

# Evolutionary Dynamics

## Exercises 2

Prof. Dr. Niko Beerenwinkel  
Johannes Gawron  
Norio Zimmermann  
Michael Schneider

3. October 2024

Exercises marked with a "☐" are programming exercises. These can be solved in a programming language of your choice. Please make sure to hand in your code along with your answers to these exercises.

### Problem 1: Equilibrium points of linear and homogeneous equations (Covered in the tutorial)

Consider the following systems of linear and homogeneous equations:

- (a) Solve the system:

$$\begin{bmatrix} 1 & -4 \\ -2 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

How many equilibrium points exist? Determine the equilibrium points.

- (b) Solve the system:

$$\begin{bmatrix} 1 & -4 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

How many equilibrium points exist? Determine the equilibrium points.

### Problem 2: Sequence space and Hamming distance

Consider a binary alphabet  $\mathcal{A} = \{0, 1\}$  and a DNA alphabet with  $\mathcal{A} = \{A, T, C, G\}$ . We are studying sequences  $S \in \mathcal{A}^L$  of length  $L$ .

- (a) Assume sequences are random with a uniform distribution. What is the average Hamming distance between two random binary sequences? What is the expected Hamming distance for two random DNA sequences? **(1 point)**
- (b) Instead of a uniform distribution over the alphabet, we assume 0 has a 2 times higher probability of appearing at a position in the sequence than 1. Furthermore, the probabilities  $p(A) = p(T)$  are twice the probabilities  $p(C) = p(G)$ . What is now the average Hamming distance between two random binary sequences? And the expected average Hamming distance for two random DNA sequences? **(2 points)**
- (c) Given a binary sequence of length  $L$ , how many sequences exist at a Hamming distance 2 from it? How many at distance  $K$  with  $K \leq L$ ? Repeat the calculation for DNA sequences. **(1 point)**

### Problem 3: Quasispecies

Consider the quasispecies equation with two genotypes 0, 1 (i.e., binary sequences of length 1). Let the fitness of genotype 0 be  $f_0 > 1$ , and the fitness of genotype 1 be  $f_1 = 1$ . Moreover, genotypes are replicated error-free with probability  $q$ . The quasispecies equation is defined as

$$\dot{x}_i = \sum_{j=1}^n \underbrace{x_j}_{\text{frequency of } j} \underbrace{f_j}_{\text{fitness of } j} \underbrace{q_{ji}}_{\text{mutation from } i \text{ to } j} - \underbrace{\phi}_{\text{average fitness}} x_i, \quad i = 0, \dots, n$$

In vector notation, this can be written as

$$\dot{x} = xW - \phi x$$

where  $W = (w_{ij}) = (f_j q_{ji})$  is the mutation-selection matrix and  $x = (x_0, \dots, x_n)$ .

- (a) Write down the mutation-selection matrix  $W$  and find its eigenvalues. **(1 point)**
- (b) What are the equilibrium points of the population? To which eigenvalue corresponds the non-trivial equilibrium point? **(1 point)**  
*Hint: Perron-Frobenius theorem.*
- (c) Examine the dynamics of the quasispecies equation and confirm the results obtained in (b). Assume that  $q = 0.7$  and  $f_0 = 1.5$ , with the initial condition  $(0.9, 0.1)$ . Modify your simulation to track the average fitness of the population over time. How does the average fitness evolve as the population approaches equilibrium? ☐ **(1 point)**
- (d) What is the equilibrium point for  $f_0 = f_1 = 1$ ? **(1 point)**
- (e) In the lecture you heard about a concept called the error threshold. Assume for this subtask a wildtype genotype with fitness  $f_{wt} > 1$  and all other genotypes have fitness  $f_m = 1$ . For a large sequence length, you can assume, that once a sequence is mutated, back-mutations are negligibly unlikely. Can you derive a condition based on  $q$  and  $f_{wt}$  for this error threshold? What happens when the threshold is crossed? **(2 points)**