

CS341 SYSTEMS BIOLOGY I

PROJECT 4 (DUE TUES. NOV. 21, 2017)

Your first goal is to analyze the performance of several genetic algorithms in the context of parameter-estimation. Your second goal is to analyze the results, i.e. what do we learn about the model? In this project, we are revisiting Goldbeter's 5-state fly clock model, whose parameters were originally chosen "by hand." You will use a cost function which ensures the oscillations in constant darkness have a period of 23.6 hours and amplitudes of at least 0.1.

In class, we talked about the difference between a genetic algorithm (GA) and an evolutionary strategy (ES). Both use selection, crossover, and mutation operators. A classic GA uses bit strings for each genome, whereas an ES uses floating point numbers. Another major difference between a GA and an ES is in the implementation of the selection operators. In a GA, every individual from generation i is eligible to be a parent of individuals in generation $i + 1$. When each individual in generation $i + 1$ is created, a parent is chosen according to its fitness. In an ES, certain individuals of generation i are elected to be the "parents". When each individual in generation $i + 1$ is created, a parent is chosen randomly (irrespective of fitness) from this set.

For this project, we are following the lead of MATLAB and implementing a GA that uses floating point genomes. Thus, our code can accommodate both GA and ES algorithms. Table 1 outlines the options available for the CS341 genetic algorithm function. There are separate columns for options that apply to GAs and those that apply to ESs.

Explanations of the selection operators may be found in: Blickle, T. and Thiele, L. 1995. A comparison of selection schemes used in genetic algorithms. Tech. Rep. TIK-Report 11, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, May.

The new row in the table indicates the number of individuals that should be considered "elite." An "elite" individual in generation i is automatically made a member of generation $i + 1$. For GA's we can choose any number of individuals to be elite. It is a good idea to use a small number. This way we don't lose our best solution without decreasing diversity too much. ESs do not use this flavor of "elitism."

TABLE 1. Algorithms implemented

	Genetic Algorithm	Evolutionary Strategy
# Parents (μ) vs. # Children (λ)	$\mu = \lambda$	$\mu < \lambda$
Selection Operator	tournament proportional ranking	indirect selection by saying only a subset of individuals are qualified to be parents
Crossover Operator	uniform	uniform
Mutation Operator	Gaussian	Gaussian
Elite Count E	$0 \leq E \leq \lambda$	$E = 0$

1. ALGORITHM PERFORMANCE

- (1) Run our algorithm on the fly model at least 5 times for each “flavor” for at least 5 generations, i.e. analyze performance for
 - (a) a GA with tournament selection, an elite count of 1, 40 parents, 40 children, and a mutation fraction of 0.05
 - (b) a GA with linear ranking selection, an elite count of 1, 40 parents, 40 children, and a mutation fraction of 0.05
 - (c) an ES with an elite count of 0, 8 parents, 40 children, and a mutation fraction of 0.05
- (2) Report your results in a concise, but informative manner. You will want to identify trends. To do this, quantify the performance of the algorithm. For example, determine the mean population cost from generation to generation.

2. IMPLICATION FOR MODEL

Using some of your output from above, analyze the effects of different parameters on the model’s performance. Below are suggestions to guide your analysis. Follow at least one of them.

- (1) How different are the parameter sets? Do wildly different parameter sets map to wildly different outputs? Are their patterns?
- (2) You may see different relationships between amplitudes of clock components. Do you see any patterns in this behavior? If not, maybe you can conclude that the relative amplitude is not a property of the model’s parameters, but that it is a property of its structure. Make a detailed case for your conclusion.
- (3) Choose several parameter sets and determine whether or not the non-intuitive behavior mentioned in Goldbeter’s paper occurs. In his paper, he showed that increasing the rate of Per protein degradation (v_d) increases the total PER in the system. Is this true for multiple parameter sets?

3. EXTENSIONS

- Implement proportional selection and include analysis of the GA with proportional selection it in your write-up.
- Include additional analysis with different values of the algorithm's parameters. What happens if the number of children per generation is different? What about increasing or decreasing the mutation fraction?
- Include simulated annealing in your set of optimization algorithms.
- Implement a simulated annealing optimization algorithm.
- Make a hybrid algorithm that uses both a GA (or simulated annealing) and a deterministic method (e.g. a hill-descending method). The idea is that the stochastic algorithms helps you find the right region of parameter space, and that in that region, the cost function will be smooth (and maybe even monotonic). Once we are in a region that is smooth and monotonic, we can use a deterministic optimization method to refine our parameter set to find the local optimum.
- Also find parameters for the Gonze/Goodwin oscillator.