Mirela Chiţaniuc, MSAI2
Gâdioi Alexandra, MOC2

# Experimental analysis of algorithms
# DNA Word Design Report

## Problem description

The DNA code design problem is: given a target **k** and a word length **n** is simply a string of length n over the alphabet A, C, G, T, and naturally corresponds to a DNA strand. The constraints that we consider are:

- **Hamming Distance Constraint (HD)** The Hamming distance between any two distinct words must be greater than d
- **Content Constrain (GC)** There is a fixed amount of G and C nucleotides
- **Reverse Complement Hamming Distance Constraint (RC)** The Hamming distance between a word and another word's reverse must be greater than d

In this report we can see differences between Stochastic Local Search and a variant of Stochastic Local Search with probabilities in solving the DNA Word Design problem.

## Stochastic Local Search

Local search is a heuristic method for solving computationally hard optimization problems. Local search can be used on problems that can be formulated as finding a solution maximizing a criterion among a number of candidate solutions.

## Stochastic Local Search with probabilities

Stochastic Local Search with probabilities is a custom-made variant of Stochastic Local Search. It operates almost in the same manner: it doesn't build up the M sets, but instead, generates a new random word with a certain probability.

## Time and input size

Here we can see how the execution time evolves accordingly to the parameter k, number of elements in the words set.

Mirela Chiţaniuc, MSAI2
Gâdioi Alexandra, MOC2

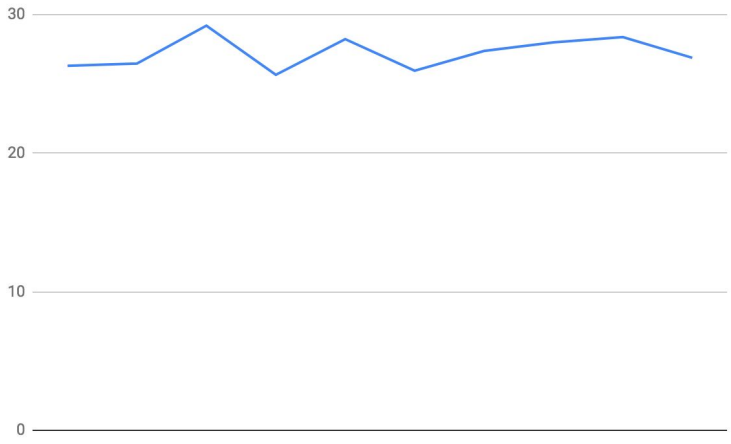


Figure 1: Execution time Stochastic local search

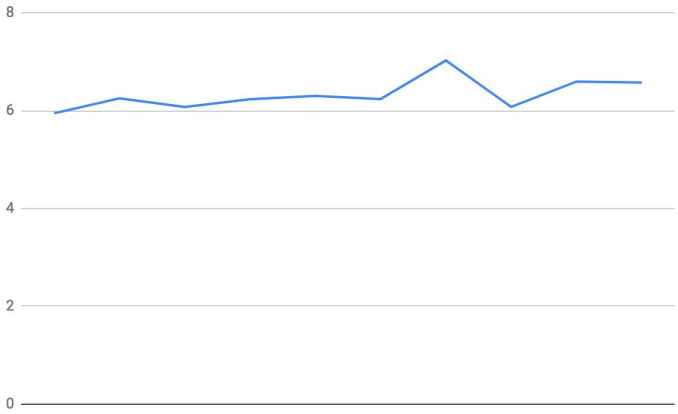


Figure 2: Execution time for Stochastic local search with optimization

Mirela Chiţaniuc, MSAI2
Gâdioi Alexandra, MOC2

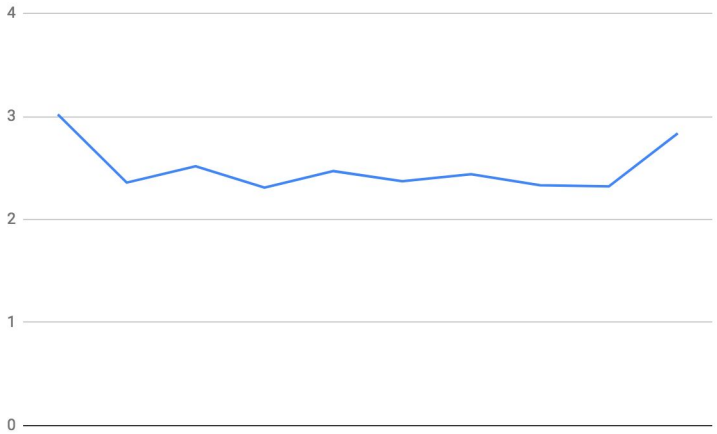


Figure 3: Execution time for Stochastic local search with probabilities

The optimization for the local search is done in the construction of the sets M1 and M2.The sets are, initially, combinations of all changes for each gene in the word. Because the constraints are parsing the words and find genes that are not according the rules, than the sets M1 and M2 can be formed only from words with all the genes the same, but with the marked gene changed.

The number of checkings done for each set is almost the same in both cases. Errors are counted each time one of the constraints is not satisfied.

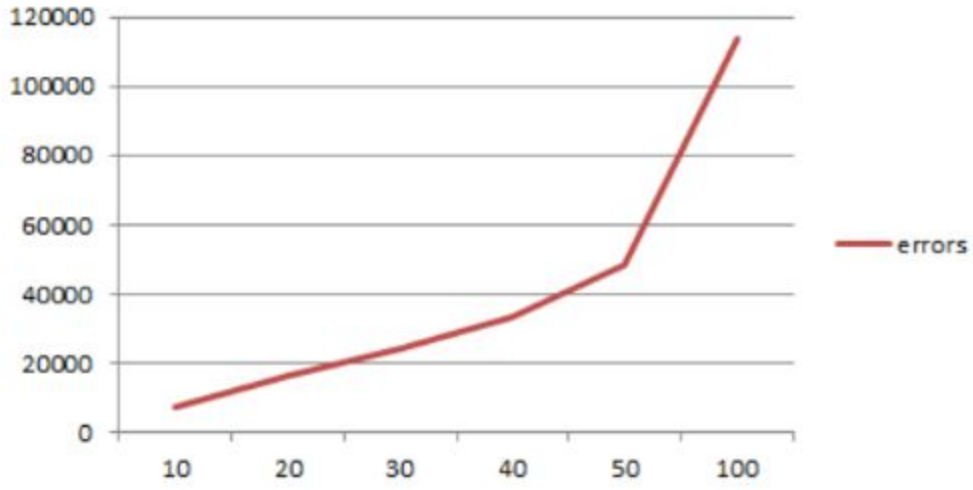


Figure 4: Errors for Stochastic Local Search

Mirela Chiţaniuc, MSAI2
Gâdioi Alexandra, MOC2

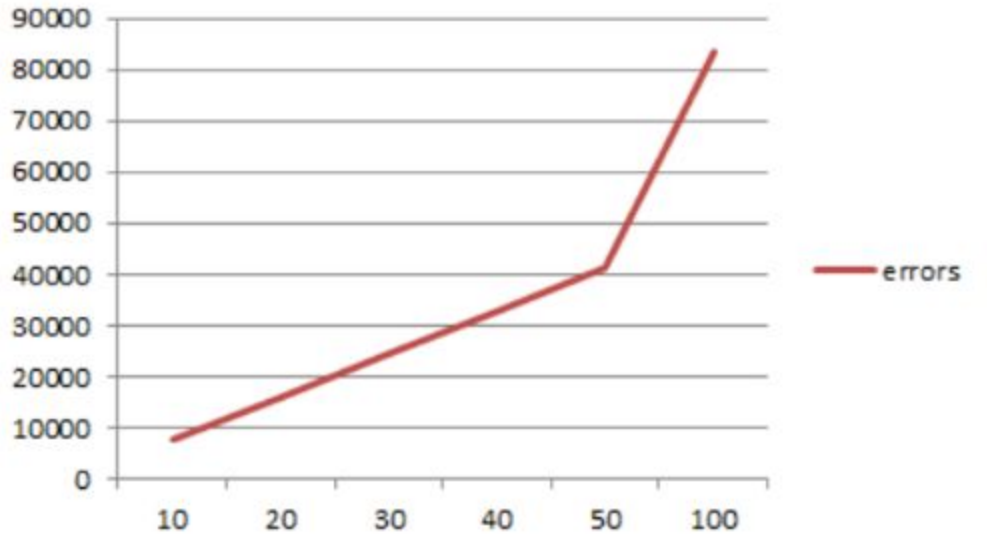


Figure 5: Errors for Stochastic Search with probabilities

**Confidence intervals**

Stochastic Local Search
('mean: ', 27.251400733000004)
('start: ', 26.40865781241807, ' end: ', 28.094143653581938)

Stochastic Local Search with optimizations
('mean: ', 6.3319474459000009)
('start: ', 6.1038655202241419, ' end: ', 6.56002937157586)

Stochastic Local Search with probabilities
('mean: ', 2.4976360321)
('start: ', 2.3255941214511595, ' end: ', 2.6696779427488404)

**T-test**

Stochastic Local Search **&** Stochastic Local Search with optimizations
T-statistic = 54.20359515232599
P-value = **2.1406307672415867e-21**

Mirela Chiţaniuc, MSAI2
Gâdioi Alexandra, MOC2

Stochastic Local Search **&** Stochastic Local Search with probabilities
T-statistic = 65.10326138309311
P-value = **8.0378874220305201e-23**

Stochastic Local Search with optimizations **&** Stochastic Local Search with probabilities
T-statistic = 30.360745850378944
P-value = **6.4906683332337147e-17**