*

*Proof* The proof is finished by deriving a lower bound on the expectation of  $f^n(x) - f^n(y)$  and then applying Chebyshev's inequality. We first consider case (1). For any  $x$  with  $\text{LO}(x) = i < n$ ,

$$\begin{aligned} E(f^n(x)) &\geq \sum_{j=1}^i (1-q)^{j-1} q \cdot (j-1) + (1-q)^i q \cdot (i+1) + (1-q)^{i+1} \cdot i, \\ E(f^n(x)) &\leq \sum_{j=1}^i (1-q)^{j-1} q \cdot (j-1) + (1-q)^i q \cdot n + (1-q)^{i+1} \cdot i. \end{aligned} \tag{43}$$

By applying these two inequalities, we get, for any  $k < i < n$ ,

$$\begin{aligned} E(f^n(x) - f^n(y) \mid \text{LO}(x) = i \wedge \text{LO}(y) = k) \\ \geq \frac{1}{q} ((i-1)(1-q)^{i+1} - i(1-q)^i) + q(1-q)^i \cdot (i+1) + (1-q)^{i+1} \cdot i \end{aligned} \tag{44}$$

$$\begin{aligned}
& -\frac{1}{q}((k-1)(1-q)^{k+1} - k(1-q)^k) - q(1-q)^k \cdot n - (1-q)^{k+1} \cdot k \\
& = (1-q)^i \left( q + 1 - \frac{1}{q} \right) - (1-q)^k \left( 1 - \frac{1}{q} + qn - qk \right) \\
& = (1-q)^k \left( \left( \frac{1}{q} - q - 1 \right) (1 - (1-q)^{i-k}) + q + qk - qn \right).
\end{aligned}$$

Thus, for any  $x$  with  $\text{LO}(x) = i$  and  $y$  with  $\text{LO}(y) = k \leq i - 5c_0 \log n$  (where  $5c_0 \log n \leq i < n$ ),

$$\begin{aligned}
\mathbb{E}(f^n(x) - f^n(y)) & \geq (1-q)^k \left( \left( \frac{1}{q} - 2 \right) (1 - (1-q)^{5c_0 \log n}) - qn \right) \\
& \geq (1-q)^k \left( \left( \frac{1}{q} - 2 \right) \left( 1 - \frac{1}{e^{5c_0 q \log n}} \right) - qn \right) \\
& \geq (1-q)^k \left( \left( \frac{1}{q} - 2 \right) \left( 1 - \frac{1}{1 + 5c_0 q \log n} \right) - qn \right) \\
& \geq (1-q)^k \left( \left( \frac{1}{q} - 2 \right) \frac{5c_0 q \log n}{2} - qn \right) \\
& \geq (1-q)^k \left( \frac{5}{2} c_0 \log n - 5 \frac{(c_0 \log n)^2}{n} - c_0 \log n \right) \\
& = (1-q)^k \left( \frac{3}{2} c_0 \log n - o(1) \right) \geq \frac{c_0 \log n}{en^{c_0}}, \tag{45}
\end{aligned}$$

where the third inequality is by  $e^x \geq 1 + x$ , the fourth is by  $5c_0 q \log n \leq 5(c_0 \log n)^2/n \leq 1$  for sufficiently large  $n$ , the fifth is by  $q \leq c_0 \log n/n$ , and the last is by

$$(1-q)^k \geq \left( 1 - \frac{c_0 \log n}{n} \right)^{n-1} \geq \left( \frac{1}{e} \right)^{\frac{n-1}{n/(c_0 \log n)-1}} \geq \left( \frac{1}{e} \right)^{c_0 \log n + 1} = \frac{1}{en^{c_0}}. \tag{46}$$

When  $i - 5c_0 \log n < \text{LO}(y) \leq i - 1 \wedge y_{i+1} = 0$  (where  $i \geq 1$ ), we calculate  $\mathbb{E}(f^n(x) - f^n(y))$  by

$$\mathbb{E}(f^n(x) - f^n(y)) = \sum_{k=\text{LO}(y)+1}^i \mathbb{E}(f^n(z^k) - f^n(z^{k-1})),$$

where  $z^i = x$ ,  $z^k$  ( $\text{LO}(y) < k < i$ ) denotes one solution  $z$  with  $\text{LO}(z) = k \wedge z_{i+1} = 0$ , and  $z^{\text{LO}(y)} = y$ . We then give a lower bound on  $\mathbb{E}(f^n(z^k) - f^n(z^{k-1}))$ , where  $\text{LO}(y) + 1 \leq k \leq i$ . For  $\mathbb{E}(f^n(z^k))$ , we directly use the lower bound in Eq. (43) to get that

$$\mathbb{E}(f^n(z^k)) \geq \sum_{j=1}^k (1-q)^{j-1} q \cdot (j-1) + (1-q)^k q \cdot (k+1) + (1-q)^{k+1} \cdot k.$$

For  $E(f^n(z^{k-1}))$ , instead of directly using the upper bound in Eq. (43), we derive a tighter one:

$$E(f^n(z^{k-1})) \leq \sum_{j=1}^{k-1} (1-q)^{j-1} q \cdot (j-1) + (1-q)^{k-1} q (q \cdot n + (1-q) \cdot i) + (1-q)^k \cdot (k-1).$$

Note that the inequality is because when the  $k-1$  leading 1-bits of  $z^{k-1}$  keep unchanged and its  $k$ -th bit (which must be 0) is flipped, if the  $(i+1)$ -th bit (which is 0) is flipped,  $f^n(z^{k-1}) \leq n$ ; otherwise,  $f^n(z^{k-1}) \leq i$ . By combining the above two inequalities, we get

$$\begin{aligned} E(f^n(z^k) - f^n(z^{k-1})) &\geq (1-q)^{k-1} (1 + q(k-i-1) - q^2(1+n-i)) \\ &= (1-q)^{k-1} (1 - o(1)) \geq 1/(2en^{c_0}), \end{aligned}$$

where the equality by  $k \geq \text{LO}(y) + 1 > i - 5c_0 \log n + 1$  and  $q = O(\log n/n)$ , and the last is by Eq. (46). Thus, we have

$$E(f^n(x) - f^n(y)) \geq (i - \text{LO}(y)) \cdot \frac{1}{2en^{c_0}} \geq \frac{1}{2en^{c_0}}. \quad (47)$$

For case (1), by combining Eqs. (45) and (47), we get a unified lower bound  $1/(2en^{c_0})$  on  $E(f^n(x) - f^n(y))$ . As the analysis for  $P(\hat{f}(x) \leq \hat{f}(y))$  (i.e., Eq. (41)) in the proof of Lemma 4, we can similarly use Chebyshev's inequality to derive that, noting that  $m = 36n^{2c_0+4}$  here,

$$P(\hat{f}(x) \leq \hat{f}(y)) \leq \sigma^2/(m\mu^2) \leq 1/n^2, \quad (48)$$

where  $\mu$  and  $\sigma^2$  are the expectation and variance of  $f^n(x) - f^n(y)$ , respectively. Thus, case (1) holds.

Then, we consider case (2), that is, we are to analyze  $P(\hat{f}(1^n) \leq \hat{f}(y))$  with  $\text{LO}(y) < n$ . We calculate  $E(f^n(1^n) - f^n(y))$  as follows:

$$E(f^n(1^n) - f^n(y)) = E(f^n(1^n) - f^n(1^{n-1}0)) + E(f^n(1^{n-1}0) - f^n(y)).$$

It is easy to derive that

$$\begin{aligned} E(f^n(1^n)) &= \sum_{j=1}^n (1-q)^{j-1} q \cdot (j-1) + (1-q)^n \cdot n, \\ E(f^n(1^{n-1}0)) &= \sum_{j=1}^{n-1} (1-q)^{j-1} q \cdot (j-1) + (1-q)^{n-1} q \cdot n + (1-q)^n \cdot (n-1). \end{aligned}$$

Then, we have

$$E(f^n(1^n) - f^n(1^{n-1}0)) = (1-q)^{n-1} (1-2q) \geq 1/(2en^{c_0}),$$

where the inequality is by Eq. (46) and  $q = O(\log n/n)$ . If  $\text{LO}(y) \leq n - 1 - 5c_0 \log n$ , by Eq. (45), we directly have  $E(f^n(1^{n-1}0) - f^n(y)) \geq \frac{c_0 \log n}{en^{c_0}} \geq 0$ . If  $\text{LO}(y) \geq n - 5c_0 \log n$ ,  $E(f^n(1^{n-1}0) - f^n(y))$  is calculated as follows:

$$E(f^n(1^{n-1}0) - f^n(y)) = \sum_{k=\text{LO}(y)+1}^{n-1} E(f^n(z^k) - f^n(z^{k-1})),$$

where  $z^{n-1} = 1^{n-1}0$ ,  $z^k$  ( $\text{LO}(y) < k < n - 1$ ) denotes one solution  $z$  with  $\text{LO}(z) = k$ , and  $z^{\text{LO}(y)} = y$ . By Eq. (44), we have, for any  $\text{LO}(y) + 1 \leq k < n$ ,

$$\begin{aligned} E(f^n(z^k) - f^n(z^{k-1})) &\geq (1 - q)^{k-1}(1 - q^2 + q(k - 1 - n)) \\ &\geq (1 - q)^{k-1}(1 - q^2 - 5c_0 q \log n) = (1 - q)^{k-1}(1 - o(1)) \geq 0, \end{aligned}$$

which implies that  $E(f^n(1^{n-1}0) - f^n(y)) \geq 0$ . Then, we get

$$E(f^n(1^n) - f^n(y)) = E(f^n(1^n) - f^n(1^{n-1}0)) + E(f^n(1^{n-1}0) - f^n(y)) \geq \frac{1}{2en^{c_0}}.$$

As the analysis in case (1) (i.e., Eq. (48)), we can get that

$$P(\hat{f}(1^n) \leq \hat{f}(y)) \leq 1/n^2.$$

Thus, case (2) holds.  $\square$

The following theorem shows that by using sampling, the (1+1)-EA can solve LeadingOnes under bit-wise noise  $(1, q)$  in polynomial time when  $q$  is in the range of  $O(\log n/n)$ .

**Theorem 20** *For the (1+1)-EA on LeadingOnes under bit-wise noise  $(1, q)$  with  $q = O(\log n/n)$ , if using sampling, there exists some  $m = O(\text{poly}(n))$  such that the expected running time is polynomial.*

*Proof* Since  $q = O(\log n/n)$ , there exists a positive constant  $c_0$  such that for all  $n$  large enough,  $q \leq c_0 \log n/n$ . We prove that if using sampling with  $m = 36n^{2c_0+4}$ , the expected running time is polynomial.

The proof is similar to that of Theorem 19. The distance function  $V(x)$  is defined as, for any  $x$  with  $\text{LO}(x) = i$ ,  $V(x) = (1 + \frac{c}{n})^n - (1 + \frac{c}{n})^i$ , where  $c = 30c_0 \log n + 7$ . Assume that currently  $\text{LO}(x) = i$ , where  $0 \leq i \leq n - 1$ . For the positive drift  $E^+$ , we consider that only the  $(i+1)$ -th bit (i.e., the first 0-bit) of  $x$  is flipped in mutation. If  $\text{LO}(x') = n$ ,  $P(\hat{f}(x') \leq \hat{f}(x)) \leq 1/n^2$  by case (2) of Lemma 5. If  $\text{LO}(x') < n$ , it must hold that  $\text{LO}(x) \leq \text{LO}(x') - 1 \wedge x_{\text{LO}(x)+1} = x'_{\text{LO}(x')+1} = 0$ , since  $x$  and  $x'$  are the same except the  $(\text{LO}(x)+1)$ -th bit. By case (1) of Lemma 5,  $P(\hat{f}(x') \leq \hat{f}(x)) \leq 1/n^2$ . Thus, the probability of accepting the offspring solution  $x'$  is  $P(\hat{f}(x') \geq \hat{f}(x)) \geq 1 - 1/n^2 \geq 1/2$ . The positive drift then can be lower bounded by

$$E^+ \geq \left(1 - \frac{1}{n}\right)^{n-1} \frac{1}{n} \cdot \frac{1}{2} \cdot \frac{c}{n} \left(1 + \frac{c}{n}\right)^i \geq \frac{c}{6n^2} \left(1 + \frac{c}{n}\right)^i.$$

For the negative drift  $E^-$ , we need to consider  $LO(x') < LO(x) = i$ . We further divide  $LO(x') < i$  into two cases. If  $LO(x') \leq i - 5c_0 \log n$  or  $i - 5c_0 \log n < LO(x') \leq i - 1 \wedge x'_{i+1} = 0$ , then  $P(\hat{f}(x') \geq \hat{f}(x)) \leq 1/n^2$  by case (1) of Lemma 5. If  $i - 5c_0 \log n < LO(x') \leq i - 1 \wedge x'_{i+1} = 1$ , we consider the probability of generating  $x'$  by mutation on  $x$ . Since it is necessary to flip the  $(i + 1)$ -th bit of  $x$  and at least one 1-bit in positions  $i - 5c_0 \log n + 2$  to  $i$  simultaneously,

$$\sum_{x': i - 5c_0 \log n < LO(x') \leq i - 1 \wedge x'_{i+1} = 1} P_{mut}(x, x') \leq \frac{5c_0 \log n}{n^2}.$$

Then, we get

$$E^- \leq \left( \left( 1 + \frac{c}{n} \right)^i - 1 \right) \cdot \left( 1 \cdot \frac{1}{n^2} + \frac{5c_0 \log n}{n^2} \cdot 1 \right) \leq \frac{5c_0 \log n + 1}{n^2} \left( 1 + \frac{c}{n} \right)^i.$$

By subtracting  $E^-$  from  $E^+$ , we have, noting that  $c = 30c_0 \log n + 7$ ,

$$E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x) \geq \left( 1 + \frac{c}{n} \right)^i \cdot \left( \frac{c}{6n^2} - \frac{5c_0 \log n + 1}{n^2} \right) \geq \frac{1}{6n^2}.$$

Note that  $V(x) \leq (1 + \frac{c}{n})^n \leq e^c = e^{30c_0 \log n + 7} = n^{O(1)}$ . By Theorem 1, we have  $E(\tau \mid \xi_0) \leq 6n^2 \cdot n^{O(1)}$ . Since each iteration of the (1+1)-EA using sampling takes  $2m = 72n^{2c_0+4}$  number of fitness evaluations, the expected running time is polynomial.  $\square$

For bit-wise noise  $(1, q)$  with  $q = \omega(\log n/n)$ , we apply the negative drift theorem (i.e., Theorem 2) to prove that using sampling still cannot guarantee a polynomial running time.

**Theorem 21** *For the (1+1)-EA on LeadingOnes under bit-wise noise  $(1, q)$  with  $q = \omega(\log n/n)$ , if using sampling with any  $m = O(\text{poly}(n))$ , the expected running time is exponential.*

*Proof* We use Theorem 2 to prove it. Let  $X_t = |x|_0$  be the number of 0-bits of the solution  $x$  after  $t$  iterations of the algorithm. We consider the interval  $[0, n/50]$ , that is, the parameters  $a = 0$  and  $b = n/50$  in Theorem 2. Then, we analyze the drift  $E(X_t - X_{t+1} \mid X_t = i)$  for  $1 \leq i < n/50$ . As in the proof of Theorem 9, we divide the drift into two parts: positive  $E^+$  and negative  $E^-$ . That is,

$$E(X_t - X_{t+1} \mid X_t = i) = E^+ - E^-.$$

For the positive drift, we can use the same analysis as that (i.e., Eq. (17)) in the proof of Theorem 9 to derive that  $E^+ \leq i/n < 1/50$ . This is because the offspring solution  $x'$  is optimistically assumed to be always accepted in the analysis of Eq. (17), and thus the change of noise and the use of sampling will not affect the analysis.

For the negative drift, we need to consider that the number of 0-bits is increased. To derive a lower bound on  $E^-$ , we only consider the  $n - i$  cases

where only one 1-bit of  $x$  is flipped, which happens with probability  $\frac{1}{n}(1 - \frac{1}{n})^{n-1}$ . Let  $x^j$  denote the solution that is generated by flipping only the  $j$ -th bit of  $x$ . Then, we have

$$E^- \geq \sum_{j:x_j=1} \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1} \cdot P(\hat{f}(x^j) \geq \hat{f}(x)) \cdot (i+1-i).$$

We then investigate  $P(\hat{f}(x^j) \geq \hat{f}(x))$ . Let  $F(y)$  denote the event that when evaluating the noisy fitness of a solution  $y$ , at least one 1-bit in its first  $n/3$  positions is flipped by noise. Note that there must exist 1-bits in the first  $n/3$  positions of  $x$ , since  $|x|_0 = i < n/50$ . For any  $x^j$  with  $j > n/3$ , its first  $n/3$  bits are the same as that of  $x$ . If both the events  $F(x)$  and  $F(x^j)$  happen,  $f^n(x) < n/3 \wedge f^n(x^j) < n/3$ , and the last  $(2n)/3$  bits of  $x$  and  $x^j$  will not affect their noisy fitness. Thus, for any  $x^j$  with  $j > n/3$ ,  $f^n(x^j)$  has the same distribution as  $f^n(x)$  conditioned on  $F(x) \cap F(x^j)$ . When estimating  $\hat{f}(y)$  of a solution  $y$  by sampling, let  $F_t(y)$  denote the event  $F(y)$  in the  $t$ -th independent noisy evaluation of  $y$ . Thus, for all  $j > n/3$ ,  $\sum_{t=1}^m f_t^n(x)$  and  $\sum_{t=1}^m f_t^n(x^j)$  have the same distribution conditioned on  $(\cap_{t=1}^m F_t(x)) \cap (\cap_{t=1}^m F_t(x^j))$ . Since  $\hat{f}(x) = \frac{1}{m} \sum_{t=1}^m f_t^n(x)$  and  $\hat{f}(x^j) = \frac{1}{m} \sum_{t=1}^m f_t^n(x^j)$ , we have

$$P(\hat{f}(x^j) \geq \hat{f}(x) \mid (\cap_{t=1}^m F_t(x)) \cap (\cap_{t=1}^m F_t(x^j))) \geq 1/2.$$

Since  $|x|_0 = i < n/50$ , there are at least  $n/3 - n/50 \geq n/4$  number of 1-bits in the first  $n/3$  positions of  $x$ , which implies that the probability  $P(F_t(x))$  of the event  $F_t(x)$  happening is at least  $1 - (1-q)^{n/4}$ . Furthermore,  $P(F_t(x^j)) = P(F_t(x))$  for  $j > n/3$ , since  $x$  and  $x^j$  have the same first  $n/3$  bits. Thus,

$$\begin{aligned} P((\cap_{t=1}^m F_t(x)) \cap (\cap_{t=1}^m F_t(x^j))) &\geq (1 - (1-q)^{n/4})^{2m} \\ &\geq \left(1 - \frac{1}{e^{nq/4}}\right)^{2m} \geq \left(\frac{1}{e}\right)^{2m/(e^{nq/4}-1)} = \left(\frac{1}{e}\right)^{2m/n^{\omega(1)}} \geq \frac{1}{e}, \end{aligned}$$

where the equality is by  $q = \omega(\log n/n)$  and the last inequality is by  $m = O(\text{poly}(n))$ . By the law of total probability, we have, for all  $j > n/3$ ,

$$\begin{aligned} P(\hat{f}(x^j) \geq \hat{f}(x)) &\geq P(\hat{f}(x^j) \geq \hat{f}(x) \mid (\cap_{t=1}^m F_t(x)) \cap (\cap_{t=1}^m F_t(x^j))) \\ &\quad \cdot P((\cap_{t=1}^m F_t(x)) \cap (\cap_{t=1}^m F_t(x^j))) \geq \frac{1}{2e}. \end{aligned}$$

Then, we can get a lower bound on the negative drift:

$$\begin{aligned} E^- &\geq \sum_{j:j>n/3 \wedge x_j=1} \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1} \cdot P(\hat{f}(x^j) \geq \hat{f}(x)) \cdot 1 \\ &\geq \left(\frac{2n}{3} - \frac{n}{50}\right) \cdot \frac{1}{en} \cdot \frac{1}{2e} \geq \frac{1}{36}. \end{aligned}$$

By subtracting  $E^-$  from  $E^+$ , we have, for  $1 \leq i < n/50$ ,

$$E(X_t - X_{t+1} \mid X_t = i) = E^+ - E^- \leq \frac{1}{50} - \frac{1}{36}.$$

Thus, condition (1) of Theorem 2 holds. It is easy to verify that condition (2) of Theorem 2 holds with  $\delta = 1$  and  $r(l) = 2$ . Note that  $l = b - a = n/50$ . By Theorem 2, we get that the expected number of iterations is exponential, and then the expected running time is also exponential.  $\square$

For the one-bit noise model, we prove in Theorem 22 that the (1+1)-EA using sampling can always solve the LeadingOnes problem in polynomial time. The proof is finished by applying the additive drift theorem. Lemma 6 gives some bounds on the probability  $P(\hat{f}(x') - \hat{f}(x))$  of accepting the off-spring solution  $x'$ , which will be used in the proof of Theorem 22.

From case (2) of Lemma 6, we can observe that when the solution is close to the optimum  $1^n$ , the probability of accepting  $1^n$  is small, which is different from the situation in bit-wise noise (as shown in case (2) of Lemmas 4 and 5). If directly using the distance function constructed in the proof of Theorems 19 and 20, this small acceptance probability will make the positive drift  $E^+$  not large enough, and then the condition of the additive drift theorem is unsatisfied. To address this issue, our idea is to re-design the distance function such that the distance from non-optimal solutions to the optimum  $1^n$  is much larger than that between non-optimal solutions. Then, the small probability of accepting  $1^n$  can be compensated by the significant decrease on the distance after accepting  $1^n$ ; thus the positive drift can still be large enough to make the condition of the additive drift theorem hold.

Note that in the proof of Lemma 6, we use Berry-Esseen inequality [31] and Hoeffding's inequality, instead of Chebyshev's inequality used in the proof of Lemmas 4 and 5. When the solution  $x$  is close to the optimum  $1^n$ , the expectation of  $f^n(1^n) - f^n(x)$  is lower bounded by a negative value. Thus, for deriving a lower bound on the probability  $P(\hat{f}(1^n) \geq \hat{f}(x))$ , Chebyshev's inequality fails, while we apply Berry-Esseen inequality [31]. The analysis shows that a moderate sample size  $m = 4n^4 \log n/15$  can make this probability not too small. With this sample size, to derive a small enough upper bound on the probability  $P(\hat{f}(x) \leq \hat{f}(y))$  for two solutions  $x$  and  $y$  with  $E(f^n(x) - f^n(y)) > 0$ , we have to use Hoeffding's inequality, which is tighter than Chebyshev's inequality.

**Lemma 6** *For the LeadingOnes problem under one-bit noise, if using sampling with  $m = 4n^4 \log n/15$ , it holds that*

- (1) *for any  $x$  with  $LO(x) = i < n$  and  $y$  with  $LO(y) \leq i - 2$  or  $LO(y) = i - 1 \wedge y_{i+1} = 0$ ,  $P(\hat{f}(x) \leq \hat{f}(y)) \leq 2n^{-\frac{2}{15}}$ ; furthermore, if  $i \leq n - 4$ ,  $P(\hat{f}(x) \leq \hat{f}(y)) \leq 2n^{-\frac{32}{15}}$ .*
- (2) *for any  $y$ , if  $LO(y) \leq n - 4$ ,  $P(\hat{f}(1^n) \leq \hat{f}(y)) \leq 2n^{-\frac{8}{15}}$ ; if  $LO(y) \in \{n - 3, n - 2, n - 1\}$ ,  $P(\hat{f}(1^n) \geq \hat{f}(y)) \geq 1/n^2$ .*



*Proof* The proof is finished by deriving a lower bound on the expectation of  $f^n(x) - f^n(y)$  and then applying Hoeffding's inequality or Berry-Esseen inequality [31]. We first consider case (1). For any  $x$  with  $\text{LO}(x) = i < n$ ,

$$\mathbb{E}(f^n(x)) \geq (1-p) \cdot i + \sum_{j=1}^i \frac{p}{n} \cdot (j-1) + \frac{p}{n} \cdot (i+1) + p \frac{n-i-1}{n} \cdot i, \quad (49)$$

$$\mathbb{E}(f^n(x)) \leq (1-p) \cdot i + \sum_{j=1}^i \frac{p}{n} \cdot (j-1) + \frac{p}{n} \cdot n + p \frac{n-i-1}{n} \cdot i.$$

By applying these two inequalities, we have, for any  $k < i < n$ ,

$$\begin{aligned} & \mathbb{E}(f^n(x) - f^n(y) \mid \text{LO}(x) = i \wedge \text{LO}(y) = k) \\ & \geq (1-p)(i-k) + \frac{p}{n} \left( (i-k) \left( n - \frac{i+k+3}{2} \right) + i+1-n \right). \end{aligned} \quad (50)$$

Thus, for any  $x$  with  $\text{LO}(x) = i$  and  $y$  with  $\text{LO}(y) = k \leq i-2$  (where  $2 \leq i < n$ ), we have

$$\mathbb{E}(f^n(x) - f^n(y)) \geq 2(1-p) + \frac{p}{n}(n-i) \geq \frac{n-i}{n}.$$

When  $\text{LO}(y) = i-1 \wedge y_{i+1} = 0$ , if we directly use Eq. (50), we will get a lower bound  $1-p$ , which is 0 for  $p = 1$ . Since  $y_{i+1} = 0$ , we actually can get an exact value of  $\mathbb{E}(f^n(y))$ :

$$\mathbb{E}(f^n(y)) = (1-p) \cdot (i-1) + \sum_{j=1}^{i-1} \frac{p}{n} \cdot (j-1) + \frac{p}{n} \cdot i + p \frac{n-i}{n} \cdot (i-1).$$

By combining this equality and the lower bound in Eq. (49), we have that, for any  $x$  with  $\text{LO}(x) = i$  and  $y$  with  $\text{LO}(y) = i-1 \wedge y_{i+1} = 0$  (where  $1 \leq i < n$ ),

$$\mathbb{E}(f^n(x) - f^n(y)) \geq 1-p + \frac{p}{n}(i-1) + \frac{p}{n} + \frac{p}{n}(n-2i) \geq \frac{n-i}{n}.$$

Thus, we have a unified lower bound  $(n-i)/n$  on  $\mathbb{E}(f^n(x) - f^n(y))$  for case (1). Denote  $\mathbb{E}(f^n(x) - f^n(y))$  as  $\mu$ . We have  $\mu \geq (n-i)/n \geq 1/n$ . Since  $\hat{f}(x)$  is the average of  $m = 4n^4 \log n/15$  independent evaluations, it is easy to verify that  $\mathbb{E}(\hat{f}(x) - \hat{f}(y)) = \mu$ . Furthermore,  $|f^n(x) - f^n(y)| \leq n$ . By Hoeffding's inequality, we get

$$\mathbb{P}(\hat{f}(x) \leq \hat{f}(y)) \leq \mathbb{P}(|\hat{f}(x) - \hat{f}(y) - \mu| \geq \mu) \leq 2e^{-\frac{2m^2\mu^2}{m(2n)^2}} \leq 2n^{-\frac{2}{15}}.$$

When  $i \leq n-4$ ,  $\mu \geq (n-i)/n \geq 4/n$ . By applying this lower bound to the above inequality, we get

$$\mathbb{P}(\hat{f}(x) \leq \hat{f}(y)) \leq 2n^{-\frac{32}{15}}.$$

Thus, case (1) holds.

For case (2), we are to analyze  $P(\hat{f}(1^n) \leq \hat{f}(y))$  or  $P(\hat{f}(1^n) \geq \hat{f}(y))$ , where  $LO(y) < n$ .  $E(f^n(1^n))$  can be calculated as follows:

$$E(f^n(1^n)) = (1-p) \cdot n + \sum_{j=1}^n \frac{p}{n} \cdot (j-1).$$

By combining this equality and the upper bound in Eq. (49), we get, for any  $y$  with  $LO(y) = k < n$ ,

$$E(f^n(1^n) - f^n(y)) \geq (n-k) \left( 1-p + \frac{p}{n} \frac{n-k-3}{2} \right).$$

When  $k \leq n-4$ ,  $E(f^n(1^n) - f^n(y)) \geq 2/n$ . By Hoeffding's inequality, we get

$$P(\hat{f}(1^n) \leq \hat{f}(y)) \leq 2n^{-\frac{8}{15}}.$$

When  $k \in \{n-3, n-2, n-1\}$ ,

$$\mu = E(f^n(1^n) - f^n(y)) \geq 1-p-p/n.$$

If  $p \leq 1-2/n$ , we have  $\mu \geq 1/n$ . By Hoeffding's inequality,

$$P(\hat{f}(1^n) \geq \hat{f}(y)) \geq 1 - 2n^{-\frac{2}{15}}. \quad (51)$$

If  $p > 1-2/n$ ,  $\mu \geq -1/n$ . We then use Berry-Esseen inequality [31] to derive a lower bound on  $P(\hat{f}(1^n) \geq \hat{f}(y))$ . Let  $Y = f^n(1^n) - f^n(y) - \mu$ . Note that  $E(Y) = 0$ . Denote the variance  $\text{Var}(Y)$  as  $\sigma^2$ . Then,  $\sigma^2 = \text{Var}(f^n(1^n)) + \text{Var}(f^n(y))$ . For  $\text{Var}(f^n(1^n))$ , it can be calculated as follows:

$$\begin{aligned} \text{Var}(f^n(1^n)) &= E((f^n(1^n))^2) - (E(f^n(1^n)))^2 \\ &= (1-p)n^2 + \frac{p}{n} \frac{(n-1)n(2n-1)}{6} - \left( (1-p)n + \frac{p}{n} \frac{n(n-1)}{2} \right)^2 \\ &\geq \frac{p(n-1)(n-1/2)}{3} - \left( 1 + \frac{p(n-1)}{2} \right)^2 \geq \frac{n^2}{14}, \end{aligned}$$

where the last inequality holds because for  $p > 1-2/n$  and  $n$  being large enough, the minimum is reached when  $p = 1$ . Thus, we get  $\sigma^2 \geq n^2/14$ . Since  $-2n \leq Y \leq 2n$ ,  $\rho = E(|Y|^3) \leq 8n^3$ . Note that  $\hat{f}(1^n) - \hat{f}(y) - \mu$  is the average of  $m$  independent random variables, which have the same distribution as  $Y$ . By Berry-Esseen inequality [31], we have

$$P\left(\frac{(\hat{f}(1^n) - \hat{f}(y) - \mu)\sqrt{m}}{\sigma} \leq x\right) - \Phi(x) \leq \frac{\rho}{\sigma^3\sqrt{m}}, \quad (52)$$

where  $\Phi(x)$  is the cumulative distribution function of the standard normal distribution. Thus, for  $p > 1-2/n$ ,

$$P(\hat{f}(1^n) \geq \hat{f}(y)) = P(\hat{f}(1^n) - \hat{f}(y) - \mu \geq -\mu) \quad (53)$$

$$\begin{aligned}
&\geq 1 - P(\hat{f}(1^n) - \hat{f}(y) - \mu \leq -\mu) \\
&\geq 1 - \Phi\left(\frac{-\mu\sqrt{m}}{\sigma}\right) - \frac{\rho}{\sigma^3\sqrt{m}} \\
&= \Phi\left(\frac{\mu\sqrt{m}}{\sigma}\right) - \frac{\rho}{\sigma^3\sqrt{m}} \geq \Phi\left(-\frac{\sqrt{m}}{n\sigma}\right) - \frac{\rho}{\sigma^3\sqrt{m}} \\
&\geq \Phi\left(-\sqrt{\frac{56}{15}\log n}\right) - O\left(\frac{1}{n^2\sqrt{\log n}}\right) \\
&\geq \frac{1}{\sqrt{2\pi}} \int_{-2\sqrt{\log n}}^{-\sqrt{\frac{56}{15}\log n}} e^{-\frac{t^2}{2}} dt - O\left(\frac{1}{n^2\sqrt{\log n}}\right) \geq \frac{1}{n^2},
\end{aligned}$$

where the second inequality is by Eq. (52), the third inequality is  $\mu \geq -1/n$ , the fourth is by  $m = 4n^4 \log n/15$ ,  $\sigma^2 \geq n^2/14$  and  $\rho \leq 8n^3$ , and the last holds with sufficiently large  $n$ . According to Eqs. (51) and (53), we get that, when  $\text{LO}(y) \in \{n-3, n-2, n-1\}$ ,

$$P(\hat{f}(1^n) \geq \hat{f}(y)) \geq 1/n^2.$$

Thus, case (2) holds.  $\square$

The following theorem shows that for one-bit noise, using sampling with  $m = 4n^4 \log n/15$  is sufficient to make the (1+1)-EA solve LeadingOnes in polynomial time.

**Theorem 22** *For the (1+1)-EA on LeadingOnes under one-bit noise, if using sampling with  $m = 4n^4 \log n/15$ , the expected running time is polynomial.*

*Proof* We use Theorem 1 to prove it. We first construct a distance function  $V(x)$  as, for any  $x$  with  $\text{LO}(x) = i$ ,

$$V(x) = \begin{cases} n - i/n^6, & i \leq n-1; \\ 0, & i = n. \end{cases}$$

Then, we investigate  $E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x)$  for any  $x$  with  $\text{LO}(x) = i < n$ . We divide the drift into two parts: positive  $E^+$  and negative  $E^-$ . That is,

$$E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x) = E^+ - E^-.$$

For  $i \leq n-4$ , the lower bound analysis on  $E^+$  is similar to that in the proof of Theorem 19. We only consider that the  $(i+1)$ -th bit of  $x$  is flipped and the other bits keep unchanged, whose probability is  $\frac{1}{n}(1 - \frac{1}{n})^{n-1}$ . The offspring solution  $x'$  is the same as  $x$  except the  $(i+1)$ -th bit, and  $\text{LO}(x') > \text{LO}(x) = i$ . According to definition of  $V(x)$ , we know that the decrease on the distance is at least  $1/n^6$ . If  $\text{LO}(x') = n$ ,  $P(\hat{f}(x') \leq \hat{f}(x)) \leq 2n^{-\frac{8}{15}}$  by case (2) of Lemma 6. If  $\text{LO}(x') < n$ , it must hold that  $\text{LO}(x) \leq \text{LO}(x') - 1 \wedge x_{\text{LO}(x')+1} = x'_{\text{LO}(x')+1} = 0$ . By case (1) of Lemma 6,  $P(\hat{f}(x') \leq \hat{f}(x)) \leq 2n^{-\frac{2}{15}}$ . Thus, the probability  $P(\hat{f}(x') \geq \hat{f}(x))$  of accepting the offspring solution  $x'$  is at least

$1 - \max\{2n^{-\frac{8}{15}}, 2n^{-\frac{2}{15}}\} \geq 1/2$ , where the inequality holds with sufficiently large  $n$ . Then,  $E^+$  can be lower bounded as follows:

$$E^+ \geq \left(1 - \frac{1}{n}\right)^{n-1} \cdot \frac{1}{n} \cdot \frac{1}{2} \cdot \frac{1}{n^6} \geq \frac{1}{6n^7}.$$

For the negative drift  $E^-$ , we need to consider  $\text{LO}(x') < \text{LO}(x) = i$ . We further divide  $\text{LO}(x') < i$  into two cases. If  $\text{LO}(x') \leq i - 2$  or  $\text{LO}(x') = i - 1 \wedge x'_{i+1} = 0$ , then  $P(\hat{f}(x') \geq \hat{f}(x)) \leq 2n^{-\frac{32}{15}}$  by case (1) of Lemma 6 (note that  $i \leq n - 4$  here), and  $V(x') - V(x) \leq V(0^n) - V(x) = i/n^6$ . If  $\text{LO}(x') = i - 1 \wedge x'_{i+1} = 1$ , then  $\sum_{x': \text{LO}(x')=i-1 \wedge x'_{i+1}=1} P_{\text{mut}}(x, x') \leq 1/n^2$  since it is necessary to flip the  $i$ -th and the  $(i + 1)$ -th bits of  $x$  in mutation, and  $V(x') - V(x) = 1/n^6$ . Then,  $E^-$  can be upper bounded by as follows:

$$E^- \leq \frac{i}{n^6} \cdot 1 \cdot \frac{2}{n^{\frac{32}{15}}} + \frac{1}{n^6} \cdot \frac{1}{n^2} \cdot 1 = o\left(\frac{1}{n^7}\right).$$

By subtracting  $E^-$  from  $E^+$ , we have

$$E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x) = E^+ - E^- \geq \frac{1}{12n^7}.$$

For  $i \in \{n - 3, n - 2, n - 1\}$ , we use a trivial upper bound  $i/n^6$  on  $E^-$ . For the positive drift, we consider that the offspring solution  $x'$  is the optimal solution  $1^n$ , whose probability is at least  $\frac{1}{n^3}(1 - \frac{1}{n})^{n-3} \geq \frac{1}{en^3}$  since at most three bits of  $x$  need to be flipped. The probability of accepting  $1^n$  is at least  $\frac{1}{n^2}$  by case (2) of Lemma 6. The distance decrease is at least  $n - \frac{n-1}{n^6} \geq \frac{n}{2}$ . Thus,  $E^+ \geq \frac{1}{en^3} \cdot \frac{1}{n^2} \cdot \frac{n}{2} \geq \frac{1}{2en^4}$ . By subtracting  $E^-$  from  $E^+$ , we have

$$E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x) \geq \frac{1}{2en^4} - \frac{i}{n^6} = \frac{1}{2en^4} - O\left(\frac{1}{n^5}\right) \geq \frac{1}{4en^4}.$$

By combing the above analyses for  $i \leq n - 4$  and  $i \in \{n - 3, n - 2, n - 1\}$ , we get a unified lower bound  $1/(12n^7)$  on  $E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x)$ . Since  $V(x) \leq n$ , we have  $E(\tau \mid \xi_0) = O(n^8)$  by Theorem 1. Each iteration of the (1+1)-EA using sampling takes  $2m = 8n^4 \log n/15$  number of fitness evaluations, thus the expected running time is polynomial.  $\square$

Therefore, we have shown that the (1+1)-EA using sampling can always solve LeadingOnes in polynomial time under bit-wise noise  $(p, \frac{1}{n})$  (i.e., Theorem 19) or one-bit noise (i.e., Theorem 22); while under bit-wise noise  $(1, q)$ , the tight range of  $q$  allowing a polynomial time is  $O(\log n/n)$  (i.e., Theorems 20 and 21). The reason why sampling is ineffective under bit-wise noise  $(1, q)$  with  $q = \omega(\log n/n)$  is similar to that observed in the analysis of OneMax under bit-wise noise  $(1, q)$  with  $q = 1/2 - 1/n^{\omega(1)}$  or  $q \geq 1/2$ . For two solutions  $x$  and  $y$  with  $f(x) > f(y)$ , we can find from the calculation of  $E(f^n(x) - f^n(y))$  in the proof of Lemma 5 that when  $q = \omega(\log n/n)$ ,  $E(f^n(x) - f^n(y))$  will be very small (since  $(1 - q)^{n-1} = 1/n^{\omega(1)}$ ) or even

negative; thus a polynomial sample size is not sufficient to make the probability of accepting the true worse solution  $y$  small enough, or it will increase the probability. For the situation where sampling is effective, the analysis on LeadingOnes is a little different from that on OneMax. On OneMax,  $E(f^n(x) - f^n(y))$  is always sufficiently large when  $f(x) > f(y)$ ; thus sampling can make the probability of accepting the true worse solution  $y$  small enough and then work. While on LeadingOnes,  $E(f^n(x) - f^n(y))$  is sufficiently large in most cases instead of all cases when  $f(x) > f(y)$ , but a few exceptions do not affect the effectiveness of sampling.

## 6 Conclusion

In this paper, we theoretically study the (1+1)-EA solving the OneMax and LeadingOnes problems under bit-wise noise, which is characterized by a pair  $(p, q)$  of parameters. We derive the ranges of  $p$  and  $q$  for the running time being polynomial and super-polynomial, respectively. The previously known parameter ranges for the (1+1)-EA solving LeadingOnes under one-bit noise are also improved. Considering that the (1+1)-EA is efficient only under low noise levels, we further analyze the robustness of sampling to noise. We prove that for both bit-wise noise and one-bit noise, using sampling can significantly enlarge the range of noise parameters allowing a polynomial running time. In the future, we shall improve the currently derived bounds on LeadingOnes, as they do not cover the whole range of noise parameters. For proving polynomial upper bounds on the expected running time by using sampling, we only give a sufficiently large sample size, the tightness of which will be studied in our future work. In our analysis, we consider the bit-wise noise model with one parameter fixed. Thus, to analyze the running time under general bit-wise noise is also an interesting future work. Note that our analysis has shown that the performance of the (1+1)-EA solving OneMax under bit-wise noise  $(1, \frac{\log n}{30n})$  and  $(\frac{\log n}{30n}, 1)$  is significantly different, which implies that  $p \cdot q$  may not be the only deciding factor for the analysis of general bit-wise noise.

**Acknowledgements** We want to thank the reviewers for their valuable comments. This work was supported by the NSFC (61603367, 61672478), the YESS (2016QNRC001), the Science and Technology Innovation Committee Foundation of Shenzhen (ZDSYS201703031748284), and the Royal Society Newton Advanced Fellowship (NA150123).

## References

1. Aizawa, A.N., Wah, B.W.: Scheduling of genetic algorithms in a noisy environment. *Evolutionary Computation* **2**(2), 97–122 (1994)
2. Akimoto, Y., Astete-Morales, S., Teytaud, O.: Analysis of runtime of optimization algorithms for noisy functions over discrete codomains. *Theoretical Computer Science* **605**, 42–50 (2015)

3. Arnold, D.V., Beyer, H.G.: A general noise model and its effects on evolution strategy performance. *IEEE Transactions on Evolutionary Computation* **10**(4), 380–391 (2006)
4. Auger, A., Doerr, B.: *Theory of Randomized Search Heuristics: Foundations and Recent Developments*. World Scientific, Singapore (2011)
5. Bäck, T.: *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press, Oxford, UK (1996)
6. Branke, J., Schmidt, C.: Selection in the presence of noise. In: *Proceedings of the 5th ACM Conference on Genetic and Evolutionary Computation (GECCO'03)*, pp. 766–777. Chicago, IL (2003)
7. Chang, Y., Chen, S.: A new query reweighting method for document retrieval based on genetic algorithms. *IEEE Transactions on Evolutionary Computation* **10**(5), 617–622 (2006)
8. Corus, D., Dang, D.C., Eremeev, A.V., Lehre, P.K.: Level-based analysis of genetic algorithms and other search processes. In: *Proceedings of 13th International Conference on Parallel Problem Solving from Nature (PPSN'14)*, pp. 912–921. Ljubljana, Slovenia (2014)
9. Dang, D.C., Lehre, P.K.: Efficient optimisation of noisy fitness functions with population-based evolutionary algorithms. In: *Proceedings of the 13th ACM Conference on Foundations of Genetic Algorithms (FOGA'15)*, pp. 62–68. Aberystwyth, UK (2015)
10. Doerr, B., Goldberg, L.A.: Adaptive drift analysis. *Algorithmica* **65**(1), 224–250 (2013)
11. Doerr, B., Hota, A., Kötzing, T.: Ants easily solve stochastic shortest path problems. In: *Proceedings of the 14th ACM Conference on Genetic and Evolutionary Computation (GECCO'12)*, pp. 17–24. Philadelphia, PA (2012)
12. Doerr, B., Johannsen, D., Winzen, C.: Multiplicative drift analysis. *Algorithmica* **64**(4), 673–697 (2012)
13. Droste, S.: Analysis of the (1+1) EA for a noisy OneMax. In: *Proceedings of the 6th ACM Conference on Genetic and Evolutionary Computation (GECCO'04)*, pp. 1088–1099. Seattle, WA (2004)
14. Droste, S., Jansen, T., Wegener, I.: On the analysis of the (1+1) evolutionary algorithm. *Theoretical Computer Science* **276**(1-2), 51–81 (2002)
15. Feldmann, M., Kötzing, T.: Optimizing expected path lengths with ant colony optimization using fitness proportional update. In: *Proceedings of the 12th ACM Conference on Foundations of Genetic Algorithms (FOGA'13)*, pp. 65–74. Adelaide, Australia (2013)
16. Friedrich, T., Kötzing, T., Krejca, M., Sutton, A.: Robustness of ant colony optimization to noise. *Evolutionary Computation* **24**(2), 237–254 (2016)
17. Friedrich, T., Kötzing, T., Krejca, M., Sutton, A.: The compact genetic algorithm is efficient under extreme gaussian noise. *IEEE Transactions on Evolutionary Computation* **21**(3), 477–490 (2017)
18. Friedrich, T., Kötzing, T., Quinzan, F., Sutton, A.: Resampling vs recombination: A statistical run time estimation. In: *Proceedings of 14th ACM Conference on Foundations of Genetic Algorithms (FOGA'17)*, pp. 25–35. Copenhagen, Denmark (2017)
19. Gießen, C., Kötzing, T.: Robustness of populations in stochastic environments. *Algorithmica* **75**(3), 462–489 (2016)
20. He, J., Yao, X.: Drift analysis and average time complexity of evolutionary algorithms. *Artificial Intelligence* **127**(1), 57–85 (2001)
21. Jin, Y., Branke, J.: Evolutionary optimization in uncertain environments—a survey. *IEEE Transactions on Evolutionary Computation* **9**(3), 303–317 (2005)
22. Ma, P., Chan, K., Yao, X., Chiu, D.: An evolutionary clustering algorithm for gene expression microarray data analysis. *IEEE Transactions on Evolutionary Computation* **10**(3), 296–314 (2006)
23. Neumann, F., Witt, C.: *Bioinspired Computation in Combinatorial Optimization: Algorithms and Their Computational Complexity*. Springer-Verlag, Berlin, Germany (2010)
24. Oliveto, P., Witt, C.: Simplified drift analysis for proving lower bounds in evolutionary computation. *Algorithmica* **59**(3), 369–386 (2011)
25. Oliveto, P., Witt, C.: Erratum: Simplified drift analysis for proving lower bounds in evolutionary computation. *CORR abs/1211.7184* (2012)
26. Prügel-Bennett, A., Rowe, J., Shapiro, J.: Run-time analysis of population-based evolutionary algorithm in noisy environments. In: *Proceedings of the 13th ACM Conference on Foundations of Genetic Algorithms (FOGA'15)*, pp. 69–75. Aberystwyth, UK (2015)

27. Qian, C., Bian, C., Jiang, W., Tang, K.: Running time analysis of the (1+1)-EA for OneMax and LeadingOnes under bit-wise noise. In: Proceedings of the 19th ACM Conference on Genetic and Evolutionary Computation (GECCO'17), pp. 1399–1406. Berlin, Germany (2017)
28. Qian, C., Yu, Y., Tang, K., Jin, Y., Yao, X., Zhou, Z.H.: On the effectiveness of sampling for evolutionary optimization in noisy environments. *Evolutionary Computation* **26**(2), 237–267 (2018)
29. Qian, C., Yu, Y., Zhou, Z.H.: Analyzing evolutionary optimization in noisy environments. *Evolutionary Computation* **26**(1), 1–41 (2018)
30. Rowe, J.E., Sudholt, D.: The choice of the offspring population size in the  $(1, \lambda)$  evolutionary algorithm. *Theoretical Computer Science* **545**, 20–38 (2014)
31. Shevtsova, I.G.: Sharpening of the upper bound of the absolute constant in the berry–esseen inequality. *Theory of Probability & Its Applications* **51**(3), 549–553 (2007)
32. Sudholt, D.: A new method for lower bounds on the running time of evolutionary algorithms. *IEEE Transactions on Evolutionary Computation* **17**(3), 418–435 (2013)
33. Sudholt, D., Thyssen, C.: A simple ant colony optimizer for stochastic shortest path problems. *Algorithmica* **64**(4), 643–672 (2012)
34. Yu, Y., Qian, C., Zhou, Z.H.: Switch analysis for running time analysis of evolutionary algorithms. *IEEE Transactions on Evolutionary Computation* **19**(6), 777–792 (2015)