Exercise 05

The Boltzmann distribution

Deadline: Please, send your solutions to **luca.donati@fu-berlin.de** by **Friday, 12 June, 10.15 a.m.**

5.1 Lagrange multiplier (3 P)

Consider the surface

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

and the constraint

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

(a) 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. (2 P)

(b) With MATLAB, plot the function $f(x, y)$ and the constraint $g(x, y) = 0$. Hint: Use the functions meshgrid(), contour3() and surf(). (1 P)

5.2 Probability theory of alphabet pasta (8 P)

Table 1 shows the relative frequency (in percent) of letters in the English language. You can download a file with this data from the website of the course. In a random experiment, N letters are drawn with probabilities given by this distribution. (Imagine sitting in front of a enormous bag of alphabet noodles and picking out noodles, one at a time with replacement). The outcome of the experiment is a sequence L of letters, e.g.

$$L = \{\text{p, r, o, b, a, b, i, l, i, t, y}\}\,.$$

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

$$C = \text{p}_1\text{r}_1\text{o}_1\text{b}_2\text{a}_1\text{i}_2\text{l}_1\text{t}_1\text{y}_1\,.$$

Swapping letters changes the sequence, but not the composition. (e.g. $L' = \{\text{p, r, o, b, i, b, a, l, i, t, y}\}$)

The alphabet system can be compared with a thermodynamics system where the N letters of a sequence correspond to the N particles of a microstate. Different letters correspond to different energy levels, equal letters are particles at the same energy level. For example, in the sequence "p, r, o, b, a, b, i, l, i, t, y" there are $N = 11$ letters (particles), but only $N_e = 9$ energy levels ("b" and "i" appear twice).

Implement a program which generates a sequence of N letters according to the given distribution. (4 P)
The program has also to calculate the following quantities:

(a) What is the most likely sequence of length N=5? (0.5 P)

(b) For an arbitrary sequence of letters, what is the number of possible anagrams W with the same composition? (1.5 P)

(c) According to the proposed distribution, what is the probability to draw a certain sequence of letters L? (0.5 P)

(d) What is the probability to draw a certain composition C? (0.5 P)

The quantity W is the number of ways one can realize the given thermodynamic microstate and can be used to compute the Boltzmann entropy:

$$S = k_B \log W$$

where we can assume $k_B = 1$.

Letter	Frequency (%)	Letter	Frequency (%)	Letter	Frequency (%)	Letter	Frequency (%)
a	8.167	h	6.094	o	7.507	v	0.978
b	1.492	i	6.966	p	1.929	w	2.360
c	2.782	j	0.153	q	0.095	x	0.150
d	4.253	k	0.772	r	5.987	y	1.974
e	12.702	l	4.025	s	6.327	z	0.074
f	2.228	m	2.406	t	9.056		
g	2.015	n	6.749	u	2.758		

Table 1: Relative frequency of letters in the English language.

(e) Compute the entropy S using the Stirling's formula:

$$\log N! \approx N \log N - N$$

(1 P)

5.3 Thermodynamics of a two-state folder (7 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

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

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

(a) Plot the potential $V(x)$ and the Boltzmann distribution $p(x)$ at body temperature (310 K) from $x = 0$ to $x = 60$. (1 P)

(b) Determine the transition state (barrier). Hint: Use the function diff() (1 P)

(c) Calculate the probabilities p_{fold} and p_{unfold} of the folded and the unfolded state from $T = 270$ K to $T = 370$ K with a step of 10 K. Plot them as a function of temperature. (2 P)

(d) Explain the changes in the population. (1 P)

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

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

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

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

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

(2 P)

(f) Determine the melting temperature of the protein. The melting temperature is the temperature at which $\Delta G = 0$. Hint: Use the function polyfit(). (1 P)