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 $\hat{f}(x') = n - i$ and $\hat{f}(x) = n - i - 1$. The probability of $\hat{f}(x') = n - i$ is $(\frac{i+2}{n})^2$, 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 $(\frac{n-i}{n})^2$, since it needs to always flip one 1-bit of $x$. Thus, $p_{i,i+2} = P_2 \cdot (\frac{i+2}{n})^2 \cdot (\frac{n-i}{n})^2$.

(3) When $d = 1$, there are three possible cases for the acceptance of $x'$: $\hat{f}(x') = n - i$ and $\hat{f}(x) = n - i - 1$. 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 $2 \cdot \frac{i+1}{n} \cdot \frac{n-i-1}{n}$, since it needs to flip one 0-bit of $x$ in one noisy evaluation and flip one 1-bit in the other
noisy evaluation. Similarly, we can derive that the probabilities of \( f(x) = n - i - 1 \) and \( f(x) = n - i \) are \( \left( \frac{n-i}{n} \right)^2 + \frac{2n-i}{n} \), respectively. Thus, \( p_{i,i+1} = P_1 \cdot \left( \left( \frac{i+1}{n} \right)^2 + \frac{2n-i}{n} \right) \).

(4) When \( d = -1 \), \( x' \) will be rejected if and only if \( \hat{f}(x') = n - i \land \hat{f}(x) = n - i + 1 \). The probability of \( \hat{f}(x') = n - i \) is \( \left( \frac{n-i}{n} \right)^2 \), 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 \( \left( \frac{i}{n} \right)^2 \), since it needs to always flip one 0-bit of \( x \). Thus, \( p_{i,i-1} = P_{-1} \cdot \left( 1 - \left( \frac{n-i}{n} \right)^2 \right) \).

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

Using these probabilities, we have

\[
\mathbb{E}\left[ X_i - X_{i+1} \mid X_i = i \right] = \sum_{d=1}^{i} d \cdot p_{i,i-d} - \sum_{d=1}^{n-i} d \cdot p_{i,i+d}
\]

\[
= \left( 1 - \left( \frac{n-i+1}{n} \right)^2 \cdot \left( \frac{1}{n} \right)^2 \right) P_{-1} + \sum_{d=2}^{i} dP_{-d} - 2 \left( \frac{n-i+1}{n} \right)^2 \cdot \left( \frac{n-i}{n} \right)^2 P_2 
\]

\[
- \left( \left( \frac{i+1}{n} \right)^2 + \frac{2n-i}{n} \right) \cdot \left( \frac{n-i}{n} \right)^2 \cdot \left( \frac{i}{n} \right) P_1 
\]

\[
\leq \left( 1 - \left( \frac{n-i+1}{n} \right)^2 \cdot \left( \frac{1}{n} \right)^2 \right) \cdot \frac{i}{n} \left( 1 - \frac{1}{n} \right)^{n-1} \cdot \frac{i}{n} \left( 1 - \frac{1}{n} \right)^{n-1} 
\]

\[
- 2 \left( \frac{i}{n} \right)^2 \cdot \frac{n-i}{n} \cdot \frac{n-i-1}{n} \cdot \frac{1}{n} \cdot \frac{1}{n} \cdot \frac{1}{n} \cdot \frac{1}{n} 
\]

\[
\cdot \left( \left( \frac{i+1}{n} \right)^2 + \frac{2n-i}{n} \right) \cdot \left( \frac{n-i}{n} \right)^2 \cdot \left( \frac{i}{n} \right) P_1 
\]

(by using the bounds of \( P_2 \) in the proof of Theorem 1)

\[
\leq \frac{i}{n} \left( 1 - \frac{1}{n} \right)^{n-1} \cdot \frac{i}{n} \left( 1 - \frac{1}{n} \right)^{n-1} + O \left( \left( \frac{i}{n} \right)^2 \right) \quad \text{(since } i < n^{1/4} \text{)}
\]

\[
\leq -0.3 \cdot \frac{i}{n} + O \left( \left( \frac{i}{n} \right)^2 \right) \quad \text{(by } 1 - \frac{1}{n} \geq \frac{1}{e} \text{)}
\]

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

Condition 2 of Lemma 4 still holds with \( \delta = 1 \) and \( r(l) = \frac{30}{n} \). The analysis procedure is the same as that in the proof of Theorem 1, because the following inequality holds:

\[
P(\lvert X_{i+1} - X_i \rvert \geq 1 \mid X_i = i) \geq p_{i,i-1} = \left( 1 - \left( \frac{n-i+1}{n} \right)^2 \cdot \left( \frac{i}{n} \right)^2 \right) \cdot P_{-1} 
\]

\[
\geq \left( 1 - \frac{n-i+1}{n} \cdot \frac{i}{n} \right) \cdot P_{-1}. 
\]

Thus, by Lemma 4, the expected running time is exponential. \( \square \)
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, $P_{i,i+d}$ will also be separated into several cases if necessary. Let

$$P_{i,i+d} = \sum_{d=1}^{n-i} d \cdot p_{i,i-d} - \sum_{d=1}^{n-i} d \cdot p_{i,i+d}.$$ (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., $\hat{f}(x) = \frac{1}{k} \sum_{i=1}^{k} f_i(x)$. 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 $p_{i,i+2} = P_2 \cdot \frac{1}{2^k}$.

(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.

$$\hat{f}(x') \geq \hat{f}(x) \iff \sum_{i=1}^{k} f_i(x') \geq \sum_{i=1}^{k} f_i(x)$$
$$\iff k_1(n-i) + (k-k_1)(n-i-2) \geq k_2(n-i-1) + (k-k_2)(n-i+1)$$
$$\iff k_1 + k_2 \geq \frac{3k}{2}.$$ (2)

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 $p_{i,i+1} = P_1 \cdot \sum_{k_1+k_2 \geq \frac{3k}{2}} \binom{k}{k_1} \binom{k}{k_2} \frac{1}{2^k} \cdot \binom{k}{k_2} \frac{1}{2^k} = P_1 \cdot \sum_{n' \geq \frac{3k}{2}} \binom{k}{n'} \frac{1}{2^k}$.
(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 \( \hat{f}(x') < f(x) \), which can be simplified as follows.

\[
\hat{f}(x') < f(x) \Leftrightarrow \sum_{i=1}^{k} f^n_i(x') < \sum_{i=1}^{k} f^n_i(x) \\
\Leftrightarrow k_1(n-i) + (k-k_1)(n-i+2) < k_2(n-i+1) + (k-k_2)(n-i-1) \\
\Leftrightarrow k_1 + k_2 > \frac{3}{2} k.
\]

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 \( p_{i,i-1} = P_{-1} \cdot (1 - \sum_{k_2 > \frac{k}{2}} \binom{k}{k_2} \frac{1}{2^k}) \).

- \( 1 < i < n \). \( p_{i,i-1} = P_{-1} \cdot (1 - \sum_{k_1+k_2 > \frac{3}{2} k} \binom{k}{k_1} \binom{k}{k_2} \frac{1}{2^k}) = P_{-1} \cdot (1 - \sum_{k_2 > \frac{2k}{3} \binom{2k}{k} \frac{1}{2^k}}) \).

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

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

By applying these probabilities to Eq. [1], we have

\[
E[V(\xi_i) - V(\xi_{i+1}) \mid \xi_i = x] \ge p_{i,i-1} - p_{i,i+1} - 2 \cdot p_{i,i+2}.
\]

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

\[
E[V(\xi_i) - V(\xi_{i+1}) \mid \xi_i = x] \ge P_{-1} \cdot \frac{1}{2} \ge \frac{i}{n} \cdot \frac{1}{2} \ge \frac{i}{2en}.
\]

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

\[
E[V(\xi_i) - V(\xi_{i+1}) \mid \xi_i = x] \ge P_{-1} \cdot \frac{1}{2} - P_1 \ge \frac{i}{2en} - \frac{1}{n} \ge 0.01 \cdot \frac{i}{n},
\]

where the last inequality holds with \( n \ge 7 \).

(3) When \( 1 \le i < n - 1, p_{i,i+2} \le P_2 \cdot \frac{1}{2^k} \) and \( p_{i,i+1} = P_1 \cdot \sum_{k_2 \ge \frac{2k}{3} \binom{2k}{k} \frac{1}{2^k}} \). Let \( X_i (1 \le i \le 2k) \) be independent random variables such that \( P(X_i = 1) = \frac{1}{2} \) and \( P(X_i = 0) = \frac{1}{2} \). Then, \( \sum_{k_2 \ge \frac{2k}{3} \binom{2k}{k} \frac{1}{2^k}} = P(\sum_{i=1}^{2k} X_i \ge \frac{3}{2} k) \le e^{-\frac{k}{2}} \), where the “\( \le \)” is by Chernoff’s inequality. Thus, we have

\[
E[V(\xi_i) - V(\xi_{i+1}) \mid \xi_i = x] \ge \frac{P_{-1}}{2} - P_1 \cdot e^{-\frac{k}{2}} - 2 \cdot \frac{P_2}{2^k} \\
\ge \frac{i}{2en} - \frac{1}{n} \cdot \frac{1}{2} \ge 0.01 \cdot \frac{i}{n},
\]

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 \( n \ge 6 \).
Thus, the condition of Lemma 2 holds with $E [ V(\xi_t) - V(\xi_{t+1}) | \xi_t = x ] \geq \frac{0.01}{n} \cdot V(x)$. We then have

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

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 $O(n \log^2 n)$. □