

Exercise 05

The Boltzmann distribution

Deadline: Please hand in your protocol as a single **pdf document** by **Thursday, 16th June, 10.15 a.m.** to **saleksic@zedat.fu-berlin.de**. The protocol should contain numerical solutions, Python code, plots, and comments if necessary.

5.1 Lagrange multiplier (15 P)

Consider the surface

$$f(x, y) = e^{-(x^2+y^2)} \quad (1)$$

and the constraint

$$g(x, y) = x + y - 1 = 0 \quad (2)$$

What is the maximum of $f(x, y)$ along the constraint $g(x, y) = 0$? Solve the problem by using the method of Lagrange multipliers.

5.2 Probability theory of a peptide sequence (40 P)

Table 1 shows the relative frequency (in percent) of amino acids acting as building blocks of human proteins. You can find a file with this data in Files section. In a random experiment, the peptide sequences with N residues are drawn with probabilities given by this distribution. The outcome of the experiment is a sequence L e.g.

$$L = \{\text{AGMMTPLKRN}\}.$$

The composition C of a sequence denotes how often a given residue occurs in the sequence

$$C = \text{A}_1\text{G}_1\text{M}_2\text{T}_1\text{P}_1\text{L}_2\text{K}_1\text{R}_1\text{N}_1.$$

Swapping residues produces a new peptide sequence, but the amino acid composition remains the same like in the previously generated sequence. (e.g. $L' = \{\text{GMAMTPLKRN}\}$)

The peptide sequence can be compared with a thermodynamics system where the N residues of a sequence correspond to the N particles of a microstate. Different residues correspond to different energy levels, equal letters are particles at the same energy level. For example, in the sequence "AGMMTPLKRN" there are $N = 10$ residues (particles), but only $N_e = 9$ energy levels ("M" appears twice).

- (a) What is the most likely decamer sequence? (5 P)
- (b) What is the total number of possible decamer sequences? (5 P)
- (c) What is the total number of possible decamer sequences containing only aromatic residues? (5 P)
- (d) Implement a program which generates a sequence of a decamer according to the given distribution. (10 P)
The program has also to calculate the following quantities:
- (e) For an arbitrary sequence of a decamer, what is the number of possible sequences W with the same composition? (5 P)
- (f) According to the proposed distribution, what is the probability to draw a certain peptidic sequence L ? (5 P)
- (g) What is the probability to draw a certain composition C ? (5 P)

Residue	Frequency (%)	Residue	Frequency (%)	Residue	Frequency (%)
A	7.4	G	7.4	P	5.0
R	4.1	H	2.9	S	8.1
N	4.4	I	3.8	T	6.2
D	5.9	L	7.6	W	1.3
C	3.3	K	7.2	Y	3.3
E	5.8	M	1.8	V	6.8
Q	3.7	F	4.0		

Table 1: Relative frequency of amino acids in the human proteome.

5.3 Thermodynamics of a two-state folder (45 P)

Heat denatures proteins, i.e. above a certain temperature proteins lose their three-dimensional fold. This is one of the reasons why a high fever is dangerous. In this exercise, you will study the heat denaturation of proteins using a one-dimensional model potential energy function for a two-state folder

$$V(x) = -10^{4.05} \left(e^{-2 \cdot (0.7x-2)^2} + e^{-0.0003 \cdot (x-25)^2} \right) \quad (3)$$

A two-state folder is a protein which fold and unfolds via a single transition state. The folded state is represented by a deep and narrow minimum, the unfolded state is represented by a shallow and broad minimum. The Boltzmann distribution of this potential energy function is

$$p(x) = \frac{\exp(-\beta V(x))}{\int \exp(-\beta V(x)) dx} \quad (4)$$

(Note that the energy is defined in J/mol and consequently $\beta = \frac{1}{RT}$, where $R = 8.314 \text{ J}/(\text{K mol})$ is the ideal gas constant. In this way numerical instabilities can be avoided.)

- Plot the potential $V(x)$ and the Boltzmann distribution $p(x)$ at body temperature (310 K) from $x = 0$ to $x = 60$. (10 P)
- Determine the transition state (barrier). Hint: Use the function `argrextrema` of `scipy` (5 P)
- Calculate the probabilities p_{fold} and p_{unfold} of the folded and the unfolded state from $T = 270 \text{ K}$ to $T = 370 \text{ K}$ with a step of 10 K. Plot them as a function of temperature. (10 P)
- Explain the changes in the population. (5 P)

The equilibrium constant between folded and unfolded state can be expressed in terms of the state probabilities

$$K = \frac{p_{unfold}}{p_{fold}} \quad (5)$$

- Compute and plot the free energy and the entropy as a function of temperature assuming that the system is isolated and at constant pressure:

$$\Delta G = -RT \ln K \quad (6)$$

$$\Delta S = -\frac{\Delta H}{T} + \frac{\Delta G}{T} \quad (7)$$

(10 P)

- Determine the melting temperature of the protein. The melting temperature is the temperature at which $\Delta G = 0$. Hint: Use the function `polyfit()`. (5 P)

5.4 Files

<https://www.dropbox.com/s/kj6zuwmlb85v6k/probability.dat?dl=0>