FIT4012: Evolutionary Simulation and Synthesis
Programming Exercise 1

Name	**Questions**	**Due Date**	**Marks**
Exercise 1	10	16 August 2013	40

1 Exercises

1. Find out when hominids are first thought to have appeared, and estimate how many generations it has taken for you to evolve. Make sure you show your working and cite any references used in calculating your answer. [4 marks]

2. Code a generational genetic algorithm that evolves a solution to the OneMax problem. The OneMax function is a binary function for a string $\bar{x}$ of length L, maximised with a global optimum $= \bar{1}$, so $f(\bar{x}) = \sum_{i=1}^{L} x_i$. Use the following parameters for the GA when developing your solution:

   - Representation: binary strings of length $L = 25$
   - Initialisation: random
   - Parent selection: fitness proportionate, implemented via roulette wheel or SUS
   - Recombination: one-point crossover with probability $p_c = 0.7$
   - Mutation: bit-flip with probability $p_m = 1/L$
   - Replacement: strict generational (no elitism)
   - Population size: 100
   - Termination condition: 100 generations or optimum found (whichever comes first)

   After every generation find the best, worst, and mean fitness in the population and plot these on a graph with generation number as the x-axis. Now do ten runs and find the mean and standard deviation of the time taken to find the optimum. [8 marks]

3. Repeat the above exercise for a bigger problem, e.g., $L = 75$. How do your results change? [2 marks]

4. Repeat the above 2 exercises using tournament selection with $k = 2$. What differences do you see? [4 marks]

5. Calculate the probability that a binary chromosome of length L will not be changed by applying bit-flip mutation with $p_m = 1/L$.[1 mark]

6. A generational GA has a population size of 100, uses fitness proportionate selection without elitism, and after t generations has a mean population fitness of 76.0. There is one copy of the current best member, which has a fitness of 157.0.

   (a) What is the expectation for the number of copies of the best individual present in the mating pool?

   (b) What is the probability that there will be *no* copies in that mating pool, using roulette wheel selection?

   (c) What is the probability if the implementation uses SUS?

   [4 marks]

7. In an Evolutionary Strategies system, mutation transforms a chromosome $\langle x_1, \ldots, x_n, \sigma_1, \ldots, \sigma_n \rangle$ into $\langle x_1^{'}, \ldots, x_n^{'}, \sigma_1^{'}, \ldots, \sigma_n^{'} \rangle$, where:

$$\sigma_i^{'} = \sigma_i \cdot (1 + \alpha \cdot N(0,1)),$$
$$x_i^{'} = x_i + \sigma_i^{'} \cdot N_i(0,1).$$

   Why is it better that the mutation parameter (σ) is changed before the object variable (x)? [2 marks]

8. Write a program that implements an Evolutionary Strategy for the Ackley function[1] with $n = 30$, suing the following set-up:

   - comma selection strategy, (μ, λ), where $\mu = 30$ and $\lambda = 200$
   - uncorrelated mutations (i.e. no α values in the chromosome)
   - separate mutation step size for each dimension, so a chromosome will be a vector of 60 real-valued numbers (30 for $\bar{x}$ and 30 for $\bar{\sigma}$)
   - termination condition: 200,000 function evaluations or optimum found
   - initialisation: uniform random in the range $-30.0 <= x_i <= 30.0$

[1] see end of this document for a definition

Make 100 independent runs (remember that independent includes correctly seeding any pseudo-random number generators), store the best value of each run and calculate the mean and standard deviation of the values. Present your results in a table. [8 marks]

9. Repeat the experiment above with the plus selection strategy and compare the results. [3 marks]

10. A researcher has developed a new evolutionary algorithm to assist in an engineering optimisation problem. The researcher wants to compare this algorithm (B) against an existing one (A). Each algorithm was run multiple times and the following results were obtained (higher fitness values are better):

Algorithm	No. of experiments	Fitness value for each run
A	10	34, 37, 44, 31, 41, 42, 38, 40, 42, 38
B	10	39, 40, 34, 45, 44, 38, 42, 39, 47, 41

Can the researcher legitimately claim algorithm B is better than algorithm A based on these results? Use statistical tests to support your answer. Show any working. [4 marks]

2 Submission

Your answers to the above questions should be placed in a single pdf file and emailed to the lecturer. You submission is due by 5pm on the due date (email: `Jon.McCormack @ monash.edu`). Your should also send an accessible link to the source code you have developed (in a tar file), or alternatively you can send the source code in a compressed tar or zip file attachment in your submission email. Code should run under any Unix-based operating system and should include a `Makefile` or build script to build the program. Ideally, you should be able to change key parameters via the command line (e.g. L, selection method, p_c, p_m, population size, termination generation), hence your answers to questions 2-4 and 8-9 should just require different runs of the same program for each group of questions. Include the results of running your experiments (plots, stats, etc.) in the pdf file of your submission.

Late submissions will attract a penalty as described in the unit guide for this unit. You should receive your mark with feedback within two weeks of submission.

Ackley Function

The Ackley function is highly multimodal, with a large number of local minima fairly evenly spread over the search space, but with one minimum at the origin $\bar{x} = \bar{0}$ with the value 0.

$$f(\bar{x}) = -20 \cdot \exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + e$$

Here is a plot of the function for $n = 2$: