

Sujay Jakka

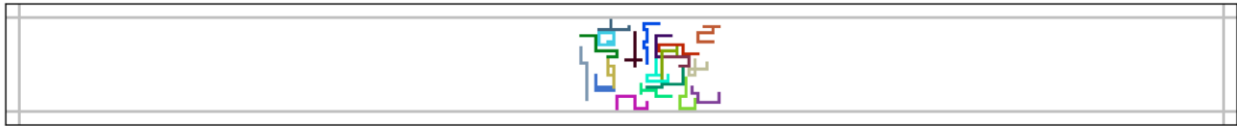
Svj0007@auburn.edu

COMP 5660 Fall 2024 Assignment 1c

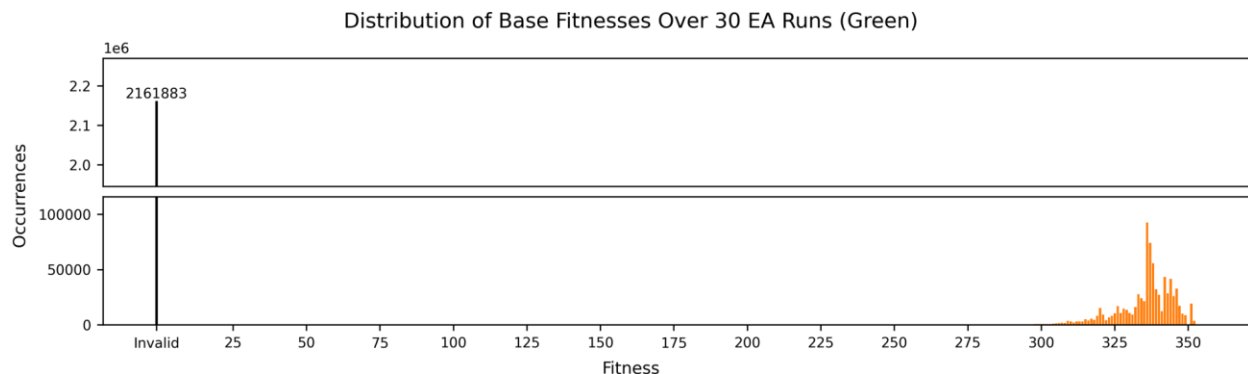
6 October 2024

### Assignment 1c Report

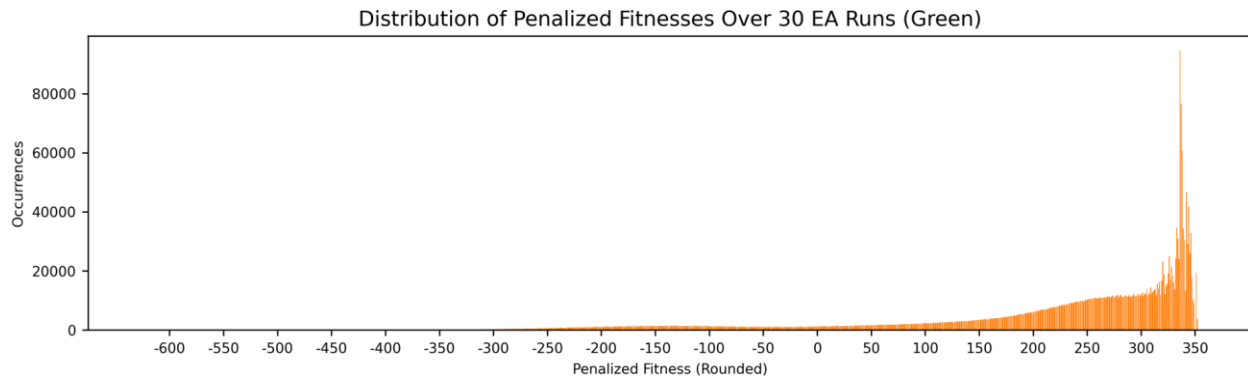
I will first introduce the results for the green deliverable. The best base fitness my EA achieved across 30 runs was 353. The image below shows the visualization of the best solution from my EA.



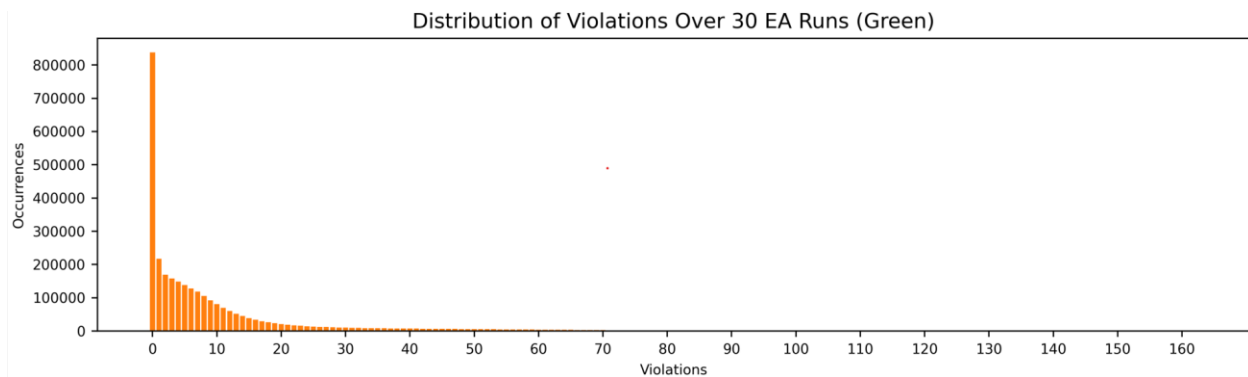
Looking at the distribution of base fitness over 30 runs of my EA, I had 2,161,883 invalid solutions which was approximately 72.06% of all solutions. The image below shows the distribution of base fitness across 30 runs.



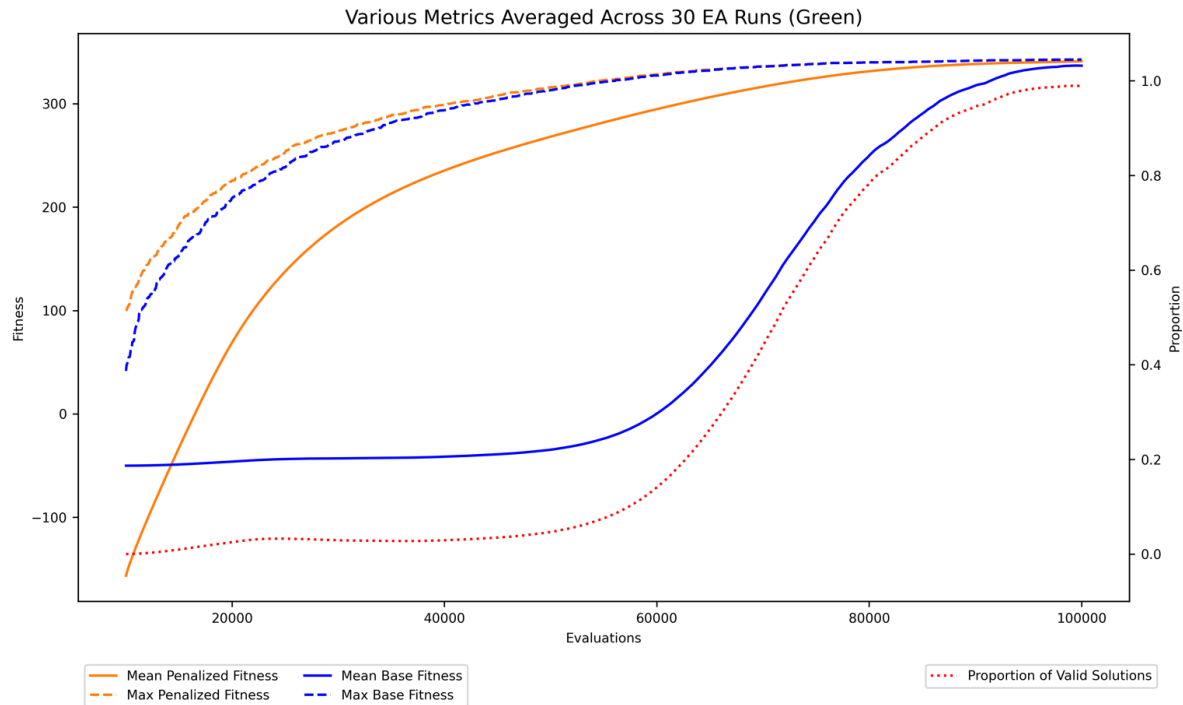
Furthermore, looking at the distribution of penalized fitness over 30 runs of my EA, I had a distribution that was skewed left. It seems that most of the individuals had penalized fitness between 200 and 350 with a sharp peak at approximately 340. The image below shows the distribution of penalized fitness across 30 runs.



According to the violations histogram below, the distribution seems to be skewed right with a sharp peak at 0 violations. Approximately, 27% of individuals across 30 runs had 0 violations.



From the plot below, it is evident that the EA did not have premature convergence of the max penalized fitness and mean penalized fitness or of the max base fitness and mean base fitness. Additionally, the proportion of valid solutions reached almost 100% only at the end of the life of the EA indicating that the EA was able to properly explore the genotype landscape without focusing only on valid solutions.



Furthermore, the mean of the sample of 30 runs(sample size 30) of the 1c EA was 342.83, while the mean of the sample of 30 runs(sample size 30) of my 1b EA was 290.87. The standard deviation of the 1c and 1b EA was 5.04 and 17.09 respectively. I performed a two sample independent t-test where the null hypothesis was that the two algorithms are equally effective at solving the problem, and the alternative hypothesis is that the two algorithms are not equally effective at solving the problem instance. I performed the t-test with a significance level(alpha) of 0.05 and a degrees of freedom of 30. I got a p-value of 2.2572e-17, which is significantly smaller than the significance level of 0.05. We can then reject the null hypothesis, and claim that the two algorithms have statistically significant differences in performance. Because it is highly unlikely that both sample distributions have the same population mean and that the 1c EA sample has a higher mean than the 1b EA sample, we can conclude that the 1c EA performed better. The parameters I used for the green deliverable are shown below.

---

[ea]

```
mu = 10_000
num_children = 50
mutation_rate = 0.15
parent_selection = k_tournament_with_replacement
survival_selection = k_tournament_without_replacement
```

```
[recombination_kwargs]
method = uniform
probab_selecting_parent_1_gene = 0.5
```

```
[parent_selection_kwargs]
k = 15
```

```
[survival_selection_kwargs]
k = 15
```

```
[fitness_kwargs]
penalty_coefficient = 4
```

```
[mutation_kwargs]
probab_creep_mutation = .15
mean_of_change_dist = 0
std_of_change_dist = 3
```

prob\_random\_reset = .15

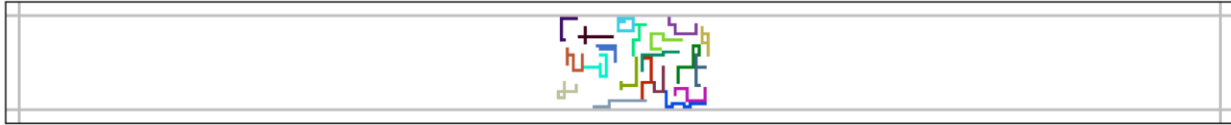
---

Before introducing the results of the yellow deliverable, I will describe the algorithm I chose for the adaptive penalty coefficient. I adopted the method from the textbook, which can be seen on page 135. I had three parameters beta 1, beta 2, and k. The algorithm changes the penalty coefficient before performing parent selection for the next generation. The algorithm would look at the individual with the best fitness in each of the last k generations. If these individuals all had 0 violations, it would reduce the penalty coefficient by a factor of beta 1. However, if these individuals all had at least one violation, I would increase the penalty coefficient by a factor of beta 2. Lastly, if there is at least one individual with 0 violations and at least one individual with 1 violation in the best fitness individuals in each of the last k generations, then we keep the penalty coefficient the same. The image below describes the algorithm.

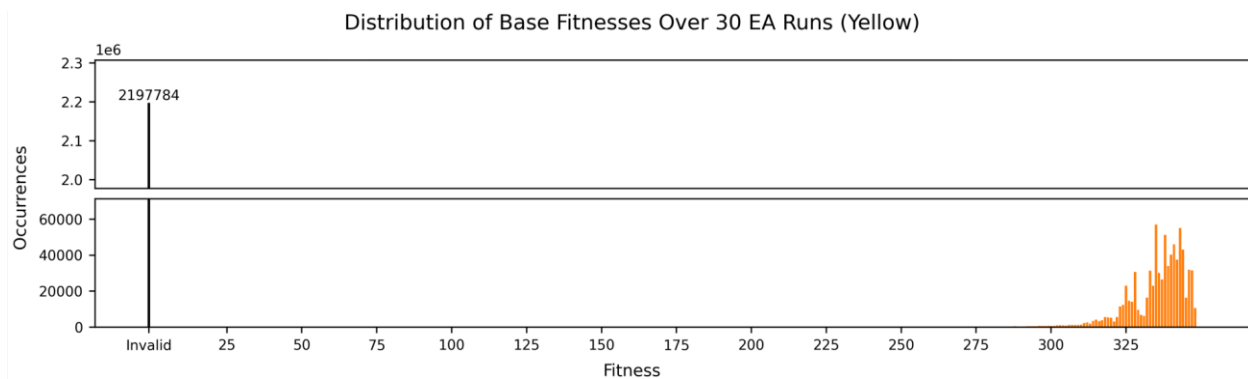
$$W(t+1) = \begin{cases} (1/\beta_1) \cdot W(t) & \text{if } \bar{b}^i \in \mathcal{F} \quad \text{for all } t-k+1 \leq i \leq t, \\ \beta_2 \cdot W(t) & \text{if } \bar{b}^i \in \mathcal{S} - \mathcal{F} \text{ for all } t-k+1 \leq i \leq t, \\ W(t) & \text{otherwise.} \end{cases}$$

In this formula,  $\mathcal{S}$  is the set of all search points (solutions),  $\mathcal{F} \subseteq \mathcal{S}$  is a set of all *feasible* solutions,  $\bar{b}^i$  denotes the best individual in terms of the function *eval* in generation  $i$ ,  $\beta_1, \beta_2 > 1$ , and  $\beta_1 \neq \beta_2$  (to avoid cycling).

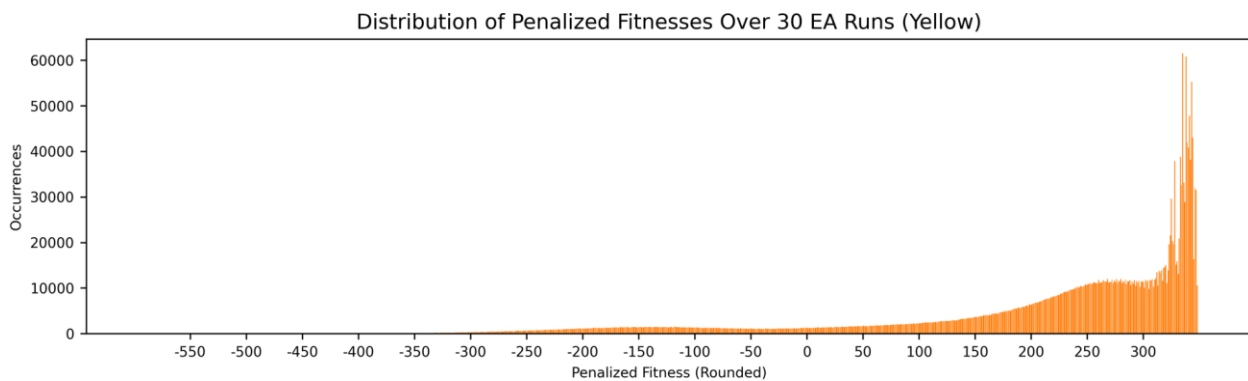
When I performed this experiment, I kept all other parameters the **same** as the green deliverable, however I added my beta 1, beta 2, and k parameters under the fitness\_kwargs. I set b\_1 = 10003/10000, b\_2 = 10001/10000, and lastly k = 20. The best base fitness my EA achieved across 30 runs was 349 which was less than my best solution of the green deliverable which was 353. The image below shows the visualization of the best solution from my 1c yellow EA.



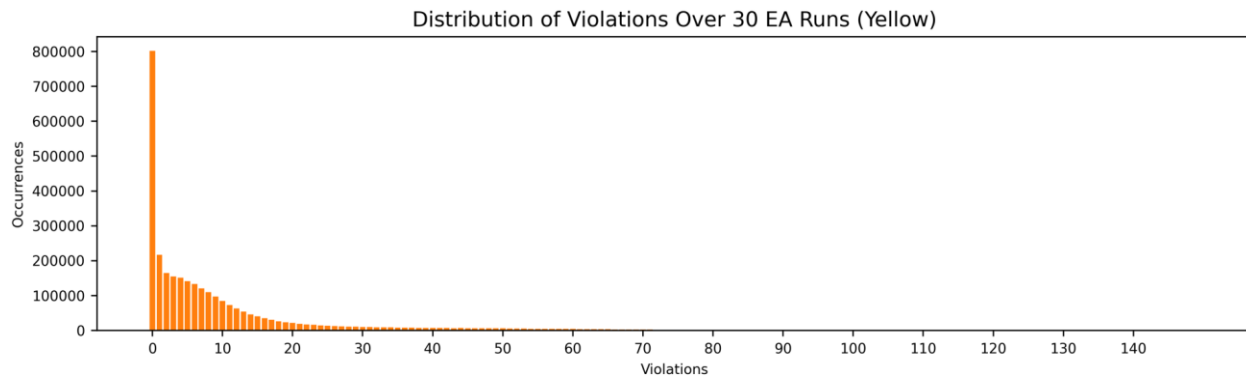
Looking at the distribution of base fitness over 30 runs of my EA, I had 2,197,784 invalid solutions which was approximately 73.26% of all solutions. The image below shows the distribution of base fitness across 30 runs.



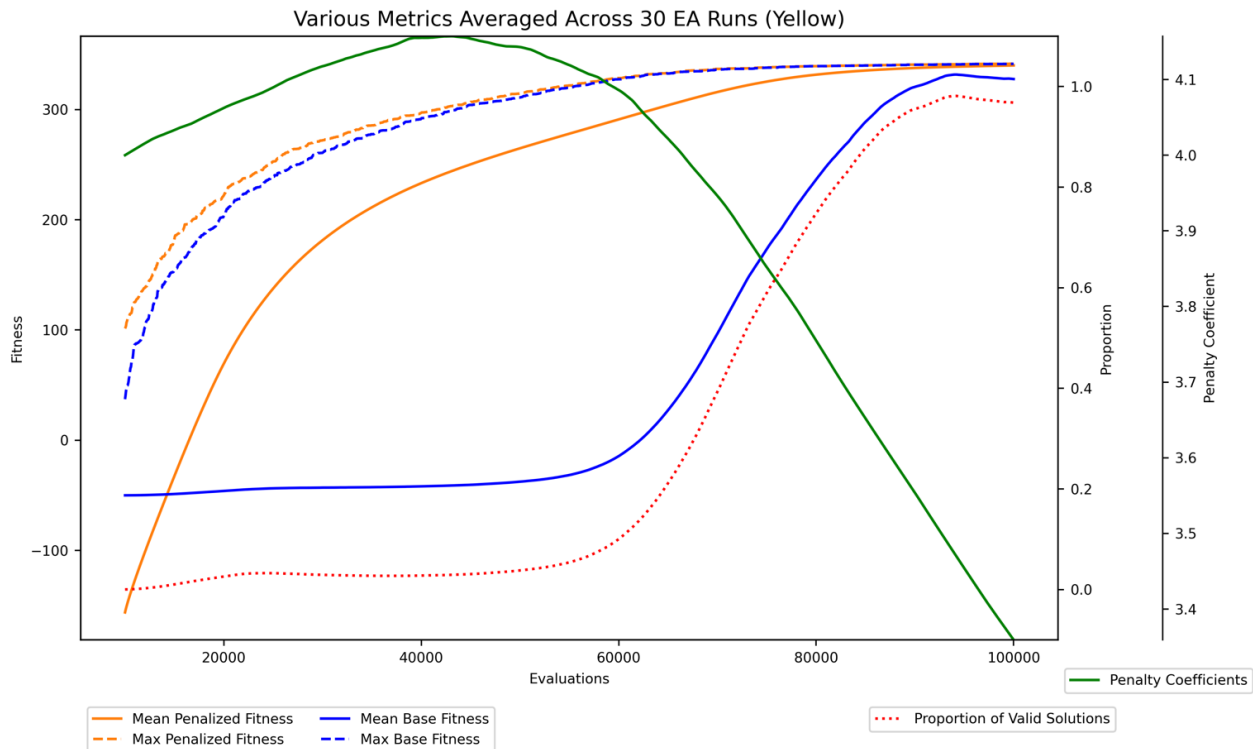
Furthermore, looking at the distribution of penalized fitness over 30 runs of my EA, I had a distribution that was skewed left. It seems that most of the individuals had penalized fitness between 200 and 340 with a sharp peak at approximately 335. The image below shows the distribution of penalized fitness across 30 runs.



According to the violations histogram below, the distribution seems to be skewed right with a sharp peak at 0 violations. Approximately, 26% of individuals across 30 runs had 0 violations.



From the plot below, it is evident that the EA did not have premature convergence of the max penalized fitness and mean penalized fitness or of the max base fitness and mean base fitness. Additionally, the proportion of valid solutions reached approximately 95% only at the end of the life of the EA indicating that the EA was able to properly explore the genotype landscape without focusing only on valid solutions. Lastly, the penalty coefficient seemed to have an inverse relationship with the proportion of valid solutions curve. This makes a lot of sense because in earlier generations most of the population is of individuals with violations. It makes sense that the EA would like to see more valid solutions so the penalty coefficient was higher in earlier generations. Once the number of valid solutions started to increase in the population, it makes sense for the EA to want to encourage exploration causing the penalty coefficient to decrease.



Lastly, the mean of the sample of 30 runs(sample size 30) of the 1c Yellow EA was 341.50, while the mean of the sample of 30 runs(sample size 30) of my 1c Green EA was 342.83. The standard deviation of the 1c Yellow and 1c Green EA was 5.35 and 5.04 respectively. I performed a two sample independent t-test where the null hypothesis was that the two algorithms are equally effective at solving the problem, and the alternative hypothesis is that the two algorithms are not equally effective at solving the problem instance. I performed the t-test with a significance level(alpha) of 0.05 and a degrees of freedom of 30. I got a p-value of 0.325, which is significantly larger than the significance level of 0.05. We **cannot** reject the null hypothesis, meaning that the two algorithms do not have statistically significant differences in performance.