Supplementary Material of "On the Effectiveness of Sampling for Evolutionary Optimization in Noisy Environments"

Chao Qian[1,2] chaoqian@ustc.edu.cn
Yang Yu[2] yuy@nju.edu.cn
Ke Tang[1] ketang@ustc.edu.cn
Yaochu Jin[3] yaochu.jin@surrey.ac.uk
Xin Yao[1,4] x.yao@cs.bham.ac.uk
Zhi-Hua Zhou[2*] zhouzh@nju.edu.cn
[1]USTC-Birmingham Joint Research Institute in Intelligent Computation and Its Applications, School of Computer Science and Technology, University of Science and Technology of China, Hefei, 230027, China
[2]National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210023, China
[3]Department of Computer Science, University of Surrey, Guildford, GU2 7XH, UK
[4]Center of Excellence for Research in Computational Intelligence and Applications, School of Computer Science, University of Birmingham, Birmingham, B15 2TT, UK

1 Detailed Proofs

This document aims to provide the detailed proofs of Theorems 2 and 10, which are omitted in our original paper due to space limitations.

Proof of Theorem 2. We use Lemma 4 to prove this theorem. We first analyze $p_{i,i+d}$ as that analyzed in the proof of Theorem 1. Note that for a solution x, the fitness value output by sampling with $k=2$ is $\hat{f}(x)=(f_1^n(x)+f_2^n(x))/2$, where $f_1^n(x)$ and $f_2^n(x)$ are noisy fitness values output by two independent fitness evaluations.
(1) When $d \geq 3$, $\hat{f}(x') \leq n-i-d+1 \leq n-i-2 < \hat{f}(x)$. Thus, the offspring x' will be discarded, then we have $\forall d \geq 3: p_{i,i+d}=0$.
(2) When $d=2$, the offspring solution x' will be accepted if and only if $\hat{f}(x')=n-i-1=\hat{f}(x)$. The probability of $\hat{f}(x')=n-i-1$ is $(\frac{i+2}{n})^2$, since it needs to always flip one 0-bit of x' in two noisy fitness evaluations. The probability of $\hat{f}(x)=n-i-1$ is $(\frac{n-i}{n})^2$, since it needs to always flip one 1-bit of x. Thus, $p_{i,i+2}=P_2 \cdot (\frac{i+2}{n})^2(\frac{n-i}{n})^2$.
(3) When $d=1$, there are three possible cases for the acceptance of x': $\hat{f}(x')=n-i \wedge \hat{f}(x)=n-i-1$, $\hat{f}(x')=n-i \wedge \hat{f}(x)=n-i$ and $\hat{f}(x')=n-i-1 \wedge \hat{f}(x)=n-i-1$. The probability of $\hat{f}(x')=n-i$ is $(\frac{i+1}{n})^2$, since it needs to always flip one 0-bit of x in two noisy evaluations. The probability of $\hat{f}(x')=n-i-1$ is $2\frac{i+1}{n}\frac{n-i-1}{n}$, since it needs to flip one 0-bit of x in one noisy evaluation and flip one 1-bit in the other

---
* Corresponding author

©200X by the Massachusetts Institute of Technology

noisy evaluation. Similarly, we can derive that the probabilities of $\hat{f}(x) = n-i-1$ and $\hat{f}(x) = n-i$ are $(\frac{n-i}{n})^2$ and $2\frac{n-i}{n}\frac{i}{n}$, respectively. Thus, $p_{i,i+1} = P_1 \cdot ((\frac{i+1}{n})^2((\frac{n-i}{n})^2 + 2\frac{n-i}{n}\frac{i}{n}) + 2\frac{i+1}{n}\frac{n-i-1}{n}(\frac{n-i}{n})^2)$.
(4) When $d=-1$, x' will be rejected if and only if $\hat{f}(x') = n-i \wedge \hat{f}(x) = n-i+1$. The probability of $\hat{f}(x') = n-i$ is $(\frac{n-i+1}{n})^2$, since it needs to always flip one 1-bit of x' in two noisy evaluations. The probability of $\hat{f}(x) = n-i+1$ is $(\frac{i}{n})^2$, since it needs to always flip one 0-bit of x. Thus, $p_{i,i-1} = P_{-1} \cdot (1-(\frac{n-i+1}{n})^2(\frac{i}{n})^2)$.
(5) When $d \le -2$, $\hat{f}(x') \ge n-i-d-1 \ge n-i+1 \ge \hat{f}(x)$. Thus, the offspring x' will always be accepted, then we have $\forall d \le -2 : p_{i,i+d} = P_d$.

Using these probabilities, we have

$$\begin{aligned}
&\mathbb{E}[\![X_t - X_{t+1} \mid X_t = i]\!] = \sum_{d=1}^{i} d \cdot p_{i,i-d} - \sum_{d=1}^{n-i} d \cdot p_{i,i+d} \\
&= \left(1-\left(\frac{n-i+1}{n}\right)^2\left(\frac{i}{n}\right)^2\right)P_{-1} + \sum_{d=2}^{i} dP_{-d} - 2\left(\frac{i+2}{n}\right)^2\left(\frac{n-i}{n}\right)^2 P_2 \\
&\quad - \left(\left(\frac{i+1}{n}\right)^2\left(\left(\frac{n-i}{n}\right)^2 + 2\frac{n-i}{n}\frac{i}{n}\right) + 2\frac{i+1}{n}\frac{n-i-1}{n}\left(\frac{n-i}{n}\right)^2\right)P_1 \\
&\le \left(1-\left(\frac{n-i+1}{n}\right)^2\left(\frac{i}{n}\right)^2\right)\frac{i}{n}\left(1-\frac{1}{n}\right)^{n-1} \cdot 1.14 + \frac{i}{n}\left(\left(1+\frac{1}{n}\right)^{i-1} - 1\right) \\
&\quad - 2\left(\frac{i+2}{n}\right)^2\left(\frac{n-i}{n}\right)^2 \frac{(n-i)(n-i-1)}{2n^2}\left(1-\frac{1}{n}\right)^{n-2} - \frac{n-i}{n}\left(1-\frac{1}{n}\right)^{n-1} \\
&\quad \cdot \left(\left(\frac{i+1}{n}\right)^2\left(\left(\frac{n-i}{n}\right)^2 + 2\frac{n-i}{n}\frac{i}{n}\right) + 2\frac{i+1}{n}\frac{n-i-1}{n}\left(\frac{n-i}{n}\right)^2\right) \\
&\quad (\text{by using the bounds of } P_d \text{ in the proof of Theorem 1}) \\
&\le \frac{i}{n}\left(1-\frac{1}{n}\right)^{n-1}(1.14-2) + O\left(\left(\frac{i}{n}\right)^2\right) \quad (\text{since } i < n^{1/4}) \\
&\le -0.3 \cdot \frac{i}{n} + O\left(\left(\frac{i}{n}\right)^2\right). \quad (\text{by } \left(1-\frac{1}{n}\right)^{n-1} \ge \frac{1}{e})
\end{aligned}$$

It is also easy to verify that $P(X_{t+1} \neq i \mid X_t = i) = \Theta(\frac{i}{n})$ for $1 \le i < n^{1/4}$. Thus, $\mathbb{E}[\![X_t - X_{t+1} \mid X_t = i]\!] = -\Omega(P(X_{t+1} \neq i \mid X_t = i))$, which implies that condition 1 of Lemma 4 holds.

Condition 2 of Lemma 4 still holds with $\delta = 1$ and $r(l) = \frac{32e}{7}$. The analysis procedure is the same as that in the proof of Theorem 1, because the following inequality holds:

$$\begin{aligned}
P(|X_{t+1} - X_t| \ge 1 \mid X_t = i) \ge p_{i,i-1} &= \left(1-\left(\frac{n-i+1}{n}\right)^2\left(\frac{i}{n}\right)^2\right) \cdot P_{-1} \\
&\ge \left(1-\frac{n-i+1}{n}\frac{i}{n}\right) \cdot P_{-1}.
\end{aligned}$$

Thus, by Lemma 4, the expected running time is exponential. □

Proof of Theorem 10. We use Lemma 2 to prove this theorem. The proof is very similar to that of Theorem 3 except that the probabilities $p_{i,i+d}$ are different due to the difference on the noise and the value of k.

We use the distance function $V(x) = |x|_0$. Let i (where $1 \leq i \leq n$) denote the number of 0-bits of the current solution x. Let $p_{i,i+d}$ be the probability that the next solution after mutation and selection has $i+d$ number of 0-bits (where $-i \leq d \leq n-i$). Thus,

$$\mathbb{E}[\![V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x]\!] = \sum\nolimits_{d=1}^{i} d \cdot p_{i,i-d} - \sum\nolimits_{d=1}^{n-i} d \cdot p_{i,i+d}. \quad (1)$$

We then analyze $p_{i,i+d}$ $(1 \leq i \leq n)$. For a solution x, the fitness value output by sampling is the average of noisy fitness values by k independent evaluations, i.e., $\hat{f}(x) = \sum_{i=1}^{k} f_i^n(x)/k$. Note that for the flipping in asymmetric one-bit noise, the probability of flipping a 0 (or 1) bit is different when $|x|_0 = 0$, $|x|_0 = n$ and $0 < |x|_0 < n$. In these three cases, the probabilities of flipping a 0 bit are 0, 1 and $\frac{1}{2}$, respectively; the probabilities of flipping a 1 bit are 1, 0 and $\frac{1}{2}$, respectively. Thus, the analysis of $p_{i,i+d}$ will also be separated into several cases if necessary. Let P_d denote the probability that the offspring solution x' generated by mutation has $i+d$ number of 0-bits.

(1) When $d \geq 3$, $\hat{f}(x') \leq n-i-d+1 \leq n-i-2 < \hat{f}(x)$. Thus, the offspring x' will be discarded, then $\forall d \geq 3 : p_{i,i+d} = 0$.

(2) When $d = 2$, x' will be accepted if and only if $\hat{f}(x') = n-i-1 = \hat{f}(x)$, that is, it needs to always flip one 0-bit of x' and flip one 1-bit of x in k noisy fitness evaluations. We then consider three cases:

- $i = n$ or $n-1$. It trivially holds that $p_{i,i+2} = 0$.
- $i = n-2$. Note that $|x'|_0 = i+2 = n$, thus the probability of flipping a 0 bit of x' in noisy evaluation is 1. Then, we have $p_{i,i+2} = P_2 \cdot 1^k \cdot \frac{1}{2^k}$.
- $1 \leq i < n-2$. We have $p_{i,i+2} = P_2 \cdot \frac{1}{2^k} \cdot \frac{1}{2^k}$.

(3) When $d = 1$, there are two possible values for $f^n(x')$: $n-i-2$ or $n-i$. Similarly, $f^n(x) = n-i-1$ or $n-i+1$. In the k independent noisy evaluations for x', let $k_1 \in [0,k]$ denote the number of times that $f^n(x') = n-i$. Similarly, let $k_2 \in [0,k]$ denote the number of times that $f^n(x) = n-i-1$. The condition for the acceptance of x' is $\hat{f}(x') \geq \hat{f}(x)$, which can be simplified as follows.

$$\begin{aligned}
\hat{f}(x') \geq \hat{f}(x) &\Leftrightarrow \sum\nolimits_{i=1}^{k} f_i^n(x') \geq \sum\nolimits_{i=1}^{k} f_i^n(x) \\
&\Leftrightarrow k_1(n-i) + (k-k_1)(n-i-2) \geq k_2(n-i-1) + (k-k_2)(n-i+1) \\
&\Leftrightarrow k_1 + k_2 \geq \frac{3}{2}k.
\end{aligned}$$

We then consider three cases:

- $i = n$. It trivially holds that $p_{i,i+1} = 0$.
- $i = n-1$. Note that $|x'|_0 = i+1 = n$, thus the probability of flipping a 0 bit of x' in noisy evaluation (i.e., $f^n(x') = n-i$) is 1, which implies that $k_1 = k$. Thus, the condition of accepting x' changes to be $k_2 \geq \frac{k}{2}$. Then, we have $p_{i,i+1} = P_1 \cdot \sum_{k_2 \geq \frac{k}{2}} \binom{k}{k_2} \frac{1}{2^k}$.
- $1 \leq i < n-1$. We have $p_{i,i+1} = P_1 \cdot \sum_{k_1+k_2 \geq \frac{3}{2}k} \binom{k}{k_1} \frac{1}{2^k} \cdot \binom{k}{k_2} \frac{1}{2^k} = P_1 \cdot \sum_{k' \geq \frac{3}{2}k} \binom{2k}{k'} \frac{1}{2^{2k}}$.

(4) When $d = -1$, $f^n(x') = n-i$ or $n-i+2$; $f^n(x) = n-i-1$ or $n-i+1$. In the k independent noisy evaluations for x', let $k_1 \in [0,k]$ denote the number of times that $f^n(x') = n-i$. Similarly, let $k_2 \in [0,k]$ denote the number of times that $f^n(x) = n-i+1$. The condition for the rejection of x' is $\hat{f}(x') < \hat{f}(x)$, which can be simplified as follows.

$$\begin{aligned}\hat{f}(x') < \hat{f}(x) &\Leftrightarrow \sum\nolimits_{i=1}^{k} f_i^n(x') < \sum\nolimits_{i=1}^{k} f_i^n(x)\\ &\Leftrightarrow k_1(n-i)+(k-k_1)(n-i+2) < k_2(n-i+1)+(k-k_2)(n-i-1)\\ &\Leftrightarrow k_1+k_2 > \frac{3}{2}k.\end{aligned}$$

We then consider three cases:

- $i = n$. Note that the probability of flipping a 0 bit of x in noisy evaluation (i.e., $f^n(x) = n-i+1$) is 1, which implies that $k_2 = k$. Thus, the condition of rejecting x' changes to be $k_1 > \frac{k}{2}$. Then, we have $p_{i,i-1} = P_{-1} \cdot (1 - \sum_{k_1 > \frac{k}{2}} \binom{k}{k_1} \frac{1}{2^k})$.
- $i = 1$. Note that $|x'|_0 = i-1 = 0$, thus the probability of flipping a 1 bit of x' in noise (i.e., $f^n(x') = n-i$) is 1, which implies that $k_1 = k$. Thus, the condition of rejecting x' changes to be $k_2 > \frac{k}{2}$. Then, we have $p_{i,i-1} = P_{-1} \cdot (1 - \sum_{k_2 > \frac{k}{2}} \binom{k}{k_2} \frac{1}{2^k})$.
- $1 < i < n$. $p_{i,i-1} = P_{-1} \cdot (1 - \sum_{k_1+k_2 > \frac{3}{2}k} \binom{k}{k_1}\frac{1}{2^k} \cdot \binom{k}{k_2}\frac{1}{2^k}) = P_{-1} \cdot (1 - \sum_{k' > \frac{3}{2}k} \binom{2k}{k'} \frac{1}{2^{2k}})$.

By combining the above three cases, we can easily derive that $p_{i,i-1} \geq P_{-1} \cdot \frac{1}{2}$.

(5) When $d \leq -2$, $\hat{f}(x') \geq n-i-d-1 \geq n-i+1 \geq \hat{f}(x)$. Thus, x' will always be accepted, then we have $\forall d \leq -2: p_{i,i+d} = P_d$.

By applying these probabilities to Eq. (1), we have

$$\mathbb{E}[\![V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x]\!] \geq p_{i,i-1} - p_{i,i+1} - 2 \cdot p_{i,i+2}. \tag{2}$$

We then analyze Eq. (2) in three cases.

(1) When $i = n$, $p_{i,i+2} = 0$ and $p_{i,i+1} = 0$. Thus, we have

$$\mathbb{E}[\![V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x]\!] \geq P_{-1} \cdot \frac{1}{2} \geq \frac{i}{n}\left(1-\frac{1}{n}\right)^{n-1} \cdot \frac{1}{2} \geq \frac{i}{2en}.$$

(2) When $i = n-1$, $p_{i,i+2} = 0$ and $p_{i,i+1} = P_1 \cdot \sum_{k_2 \geq \frac{k}{2}} \binom{k}{k_2} \frac{1}{2^k} < P_1 \leq \frac{n-i}{n} = \frac{1}{n}$. Thus,

$$\mathbb{E}[\![V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x]\!] \geq P_{-1} \cdot \frac{1}{2} - P_1 \geq \frac{i}{2en} - \frac{1}{n} \geq 0.01 \cdot \frac{i}{n},$$

where the last inequality holds with $n \geq 7$.

(3) When $1 \leq i < n-1$, $p_{i,i+2} \leq P_2 \cdot \frac{1}{2^k}$ and $p_{i,i+1} = P_1 \cdot \sum_{k' \geq \frac{3}{2}k} \binom{2k}{k'} \frac{1}{2^{2k}}$. Let X_i $(1 \leq i \leq 2k)$ be independent random variables such that $P(X_i = 1) = \frac{1}{2}$ and $P(X_i = 0) = \frac{1}{2}$. Then, $\sum_{k' \geq \frac{3}{2}k} \binom{2k}{k'} \frac{1}{2^{2k}} = P(\sum_{i=1}^{2k} X_i \geq \frac{3}{2}k) \leq e^{-\frac{k}{12}}$, where the "$\leq$" is by Chernoff's inequality. Thus, we have

$$\begin{aligned}\mathbb{E}[\![V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x]\!] &\geq \frac{P_{-1}}{2} - P_1 \cdot e^{-\frac{k}{12}} - 2 \cdot \frac{P_2}{2^k}\\ &\geq \frac{i}{2en} - \frac{1}{n^2} - \frac{1}{n^{12}} \geq 0.01 \cdot \frac{i}{n},\end{aligned}$$

where the second inequality is by $k = \lceil 24 \log n \rceil$ (note that $\log$ corresponds to the natural logarithm, i.e., the base is e), and the last inequality holds with $n \geq 6$.

Thus, the condition of Lemma 2 holds with $\mathbb{E}[\![V(\xi_t) - V(\xi_{t+1}) \mid \xi_t = x]\!] \geq \frac{0.01}{n} \cdot V(x)$. We then have

$$\mathbb{E}[\![\tau \mid \xi_0]\!] \leq \frac{n}{0.01} \cdot (1 + \log V(\xi_0)) \in O(n \log n),$$

i.e., the expected iterations for finding the optimal solution is upper bounded by $O(n \log n)$. Because the cost of each iteration is $2k = 2 \cdot \lceil 24 \log n \rceil$, the expected running time is $O(n \log^2 n)$. □