CS5401 FS2018 Assignment 1d

William Lorey
wwlytc@mst.edu

Contents

Introduction

Assignment 1d involved implementing a Multi-Objective Evolutionary Algorithm (MOEA) to more effectively solve Light Up puzzles by balancing the fulfillment of three objectives:

1. maximize the number of cells lit up (represented in this implementation as a ratio of lit cells to the total number of white cells)
2. minimize the number of bulbs shining on each other
3. minimize the number of black cell adjacency constraint violations

For BONUS #1, a fourth objective was added, namely minimizing the number of bulbs placed on the board.

This report outlines this solution's particular implementation of a MOEA, the impact of initialization strategies on the MOEA's performance, a comparison between parent selection, survival strategy, and survival selection strategies on MOEA performance, as well as the impact of increasing the number of objectives on non-domination and MOEA performance (BONUS #1).

MOEA Overview

The MOEA implemented in this assignment is based on the NSGA-II algorithm. It begins, similar to a standard evolutionary algorithm, by creating an initial population using either uniform random or validity enforced plus uniform random initialization, the settings for which are specified in the algorithm configuration file. That population is evaluated and the subfitnesses are determined and assigned to each individual in the population.

The population is then evaluated on the basis of non-domination. A list of Pareto fronts is created from the initial population where all genotypes in a given front are not dominated by any other genotypes in that front while genotypes in higher level fronts are dominated by genotypes in lower level fronts. The 'best' genotypes, those in the best level of non-domination, are assigned to level number one. Subsequent levels increase in increments of one for other levels of non-domination.

The fitness of each genotype is then set to its level in the list of Pareto fronts, with individuals exhibiting a smaller fitness (level number) are more fit. A binary tournament selection is performed to choose breeding parents. Then offspring are created using an n-point crossover recombination (with n determined in the configuration file). Following that, mutation is performed, completing the child population.

For the standard NSGA-II configuration (exhibited in the deliverables configuration folder), the plus survival strategy is exhibited, combining the children and parent populations into one large population from which to choose the new population. Individuals are then selected for survival using a binary tournament selection and the process is repeated using the new population until the end of the experiment.

Impact of Initialization on MOEA Performance

The effect of Validity Enforced plus Uniform Random versus Uniform Random initialization was examined in this experiment for both the provided puzzle and randomly generated puzzles. One would assume that an initialization method utilizing Validity Enforced initialization would outperform a solely Uniform Random initialization. After performing statistical analysis on the experiment data, it was concluded that there is in fact no tangible difference between the initialization methods for the tested puzzles. Table 1 and Table 2 each display statistical analysis supporting this finding.

This statistical analysis for this section was performed on the average of all last best subfitnesses, giving thirty data points to compare for each MOEA initialization method. The analysis consisted of performing an f-test, which determined if variances could be treated as equal. In both cases, the f-test yielded that unequal variances should be assumed. Following the f-test, the two-tailed t-test was performed assuming unequal variances. This test yielded (in both cases) that neither initialization method was statistically better for the set of Light Up puzzles tested.

To visually interpret the data, plots of evaluations versus average local subfitness and evaluations versus local best subfitness were graphed for each of the subfitnesses collected in this experiment (not including the bonus): ratio of lit cells to total number of white cells, number of bulbs shining on each other, and number of black cell constraints not met. For concision, figures pertaining to only the provided puzzle are discussed in this section as the randomly generated puzzle results are quite similar. Figures 1, 2, 3, 7, 8, and 9 depict experiment plots for the provided puzzle while figures 4, 5, 6, 10, 11, and 12 depict experiment plots for the randomly generated puzzle.

Figure 1 and Figure 7 depict the lit cell ratio subfitness plots for the Validity Enforced plus Uniform Random and Uniform Random initialized experiments, respectively. In the uniform random case (Figure 7), both the average local list cell ratio and the local best lit cell ratio started lower than those of the validity enforced experiment. However, as the experiments progressed, both the average and best subfitnesses of each experiment reached appreciatively the same point without the best subfitness plateauing. This implies that letting the experiment run for longer would produce a more fit solution, with respect to the lit cell ratio. Note that the lit cell ratio subfitness, the metric was maximized.

Figure 2 and Figure 8 depict the evaluations versus subfitness plots for the bulb shine constraint violations. These plots behaved quite similarly to the lit cell ratio plots in that the Validity Enforced plus Uniform Random plot started with a higher average number of bulb shine constraints violated while the plain Uniform Random plot had a lower number of initial constraint violations. This is logical as enforcing validity has the potential to place more bulbs, which creates opportunity for more bulbs to shine on each other, further adding to the number of bulb shine violations. The local best for both plots stayed right at zero bulb constraint violations, implying that most, if not all, experiments always had at least one member of the population with no bulb shine constraints. Because this subfitness is to be minimized as part of the MOEA, it is peculiar that the average number of bulb constraints increases as the experiments continue. This implies that as more bulbs are placed on the board, the multi-objective nature of the algorithm allows for more and more bulbs to shine on each other, keeping those individuals with more bulb-shine in the population so long as the levels of non-domination dictate it.

The third subfitness, black cell constraint violations, was examined for both initialization schemes in Figure 3 and Figure 9. Before examining these plots, it was hypothesized that the Validity Enforced plus Uniform Random initialization method would be superior with respect to minimizing the black cell constraint violations when compared to the Uniform Random only initialization. While the statistical analysis was not granular enough to definitively prove this, the graphs provide anecdotal evidence that the Validity Enforced plus Uniform Random initialization method is in fact superior when it comes to the black cell constraint violations. This conclusion was drawn because the Validity Enforced plus Uniform Random method had lower average and local best black cell constraints violated at each point on the graph, across all experiments.

Comparison of Parent Selection Strategy, Survival Strategy, and Survival Selection Strategy MOEA Configurations

All combinations of the following MOEA configurations were compared, for both randomly generated puzzles and the provided puzzle, to determine which configuration combination was optimal.

- Parent Selection
  - Fitness Proportional
  - Binary Tournament
- Survival Selection

- Fitness Proportional
- Binary Tournament

- Survival Strategy
  - Plus
  - Comma

Each combination of configurations resulted in $n = 2^3 = 8$ distinct configurations. Each configuration was tested against each other configuration (excluding itself), resulting in

$$n * (n - 1) = 8 * 7 = 56$$

comparisons per puzzle type. 56 comparisons were performed for both the provided puzzle and the randomly generated puzzle, resulting in 112 comparisons total.

The explicit statistical analysis for each individual comparison is not tabulated in this report for brevity, however, the statistical analysis methodology is described as follows: First, the last best subfitness for each of the three subfitness were aggregated and written to separate files for each of the configuration schemes listed above. This resulted in $3 * 8 = 24$ data files to drive the statistical analysis for each the provided puzzle and the randomly generated puzzle (48 in total). Following this, each possible pairing (in which order does matter) was performed, comparing from each configuration each of the three last best subfitness lists using first an f-test to determine whether or not equal variances could be assumed followed by a two-tailed t-test assuming either equal or unequal variances (depending on the outcome of the f-test). The f-test would then yield if a configuration was better with respect to each subfitness. If the number of a configuration's subfitness' superiorities outnumbered the number of a configuration's subfitness' subordinates for a given comparison, that configuration was marked as being better than the other, incrementing a dominance count associated with that configuration. The configuration(s) with the highest dominance count are statistically better than all other configurations tested for the given test puzzles.

The randomly generated puzzle configurations yielded two configurations that were both equally as optimal (they both dominated two other configurations, while all other configurations dominated no other configurations). These were MOEAs configured with (1) fitness proportional parent selection, tournament survival selection, and plus survival strategy and (2) tournament parent selection, tournament survival selection, and plus survival strategy. Both of these solutions each were dominant over two other configurations out of all the other configurations. They were either at least as good as or subordinate to the other configurations. The outcome of these comparisons proved that for this MOEA tested against randomly generated puzzles, the plus method for survival strategy is optimal, as both dominant configurations employed this strategy for at least one selection. The results also showed that relatively *low* selection pressure for parent and survival selections led to better algorithm results. The binary tournament selection applies low selection pressure and is prevalent in both configurations - the first configuration used binary tournament selection for survival selection while the second used a binary tournament selection for both the parent and survival selections. This method of selection favors exploration over exploitation in traversing the solution search space for optimal bulb placements.

The provided puzzle configurations yielded one configuration that was optimal when compared to all other solutions. In fact, this configuration was the only solution found to be statistically better than all other configurations. This configuration involved a tournament parent selection and tournament survival selection with a plus survival strategy. Similarly to the above dominance results, the results for the provided puzzle showed that configurations employing a low selection pressure and a plus survival strategy produced better results. This dominance was more prevalent for the provided puzzle, as the winning configuration dominated all other configurations. Again, this proves that for this particular MOEA, an approach of exploration over exploitation is integral in achieving more optimal results.

BONUS #1: Impact of Increasing Number of Objectives on MOEA Performance

For the first bonus, a fourth subfitness was added to the MOEA, namely a constraint minimizing the number of bulbs placed on the board. This constraint directly contradicted one of the main measures of an EA's performance - the ratio of lit cells to the total number of white cells. For an MOEA to be successful with only the lit cell ratio, bulb shine constraint violations, and black cell constraint violations subfitnesses, it did not explicitly need to minimize the number of bulbs placed. In fact, the only thing *limiting* the number of bulbs placed on the board was the bulb shine constraint violation metric. By adding the fourth subfitness of minimizing the number of placed bulbs, the MOEA is more likely to be conservative regarding the placement of bulbs on a board, as now half of the four total subfitness metrics actively limit the number of placed bulbs. Although statistical analysis was not performed regarding the implication of an increased number of subfitness objectives and its impact on MOEA performance, the following hypothesis is offered instead: If a fourth subfitness objective is added which minimizes the number of bulbs placed on the board, the best level of non-domination will (overall, across many runs) prove to contain genotypes with lower lit-cell ratios as the incentive for placing more bulbs will be limited while the incentive for lighting up many cells will stay the same.

Appendix: Figures and Tables

Note that figures and tables found in this section are referenced above in the body of the document.

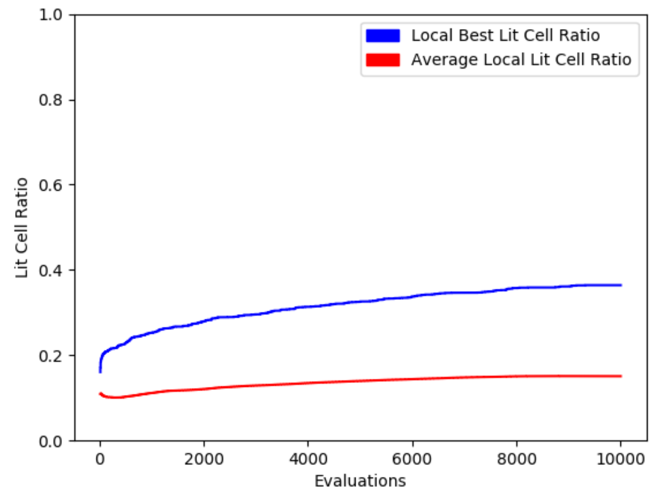

Figure 1: Evaluations versus Average Local Subfitness and Evaluations versus Local Best Subfitness for the **Lit Cell Ratio Subfitness, Validity Enforced plus Uniform Random Initialized, Provided Puzzle**, Averaged Over All Runs

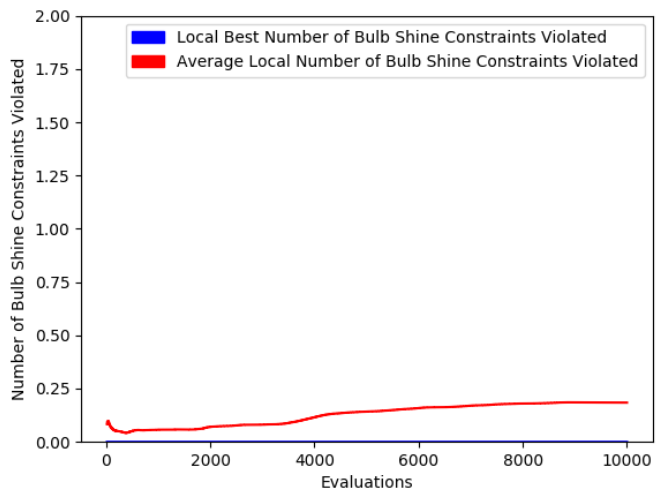

Figure 2: Evaluations versus Average Local Subfitness and Evaluations versus Local Best Subfitness for the **Bulb Shine Constraint Subfitness, Validity Enforced plus Uniform Random Initialized, Provided Puzzle**, Averaged Over All Runs

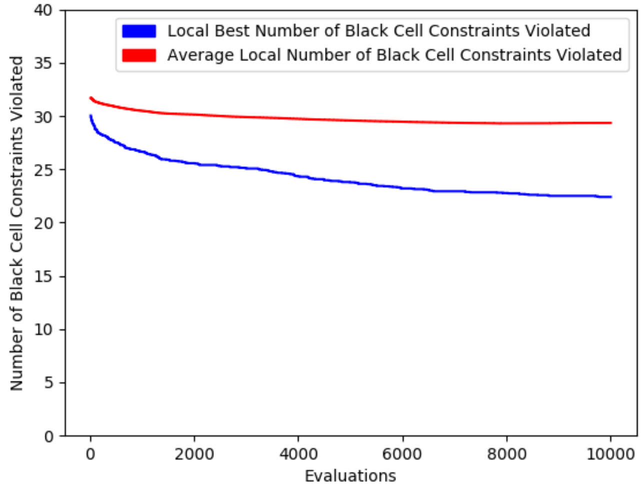

Figure 3: Evaluations versus Average Local Subfitness and Evaluations versus Local Best Subfitness for the **Black Cell Constraint Subfitness, Validity Enforced plus Uniform Random Initialized, Provided Puzzle**, Averaged Over All Runs

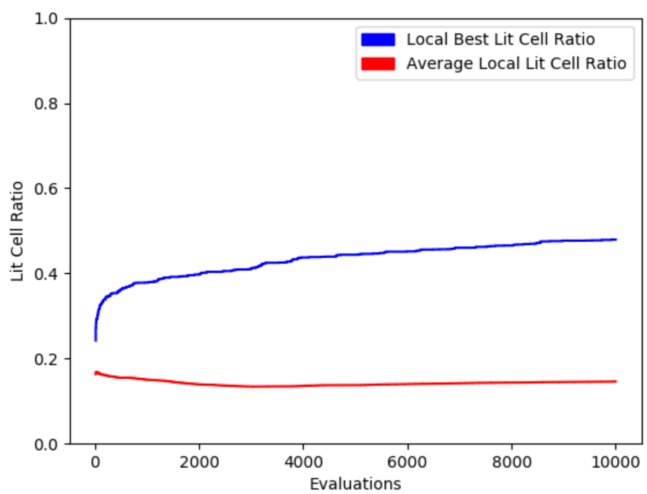

Figure 4: Evaluations versus Average Local Subfitness and Evaluations versus Local Best Subfitness for the **Lit Cell Ratio Subfitness, Validity Enforced plus Uniform Random Initialized, Randomly Generated Puzzle**, Averaged Over All Runs

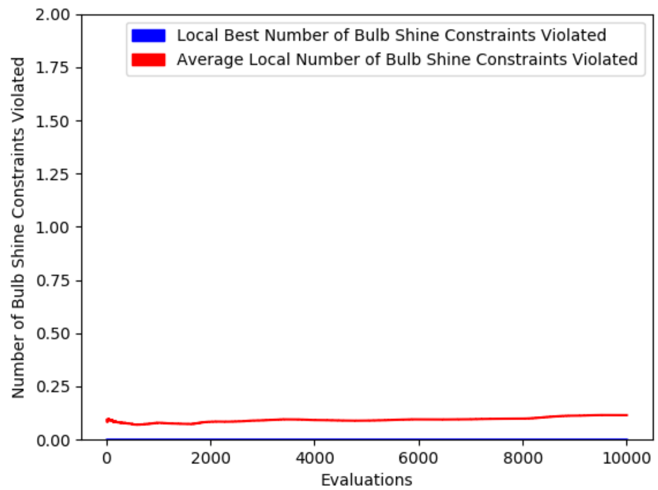

Figure 5: Evaluations versus Average Local Subfitness and Evaluations versus Local Best Subfitness for the **Bulb Shine Constraint Subfitness, Validity Enforced plus Uniform Random Initialized, Randomly Generated Puzzle**, Averaged Over All Runs

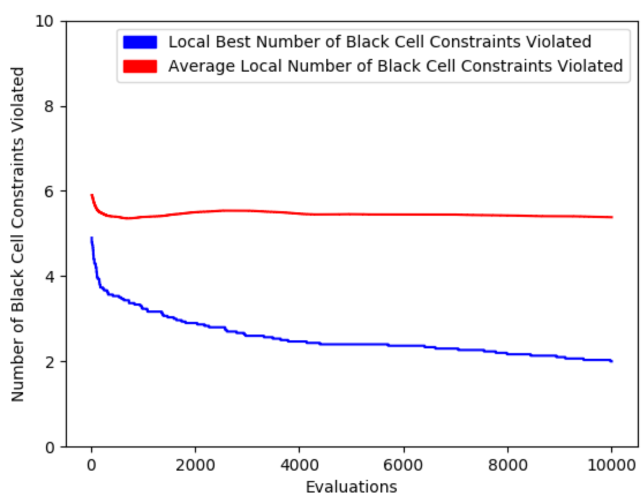

Figure 6: Evaluations versus Average Local Subfitness and Evaluations versus Local Best Subfitness for the **Black Cell Constraint Subfitness, Validity Enforced plus Uniform Random Initialized, Randomly Generated Puzzle**, Averaged Over All Runs

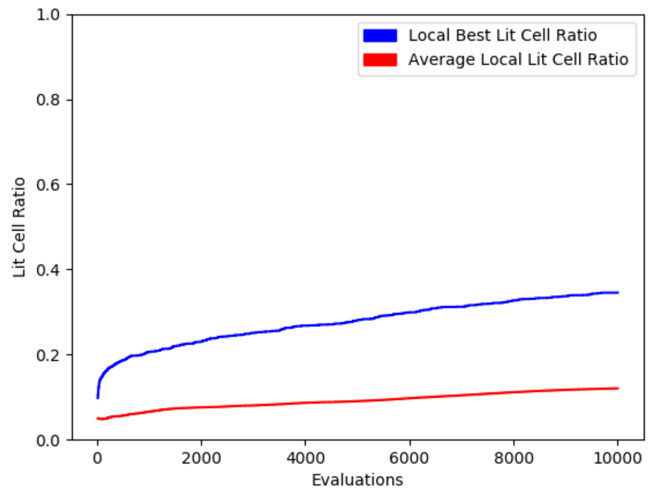

Figure 7: Evaluations versus Average Local Subfitness and Evaluations versus Local Best Subfitness for the **Lit Cell Ratio Subfitness, Uniform Random Initialized, Provided Puzzle**, Averaged Over All Runs

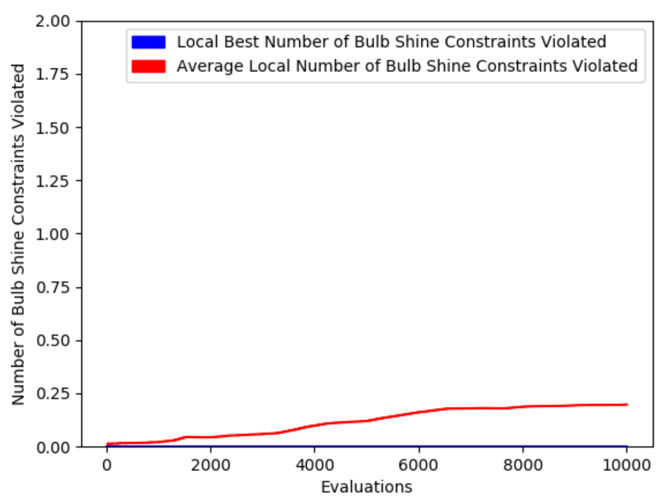

Figure 8: Evaluations versus Average Local Subfitness and Evaluations versus Local Best Subfitness for the **Bulb Shine Constraint Subfitness, Uniform Random Initialized, Provided Puzzle**, Averaged Over All Runs

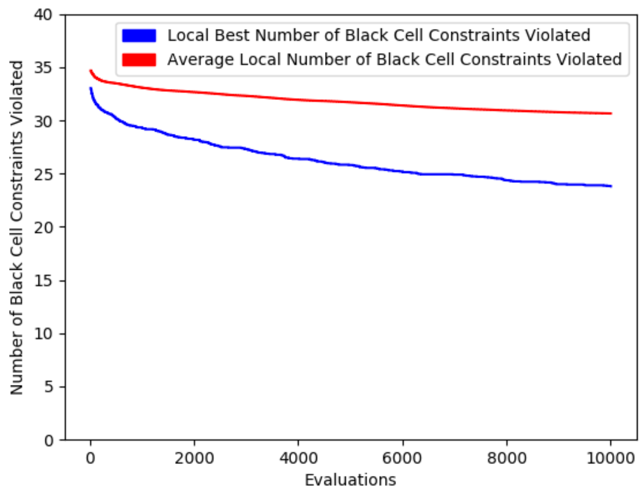

Figure 9: Evaluations versus Average Local Subfitness and Evaluations versus Local Best Subfitness for the **Black Cell Constraint Subfitness, Uniform Random Initialized, Provided Puzzle**, Averaged Over All Runs

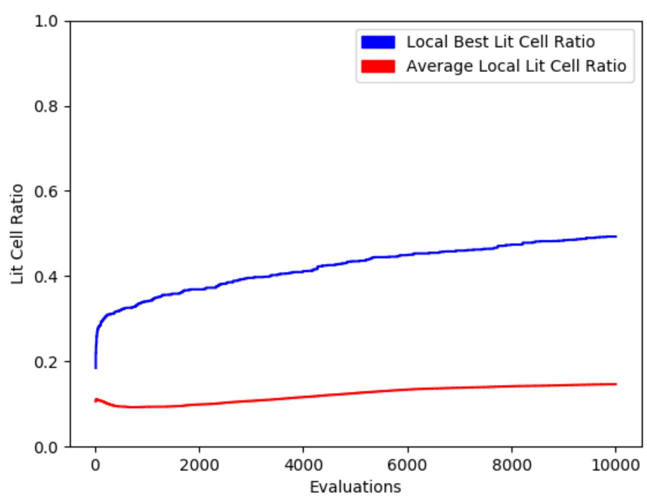

Figure 10: Evaluations versus Average Local Subfitness and Evaluations versus Local Best Subfitness for the **Lit Cell Ratio Subfitness, Uniform Random Initialized, Randomly Generated Puzzle**, Averaged Over All Runs

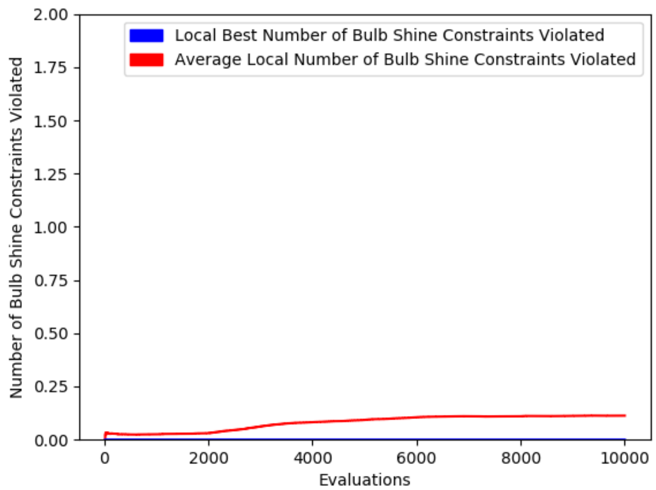

Figure 11: Evaluations versus Average Local Subfitness and Evaluations versus Local Best Subfitness for the **Bulb Shine Constraint Subfitness, Uniform Random Initialized, Randomly Generated Puzzle**, Averaged Over All Runs

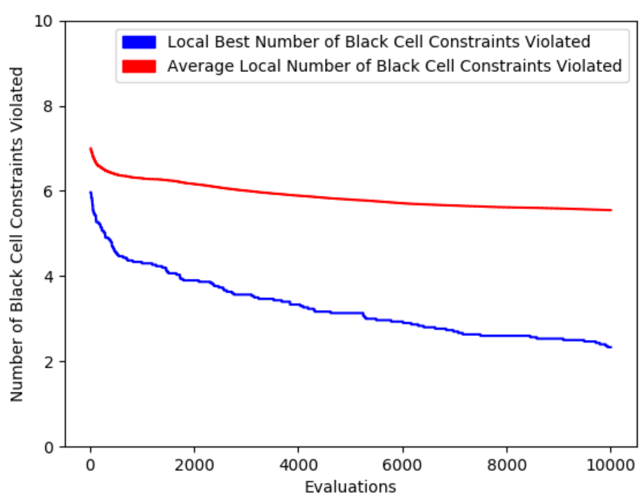

Figure 12: Evaluations versus Average Local Subfitness and Evaluations versus Local Best Subfitness for the **Black Cell Constraint Subfitness, Uniform Random Initialized, Randomly Generated Puzzle**, Averaged Over All Runs

Table 1: Statistical Analysis performed on the Uniform Random and Validity Enforced Uniform Random Initialized, Randomly Generated Puzzle, EA configurations

	random_gen	random_gen_uniform_random_init
mean	1.10325284795678	1.0764116503308314
variance	0.04970344216568983	0.04355857958196582
standard deviation	0.22294268807406498	0.20870692269775296
observations	30	30
df	29	29
F	1.141071234248146	
F critical	0.5373999648406917	
Unequal variances assumed		
observations	30	
df	31	
t Stat	0.47331304656977363	
P two-tail	0.6377741699825987	
t Critical two-tail	2.0395	
Nether random_gen_uniform_random_init nor random_gen is statistically better		

Table 2: Statistical Analysis performed on the Uniform Random and Validity Enforced Uniform Random Initialized, Provided Puzzle, EA configurations

	website_puzzle	website_puzzle_uniform_random_init
mean	0.8031737242867948	0.7959489651519909
variance	0.000603666878279215	0.00058947613524421
standard deviation	0.024569633254878164	0.02427912962287178
observations	30	30
df	29	29
F	1.02407348183676	
F critical	0.5373999648406917	
Unequal variances assumed		
observations	30	
df	31	
t Stat	1.1263571505364551	
P two-tail	0.26465388827990055	
t Critical two-tail	2.0395	
Nether website_puzzle_uniform_random_init nor website_puzzle is statistically better		