This is the supplementary material for the paper "Initial Results on Runtime Analysis of $(1+1)$ Evolutionary Algorithm Controlled with Q-learning using Greedy Exploration Strategy on OneMax+ZeroMax Problem", submitted to the PPSN conference. It contains a proof of the $O(N \log N)$ bound on the number of fitness evaluations of the algorithm described in the paper on the OneMax+ZeroMax problem.

1 Excerpts from the Paper

For the sake of convenience, we outline here the necessary results from the paper.

The algorithm considered in the paper is the $(1+1)$ evolutionary algorithm solving the OneMax+ZeroMax problem. The objective selection is done using the EA+RL approach, which, in this paper, uses the greedy Q-learning reinforcement learning algorithm.

A Markov chain that is used to model this algorithm has $2^N \cdot N + 1$ states of the form $\langle B, V \rangle$, where B is the number of one-bits in the current bit vector and V is the set of reinforcement learning states, which have been left at least once (i.e. the reinforcement learning states where OneMax will always be selected). For convenience, we merge all the states with $B = N$, as they all are terminal states for the optimization process.

The mutation operator is used which flips each bit with the probability of $1/N$. The probability $P^{i,j}$ of going from the state that has i one-bits to the state that has j one-bits is:

$$P^{i,j} = \begin{cases} \sum\limits_{k=0}^{\min(N-j,i)} \binom{N-i}{j-i+k}\binom{i}{k}\left(\frac{1}{N}\right)^{j-i+2k}\left(1-\frac{1}{N}\right)^{N-(j-i+2k)} & \text{if } i < j, \\ \sum\limits_{k=0}^{\min(N-i,j)} \binom{i}{i-j+k}\binom{N-i}{k}\left(\frac{1}{N}\right)^{i-j+2k}\left(1-\frac{1}{N}\right)^{N-(i-j+2k)} & \text{if } i > j, \\ \sum\limits_{k=0}^{\min(N-i,i)} \binom{N-i}{k}\binom{i}{k}\left(\frac{1}{N}\right)^{2k}\left(1-\frac{1}{N}\right)^{N-2k} & \text{if } i = j. \end{cases} \quad (1)$$

In a state with the known optimal action ("gray states"), OneMax is selected with the probability of 1.0, and the mutation is accepted only if the number of ones does not decrease. This sums up to the following probabilities:

$$P_G^{i,j} = \begin{cases} P^{i,j} & \text{if } i < j, \\ 1 - \sum_{k=i+1}^{j} P^{i,j} & \text{if } i = j, \\ 0 & \text{if } i > j. \end{cases} \quad (2)$$

In a state with no known optimal action ("white states"), OneMax is selected with the probability of 0.5, as well as ZeroMax. In the case of OneMax, the mutations that do not decrease the number of ones are accepted, and in the case of ZeroMax, the opposite mutations are accepted. This results in the following:

$$P_W^{i,j} = \begin{cases} \frac{1}{2} P^{i,j} & \text{if } i < j, \\ \frac{1}{2} P^{i,j} & \text{if } i > j, \\ 1 - \frac{1}{2} \sum_{k \neq i} P^{i,k} = \frac{1}{2} \left(1 + P^{i,i}\right) & \text{if } i = j. \end{cases} \tag{3}$$

We define $E(B, V)$ as the expectation of the number of fitness evaluations needed to get to the terminal state from the state $\langle B, V \rangle$. It is assumed in this part of the paper that for any set V and any two $B_1 < B_2$ the inequality $E(B_1, V) > E(B_2, V)$ holds (the *non-increasing assumption*).

2 Proof of $O(N \log N)$ Bound Under Non-Increasing Assumption

This section is dedicated to the proof that the expectation of the number of fitness evaluations of the algorithm is $O(N \log N)$ provided the non-increasing assumption is true.

First, we will give lower and upper bounds on $P^{i,j}$, the components of transition probabilities in the Markov chain. Next, we show that if the algorithm never accepts mutations that increase the number of ones by more than one, the required expectation does not decrease. Finally, we simplify the Markov chain and prove the result.

2.1 Lower and Upper Bound on $P^{i,j}$

The following theorem gives a lower and an upper bound on $P^{i,j}$.

Theorem 1. *Assume that $i \neq j$. Let $S^{i,j}$ be the following:*

$$S^{i,j} = \begin{cases} \binom{N-i}{j-i} \left(\frac{1}{N}\right)^{j-i} \left(1 - \frac{1}{N}\right)^{N-(j-i)} & \textit{if } i < j, \\ \binom{i}{i-j} \left(\frac{1}{N}\right)^{i-j} \left(1 - \frac{1}{N}\right)^{N-(i-j)} & \textit{if } i > j. \end{cases} \tag{4}$$

Then $S^{i,j} \leq P^{i,j} \leq \frac{8}{7} S^{i,j}$.

Proof. The lower bounds are proven easily, since $S^{i,j}$ are the addends for $k = 0$ in (1), and all these addends are positive.

Now we are proving the upper bound. We denote as $S_k^{i,j}$ the k-th addend of the sum in (1) corresponding to $P^{i,j}$. Specifically, $S^{i,j} = S_0^{i,j}$.

Consider the case of $i < j$. The ratio of the k-th addend to the $(k+1)$-th addend is:

$$\frac{S_k^{i,j}}{S_{k+1}^{i,j}} = \frac{\binom{N-i}{j-i+k} \binom{i}{k} \left(\frac{1}{N}\right)^{j-i+2k} \left(1 - \frac{1}{N}\right)^{N-(j-i+2k)}}{\binom{N-i}{j-i+k+1} \binom{i}{k+1} \left(\frac{1}{N}\right)^{j-i+2k+2} \left(1 - \frac{1}{N}\right)^{N-(j-i+2k)-2}} =$$

$$= \frac{(j-i+k+1)(k+1)}{(N-j-k)(i-k)} N^2 \left(1 - \frac{1}{N}\right)^2 = \frac{(j-i+k+1)(k+1)}{(N-j-k)(i-k)} (N-1)^2.$$

When i and j are fixed, this ratio grows as k grows, so

$$\frac{S_k^{i,j}}{S_{k+1}^{i,j}} \geq \frac{S_0^{i,j}}{S_1^{i,j}} = \frac{j-i+1}{(N-j)i}(N-1)^2.$$

When i is fixed, this ratio grows as j grows, so we replace j with its minimum possible value $i+1$ and then minimize the result with $i = \frac{N-1}{2}$:

$$\frac{S_k^{i,j}}{S_{k+1}^{i,j}} \geq \frac{2(N-1)^2}{(N-i-1)i} \geq \frac{2(N-1)^2}{\frac{N-1}{2}\frac{N-1}{2}} = \frac{8(N-1)^2}{(N-1)^2} = 8.$$

This means that $P^{i,j}$ can be bounded by a sum of geometric progression:

$$P^{i,j} = \sum_{k=0}^{\min(N-j,i)} S_k^{i,j} \leq \sum_{k=0}^{\min(N-j,i)} \left(\frac{1}{8}\right)^k S_0^{i,j} \leq \sum_{k=0}^{\infty} \left(\frac{1}{8}\right)^k S_0^{i,j} = \frac{8}{7} S_0^{i,j} = \frac{8}{7} S^{i,j}.$$

The case of $i > j$ is proven in the same way. □

2.2 Model Simplification