

Laboratory practice for chemical computations (2)

notes for practice

BSc in Chemistry, computational lab in 2nd semester

2 hours/week, 2 credits

Gergely Tóth

Loránd Eötvös University, Institute of Chemistry



PART I

<u>Solution of non-linear equations, search for minima and maxima</u>	2
<u>Statistical tests</u>	4
<u>Analysis of variance – one-way (single factor)</u>	7
<u>Matrices – basic manipulation</u>	8
<u>Solution of set of inhomogeneous linear equations</u>	9
<u>Parameter estimation</u>	11
Sample test for Part I	14

PART II

<u>Basics of Octave</u>	16
<u>Data storage – matrices – statistics using Octave</u>	18
<u>Non-linear equations, minima and maxima using Octave</u>	20
<u>Solution of a set of linear equations using Octave</u>	20
<u>Distribution of a function of a random variable</u>	21
<u>Error propagation (by Gauss)</u>	22
<u>Integration, polynomials, interpolation using Octave</u>	23
<u>Fourier-transformation, frequency and noise filtering</u>	24
<u>Modelling</u>	26
Sample test for Part II	28
References	29

version 2016.05.06

"Supported by the Higher Education Restructuring Fund allocated to ELTE"

PART I

Solution of non-linear equations, search for minima and maxima

Importance of numerical methods

Iterative methods

initial value(s)

convergent-divergent methods

stop criteria: $|x_{k+1}-x_k| < \epsilon$, vagy $|f(x_{k+1})-f(x_k)| < \epsilon$,

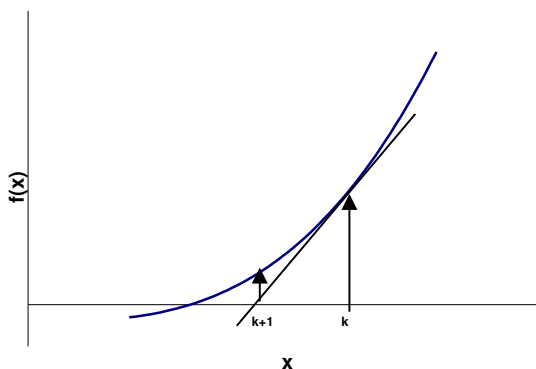
Numerical solution of nonlinear equations

Purpose: $y=h(x)$ function, for a given y_0 value search the corresponding x_0 , if x cannot be obtained explicitly as $x=g(y)$.

rearrange to $f(x)=h(x)-y_0=0$ (with EXCEL not necessary)

One of the basic methods:

Newton method



1 initial value, might diverge, iteration step:

$$x_{k+1} = x_k - f(x_k) / f'(x_k)$$

Another method: Bracketing and half of intervals (2 initial values, safe convergence)

x_a and x_f chosen as $f(x_a) * f(x_f) < 0$

$x_k = (x_a + x_f) / 2$, if $f(x_k) * f(x_a) < 0$, new $x_f = x_k$, otherwise $x_a = x_k$

repeating as $[x_a, x_f] < \epsilon$

Solution with EXCEL:

Solver (or Goal seeker)

Variable cell (x with initial value), target cell ($f(x)$) function pointing to the variable cell

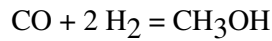
Set of minima/maxima/non linear solution Constraints

Task: Lennard-Jones interaction potential: minima and $y=0$ search

Exercises:

Calculation of chemical equilibrium [1]

Syntheses of methanol starts with 25 % CO, 55 % H₂ and 20 % inert gas mixture (in mole %). The equilibrium constant using the partial pressures of the reaction



is at 350 °C:

$$K_p = \frac{P_{\text{CH}_3\text{OH}}}{P_{\text{CO}} P_{\text{H}_2}^2} = 1.4 \cdot 10^{-14} \text{ Pa}^{-2} \quad (1)$$

Let x the equilibrium conversion, and start from 100 mol mixture. Writing the initial and equilibrium quantities of the components:

Component	Initial (mol)	Equilibrium (mol)
CO	25	25-25x
H ₂	55	55-50x
CH ₃ OH	0	25x
inert gas	20	20
sum	100	100-50x

The partial pressures at the equilibrium:

$$P_{\text{CH}_3\text{OH}} = \frac{25x}{100 - 50x} P$$

$$P_{\text{CO}} = \frac{25 - 25x}{100 - 50x} P$$

$$P_{\text{H}_2} = \frac{55 - 50x}{100 - 50x} P$$

,where P denotes the total pressure. Determine the equilibrium conversion using EXCEL according to equation (1)!

Use the following total pressures:

- 1) $P = 3.0 \cdot 10^7 \text{ Pa}$
- 2) $P = 3.5 \cdot 10^7 \text{ Pa}$
- 3) $P = 2.0 \cdot 10^7 \text{ Pa}$
- 4) $P = 2.5 \cdot 10^7 \text{ Pa}$

Simplify the calculation with the 10^7 magnitude of the pressure to avoid the optimization of numbers around 10^{-14} .

Determination of dissociation degree [1]

The dissociation of nitrogen-dioxide is due to the equation $2 \text{NO}_2 \rightleftharpoons 2 \text{NO} + \text{O}_2$.

If the reaction is performed at constant volume, the equilibrium partial pressures of the components can be described using the initial pressure of nitrogen-dioxide ($p_{\text{NO}_2}^0$) and the degree of dissociation (α):

$$p_{\text{NO}_2} = (1 - \alpha) p_{\text{NO}_2}^0$$

$$p_{\text{NO}} = \alpha p_{\text{NO}_2}^0$$

$$p_{\text{O}_2} = 0.5\alpha p_{\text{NO}_2}^0$$

Combining these equations into the equation of the equilibrium constant using partial pressures:

$$K_a = K_p \left[\frac{1}{p^\ominus} \right]^{\Delta v} = \frac{0.5\alpha p_{\text{NO}_2}^0 (\alpha p_{\text{NO}_2}^0)^2}{(1 - \alpha)^2 (p_{\text{NO}_2}^0)^2} \frac{1}{p^\ominus} \quad (1)$$

where $p^\ominus = 100 \text{ kPa}$ is the standard pressure, Δv is the net change of stoichiometric numbers (here equals to 1). Equation (1) can be solved for α . Calculate the degree of dissociation!! Start with 150 kPa pure nitrogen-dioxide at 700 K. The equilibrium constant is: $K_a = 0,18$

Statistical tests

Decision on the validity of a statement (hypothesis), if we have limited number of data. Data are random variables (“probability” variable in Hungarian), decision will be true also with a probability. Statement = hypothesis (H_0), opposite statement = counter hypothesis (H_a): Which one is true? General method: The hypothesis can be generally described by a distribution. Where is the position of the value calculated from our data in this distribution?

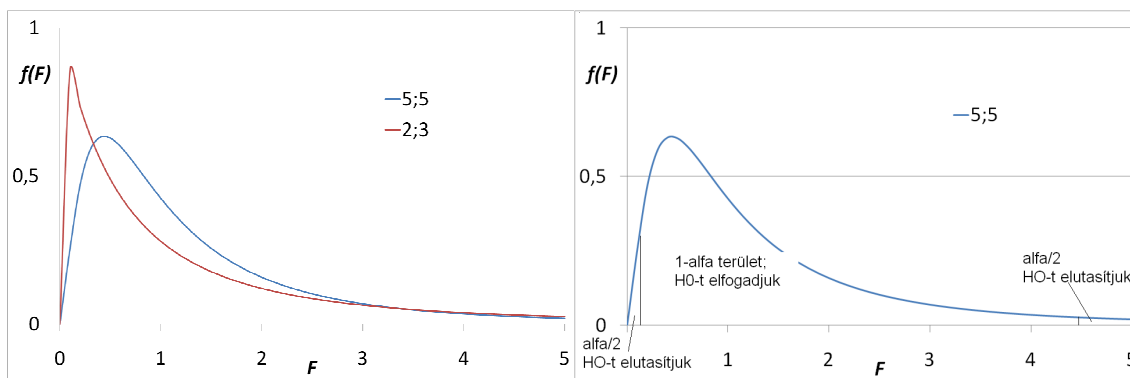
F-test – comparison of two variances

$H_0: \sigma_1 = \sigma_2$ $H_a: \sigma_1 \neq \sigma_2$,

Number of measurements: N_1 and N_2 .

For example: two instruments, two students, two methods: their precision?

Background: F -distribution ratio of two variances obtained for two samples of a standard normal distribution. Two degrees of freedom: N_1-1 and N_2-1 . Estimated standard deviations: s_1 and s_2 .



F-distribution for two pair of degrees of freedom

limits of the two-sided test

Theory: calculation of $\xi_F = \frac{s_1^2}{s_2^2}$, comparison to N_1-1, N_2-1 F -distribution, where is the actually calculated value.

Supposing a significance α (e.g., 0.05).

EXCEL failure: F.DISTRIBUTION and INVERSE.F calculate $1-F(\xi)$!!

Solution with EXCEL:

F.TEST(data1, data2) gives p -value of H_0 :

if $\alpha \leq F.TEST$, H_0 -t accepted

if $F.TEST < \alpha$, H_0 -t is rejected (H_a is accepted).

One sample t-test – comparison of the expected value of the sample and a theoretical mean

E – expected value of the population

\hat{y} – sample mean = estimated expected value

E_0 – theoretical mean (e.g. official data...)

H₀: E=E₀

H_a: E≠E₀

Background: $\xi = \frac{\hat{y} - E}{s / \sqrt{N}}$ follows $N-1$ degrees of freedom t-distribution

theoretical solution: calculation of $\xi = \frac{\hat{y} - E_0}{s / \sqrt{N}}$ and comparison to $F(\xi)$ at α .

Two sided: if $\alpha/2 \leq F(\xi) \leq (1-\alpha/2)$, H₀ is accepted

if $F(\xi) < \alpha/2$ or $(1-\alpha/2) < F(\xi)$, H₀ is rejected (H_a is accepted)

One sided, e.g. only larger is “wrong”: if $F(\xi) \leq (1-\alpha)$, H₀-t is accepted

if $(1-\alpha) < F(\xi)$, H₀-t is rejected (H_a is accepted)

EXCEL solution for two sided:

if $\alpha \leq \text{T.DISTRIBUTION}(\text{abs}(\xi); N-1; 2)$, H₀ is accepted

if $\text{T.DISTRIBUTION}(\text{abs}(\xi); N-1; 2) < \alpha$, H₀ is rejected (H_a is accepted)

$30 < N$, use normal distribution:

if $\alpha/2 \leq \text{Z.TEST}(\text{data}, E_0) \leq (1-\alpha/2)$, H₀ is accepted

otherwise, H₀ is rejected (H_a is accepted)

Two-sided t-test – comparison of two sample means

E_1, E_2 – expected value of two populations

\hat{y}_1, \hat{y}_2 – sample means = estimated expected values

H₀: E₁=E₂

H_a: E₁≠E₂

general scheme: $\xi = (\text{estimated parameter} - \text{theoretical parameter}) / (\text{estimated standard deviation of the parameter})$

estimated parameter: $\hat{y}_1 - \hat{y}_2$, theoretical parameter: $E_1 - E_2$

standard deviation: $s = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$, if $30 < N_1, N_2$

maybe used „pooled” variance, if $N_1, N_2 \leq 30$

$$s_p^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \quad s = s_p \left(\frac{1}{\sqrt{N_1}} + \frac{1}{\sqrt{N_2}} \right)$$

Calculation, acceptance and rejection as at one-sided t-test or

if $\alpha \leq \text{T.TEST}(\text{data1}; \text{data2}; 2; 2 \text{ or } 3)$, H₀ is accepted

if $\text{T.TEST}(\text{data1}; \text{data2}; 2; 2 \text{ or } 3) < \alpha$, H₀ is rejected (H_a is accepted)

χ^2 -test – goodness-of-fit test

χ^2 -distribution: $\xi_1, \xi_2, \xi_3, \xi_4, \dots$ standard normal random variables, then $\xi = \xi_1^2 + \xi_2^2 + \dots + \xi_N^2$ follows N degrees of freedom χ^2 -distribution. If centered data (shift to zero mean), degrees of freedom = $N-1$.

$E(\xi)$ =degrees of freedom

$\sigma^2(\xi)$ = 2*degrees of freedom

Connection to kinetic energy distribution of particles: v_x, v_y, v_z velocities normally distributed

$$v^2 = v_x^2 + v_y^2 + v_z^2 \rightarrow E_{\text{kinetic}} = mv^2/2 \rightarrow \text{follows Maxwell-Boltzmann distribution} = \chi^2\text{-distribution } N=3$$

χ^2 -test is good for: e.g. Check the difference between two curves, is it only statistical fluctuation or more?

χ^2 -test, to test if something follows a theoretical distribution:

$$\mathbf{H}_0: p_1^{\text{theoretical}} = p_1^{\text{experimental}}, p_2^{\text{theoretical}} = p_2^{\text{experimental}}, \dots, p_N^{\text{theoretical}} = p_N^{\text{experimental}}$$

\mathbf{H}_a : if there is at least one inequality in \mathbf{H}_0 -ban

Theoretical solution:

$$\xi = \sum_{i=1}^N \frac{(y_i - e_i)^2}{e_i}, \text{ where } y_i = \text{the number of the } i\text{-th type realization, } e_i = kp_i, k = \sum_{i=1}^N y_i$$

if $\xi \leq \chi^2(\alpha, N-1)$, \mathbf{H}_0 is accepted

if $\chi^2(\alpha, N-1) < \xi$, \mathbf{H}_0 is rejected (\mathbf{H}_a is accepted)

The scheme is good to compare any problem, if it can be converted to two discrete functions: $g(x) \rightarrow g(x_i)$ és $f(x) \rightarrow f(x_i)$

EXCEL solution:

if $\alpha \leq \text{CHISQ.TEST}(\text{data1}_{\text{real}}; \text{data2}_{\text{expected}})$, \mathbf{H}_0 is accepted

if $\text{CHISQ.TEST}(\text{data1}_{\text{real}}; \text{data2}_{\text{expected}}) < \alpha$, \mathbf{H}_0 is rejected (\mathbf{H}_a is accepted)

Exercises

1) Set statements on the expected value and the standard deviation of these data and check them in statistical tests. In the case of one sample t-test use $E_0 = 1.6$ and $E_0 = 1.8$.

2.1	1.6
2.2	2.2
2.3	1.4
1.9	2.2
2.2	1.8
3.2	1.3

2) The following data were measured for the radiation of construction materials in the schools of Ajka city (the dose are given without dimensions). Is it possible to model the data with Poisson-distribution?

dos e	frequenc y
0	4
1	9
2	13
3	12
4	2
5	0

3) The conductivity of AgNO₃ solution was measured by three students at $T=298\text{ K}$ and $c=0.05\text{ mol/dm}^3$. Evaluate statistically the data (expected values, standard deviations, confidence intervals, t- and F-tests, one sample t-test with $E_0=115,2\text{ cm}^2\Omega^{-1}\text{dm}^{-3}$)! The inverse-t value can be obtained using the INVERZ.T($\alpha,N-1$) EXCEL function.

conductivity ($\text{cm}^2\Omega^{-1}\text{dm}^{-3}$)		
1.studen t	2.studen t	3.studen t
115.28	114.42	115.36
115.52	115.28	116.16
114.76	115.58	115.68
115.80	114.90	115.50

Analysis of variance – one-way (single factor)

Purpose: The sample can be divided into classes, e.g. different labs, male/female. Is there any significant difference among the classes?

Background:

$$SS_T = \sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 = SS_{\text{withingroups}} + SS_{\text{betweengroups}} = \sum_i \sum_j (x_{ij} - \bar{x}_{.j})^2 + \sum_j n(\bar{x}_{.j} - \bar{x}_{..})^2,$$

where "." denotes averaging on that index. The total variance is divided into two: within one group and between group variances. The variances: $SS_T/(nq-1)$, $SS_{\text{within}}/q(n-1)$, $SS_{\text{between}}/(q-1)$, where n denotes the number of data within a group, and q is the number of the groups.

Hypothesis on the expected value of the groups:

$$\mathbf{H}_0: \mu_1 = \mu_2 = \dots = \mu_q$$

\mathbf{H}_a : at least one inequality in H_0

Accepted, if smaller than the corresponding critical value:

$$F = \frac{SS_{\text{betweengroups}} / (q-1)}{SS_{\text{withingroup}} / q(n-1)} < F_{q-1, q(n-1)}^{\alpha=0,05}$$

EXCEL solution: Data analysis/Single-factor ANOVA

Exercises:

Aflatoxin content of milk measured in different labs

The milk-farm ordered aflatoxin analyzation from 7 different labs. Each lab performed 5 measurements. Is there any significant difference in the choice of the labs at 0.05 significance level?

a	b	c	d	e	f	g
1.6	4.6	1.2	1.5	6	6.2	3.3
2.9	2.8	1.9	2.7	3.9	3.8	3.8
3.5	3	2.9	3.4	4.3	5.5	5.5
1.8	4.5	1.1	2	5.8	4.2	4.9
2.2	3.1	2.9	3.4	4	5.3	4.5

Statistical control of instruments:

Five pH instrument of the student lab were checked before the semester. 7-7 aliquot samples of the standard solution were measured for each instrument Evaluate statistically the results (mean, variance, standard deviation, ANOVA, t- and F-tests...). Which instruments should be recalibrated (shift on the scale) and which ones should be mended (large uncertainty during the measurement)?

Data:

A	B	C	D	E
7.137	7.120	7.087	7.317	7.091
7.113	7.151	7.110	7.286	7.028
7.103	7.104	7.098	7.315	7.080
7.127	7.147	7.075	7.283	7.117
7.100	7.144	7.092	7.298	7.091
7.095	7.108	7.117	7.285	7.162
7.104	7.133	7.103	7.272	7.114

Matrices – basic manipulation

Matrices: position and value

adding $C=A+B$, $\leftrightarrow c_{ik}=a_{ik}+b_{ik}$

times a constants: $\text{const } A \leftrightarrow \text{const } a_{ik}$

multiplication (row - column): $C_{ln}=A_{lm}B_{mn} \leftrightarrow c_{ij} = \sum_{k=1}^m a_{ik} b_{kj}$

identity matrices: $I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$

inverse: $AA^{-1}=A^{-1}A=I$

determinant of $n \times n$ matrices:

$\det A = \sum_p (-1)^l a_{1k_1} a_{2k_2} \dots a_{nk_n}$, sum of all possible product taking one element of every row being in

different columns. Multiplication with +1/-1, due to the number of permutations.

$\det A = 0$, if all elements = 0 in one row,

if a row can be got by linear combination of other rows

valid for columns as well

matrices in EXCEL:

MDETERM, MMULT, MINVERSE

(input of functions with output of more cells: see at the help of the FREQUENCY function).

Solution of set of inhomogeneous linear equations

mathematical background:

general description:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ &\dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_{n1} \end{aligned} \leftrightarrow \underline{A}\underline{x}=\underline{b}_.$$

where $\underline{A} \in \mathbb{R}^{n \times n}$, $\underline{x} \in \mathbb{R}^n$, $\underline{b} \in \mathbb{R}^n$, therefore

$$\underline{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} \text{ and } \underline{b} = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix}.$$

Inhomogeneous, if exist at least one $b_i \neq 0$. If all $b_i = 0$, homogeneous (not treated here) Unique solution if $\det A \neq 0$.

If $\det A = 0$, singular matrix. (similar to rank or independence of vectors)

Task: purchase at a market (3 sort of fruits, 3 customers; 3 sort of fruits + box(=intercept) 4 customers,) The amount of fruits and the total bills are known, the unknowns are the prices of the fruits.

EXCEL: LINEST function

Solution of set of inhomogeneous linear equations – overdetermined case

Background: More independent rows (n) than columns (m). Change in roles: a_{ij} is the value of the j -th variable in the i -th measurement, denoted earlier with x_j equals to the j -th slope or the intercept. Regression in multi dimensions: purpose the determination of slopes and intercept.

Merit function: $\sum_{i=1}^n (b_i^{\text{mért}} - b_i^{\text{számolt}})^2$ - find the minimum.

Solution: $\underline{x} = (\underline{A}^T \underline{A})^{-1} \underline{A}^T \underline{b}$

With EXCEL: LINEST function, or Data/Dataanalysis/Regression (activate!) Goodness of fit = R^2 close to 1.

Task: purchase at a market (3 sort of fruits, 4-5 customers; 3 sort of fruits + box(=intercept) 5-6 customers)

Exercises:

Determination of concentrations from spectroscopic data

A solution contains different organic compounds. The concentrations of compounds A, B, C, D and E can be

determined by measuring at 5 different wavelengths according to the equation $\lg\left(\frac{I_f}{I_0}\right) = -\sum_i \epsilon_i c_i l$. The solvent does

not absorb at the selected wavelengths. The molar absorption coefficients:

	$\epsilon_A / (dm^3 * mol^{-1} * cm^{-1})$	$\epsilon_B / (dm^3 * mol^{-1} * cm^{-1})$	$\epsilon_C / (dm^3 * mol^{-1} * cm^{-1})$	$\epsilon_D / (dm^3 * mol^{-1} * cm^{-1})$	$\epsilon_E / (dm^3 * mol^{-1} * cm^{-1})$
$\lambda = 300 \text{ nm}$	114.3	10.1	2.0	26.7	56.3
$\lambda = 400 \text{ nm}$	3.0	89.1	4.2	22.1	19.8

$\lambda=500$ nm	10.0	9.7	160.1	30.1	2.0
$\lambda=600$ nm	64.5	5.6	20.1	230.4	11.4
$\lambda=700$ nm	19.4	4.5	8.7	10.8	132.3

The measured absorbances:

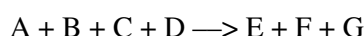
	$-\lg(I/I_0)$
$\lambda=300$ nm	0.269
$\lambda=400$ nm	0.197
$\lambda=500$ nm	0.331
$\lambda=600$ nm	0.297
$\lambda=700$ nm	0.231

The cuvette is 1 cm wide.

(results: 0.00138; 0.00163; 0.00175; 0.00065; 0.00132)

Determination of rate constant (k)

The task is the determination of the rate constant of the reaction:



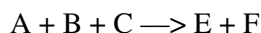
The reaction path contains many steps, therefore, the order of the reaction is not known: the determination of the exponents is also our task. The reaction is followed by taking samples from the solution at different time steps and the concentration of the components are determined for each sample and component. The concentration change is calculated for the time grids and its averages are assigned as reaction speed. The results are summarized here:

reaction speed	$c_A,$ (mol/dm ³)	$c_B,$ (mol/dm ³)	$c_C,$ (mol/dm ³)	$c_D,$ (mol/dm ³)
7.74E-04	2.3	0.9	1.9	2.2
5.90E-04	1.7	0.8	1.7	2.1
3.87E-04	1.4	0.7	1.4	1.8
2.60E-04	0.7	0.6	1.3	1.3
1.51E-04	0.3	0.5	1.0	1.2

Write the set of linear equations on the logarithmic form of equation $v=k[A]^a[B]^b[C]^c[D]^d!$ Determine $k!$ (result: 2.84e-4)

Determination of rate constant (k) – regression case

The task is the determination of the rate constant of the reaction:



The reaction path contains many steps, therefore the order of the reaction is not known: the determination of the exponents is also our task. The reaction is followed by taking samples from the solution at different time steps and the concentration of the components are determined for each sample and component. The concentration change is calculated for the time grids and its averages are assigned as reaction speed. The results are summarized in the table:

reaction speed	$c_A,$ (mol/dm ³)	$c_B,$ (mol/dm ³)	$c_C,$ (mol/dm ³)
7.74E-04	0.27	0.15	0.36
5.90E-04	0.24	0.13	0.32

3.87E-04	0.20	0.10	0.28
2.82E-04	0.17	0.07	0.24
2.40E-04	0.16	0.07	0.23
2.20E-04	0.15	0.06	0.22
1.82E-04	0.14	0.06	0.20
1.51E-04	0.12	0.05	0.19

Write the set of linear equations on the logarithmic form of equation $v=k[A]^a[B]^b[C]^c$. Overdetermined case: more equations than unknowns. Determine $k!$ (result: 0.0107)

Linear regression, confidence interval for the parameters

y depends linearly on the concentration of A, B and C compounds. Determine the linear coefficients (slopes - m_j -s) of the three variables. The intercept can be different from zero (b constant) Give 95% two-sided confidence intervals for the parameters using the equation $m_j \pm t_{(n-p, \alpha)} * s_j$; The standard deviations of the parameters are the square roots of the

diagonal elements of $S_r^2 (X^T X)^{-1}$, where $S_r^2 = \frac{\sum_{i=1}^n (y_i^{mért} - y_i^{illesztett})^2}{n - p}$ is the residual variance, n is the number of the

measurements and p is the number of the parameters. X , the so called predictor matrix of the independent variables is extended with a column vector of 1 for the intercept parameter. The inverse t-value can be approximated with the Gaussian 1,96 factor of 95% two-sided case. The measured data:

C_A	C_B	C_C	y
0.050	0.030	0.020	0.174
0.020	0.014	0.023	0.052
0.012	0.014	0.013	0.050
0.012	0.015	0.012	0.072
0.010	0.034	0.034	0.097
0.005	0.005	0.005	0.042
0.067	0.002	0.001	0.150

Parameter estimation

Parameter estimation of randomly generated functions biased with error

A-B columns: Fill with 10-10 random numbers (e.g. integers)

C column: $C1=1.2*\cos(A1)-2/B1$

D column: Random numbers from standard normal distribution (Analysis Toolpack)

E column: $E1=C1+0.05*D1$ (adding random error to the data)

F1; F2 starting values for $p1$ and $p2$

G column: equation of C column, but instead of 1.2 referenced to $\$F\1 and instead of 2 referenced to $\$F\2

$F3=\text{SUMXMY2}(C1:C10;G1:G10)$

Search for the minimum of F3 using Solver, 'variables' = $p1$ and $p2$

repeat for columns E and G

Determination of the rate constant of a reaction [1]

The differential equation of the $A \xrightarrow{k_1} B \xrightarrow{k_2} C$ reaction can be analytically solved with $A(t=0)=1\text{ mol/m}^3$, $B(t=0)=0\text{ mol/m}^3$, and $C(t=0)=0\text{ mol/m}^3$ are the initial values.

$$C(t) = 1 - \frac{k_2}{k_2 - k_1} e^{-k_1 t} + \frac{k_1}{k_2 - k_1} e^{-k_2 t}$$

Determine the k_1 and k_2 rate constants by parameter estimation. Proceed the fit on the exact simulated and on the biased data, as well. Starting values of the parameters are $k_1=4$ and $k_2=5$.

$t(\text{s})$	$C(t)$ (mol/m ³) (exact)	$C(t)$ (mol/m ³) (biased)
0.5	0.2018	0.2042
1	0.4731	0.4025
2	0.7982	0.8534
3	0.9254	0.9095
4	0.9725	0.9733
5	0.9898	0.9072
6	0.9962	0.9740
7	0.9986	0.9877
8	0.9995	1.0233
9	0.9999	1.0370

Estimation of the van der Waals constants of ammonia

The van der Waals equation of state has the form:

$$\left(p + \frac{an^2}{V^2} \right) (V - nb) = nRT \quad (1).$$

Rearranging to get the pressure explicitly:

$$p = \frac{nRT}{V - nb} - \frac{an^2}{V^2} \quad (2)$$

The constants can be related to the critical values of ammonia:

$$p_{kr} = \frac{a}{27b^2} \quad (3) \text{ és } T_{kr} = \frac{8a}{27bR} \quad (4)$$

Tasks:

- 1) Determine the a and b constants using the critical values and equations 3-4. $T_{kr} = 405\text{ K}$, $p_{kr} = 11,298 \cdot 10^6\text{ Pa}$.
- 2) Determine the a and b constants using experimental data on ammonia and equation 2. The starting values of the parameters can be taken from Task 1. Compare the pressures (experimental, fitted).

Experimental molar volumes of ammonia at 323.15 K

V_m (m ³ /mol)	p (Pa)
--------------------------------	----------

22.13E-3	1.202E5
5.110E-3	5.046E5
2.550E-3	9.697E5
1.456E-3	1.594E6

3) Compare graphically and tabulated the different pressures (experimental, fitted in Task 1, fitted in Task 2). Comment the applicability domain of this equation of state.

Sample test for Part I

1) (1 point) Compare the means of \bar{x}_1 and \bar{x}_2 in a statistical test. Use two sided-test with 95% confidence interval. Which function do you use? What is the result of the function? What is the final decision?

\bar{x}_1	\bar{x}_2
2.10	1.97
2.09	2.23
2.33	2.23
2.10	1.98
2.13	2.45
2.23	2.35

2) (3 points) 60 dosimetric measurements were performed in a school building constructed from ash-concrete. Using a time interval of 10 seconds the following frequency table was obtained in the Geiger-Müller measurements. Please check with statistical test, that the data can be modelled with Poisson-distribution. Give the calculation details and correctly justify your decision with numbers. What is the final decision with a significance of 0.05?

number of observations in 10 s	frequency
0	6
1	12
2	17
3	16
4	7
5	2

3) (3 points) A solution contains different organic compounds. The concentrations of compounds A, B and C can be determined by measuring at different wavelengths according to equation $\lg\left(\frac{I_f}{I_0}\right) = -\sum_i \epsilon_i c_i l$. The solvent does not absorb at the selected wavelengths. The molar absorption coefficients:

	$\epsilon_A / (\text{dm}^3 \cdot \text{mol}^{-1} \cdot \text{cm}^{-1})$	$\epsilon_B / (\text{dm}^3 \cdot \text{mol}^{-1} \cdot \text{cm}^{-1})$	$\epsilon_C / (\text{dm}^3 \cdot \text{mol}^{-1} \cdot \text{cm}^{-1})$
$\lambda=300 \text{ nm}$	110.3	26.7	56.3
$\lambda=400 \text{ nm}$	64.5	70.2	5.6
$\lambda=500 \text{ nm}$	19.4	10.8	120.1
$\lambda=350 \text{ nm}$	80.2	115.1	20.0
$\lambda=450 \text{ nm}$	50.2	201.3	40.3
$\lambda=550 \text{ nm}$	10.2	11.3	111.2

The measured absorbances:

	$-\lg(I_f/I_0)$
$\lambda=300 \text{ nm}$	0.250
$\lambda=400 \text{ nm}$	0.299
$\lambda=500 \text{ nm}$	0.270
$\lambda=350 \text{ nm}$	0.332
$\lambda=450 \text{ nm}$	0.384
$\lambda=550 \text{ nm}$	0.284

The holder is 1 cm wide. Determine the concentrations using Excel!

4) (1 point) Multiply the matrix with itself!

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 5 & 6 \\ 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 9 \end{pmatrix}$$

5) (3 points) The differential equation of the $A \xrightarrow{k_1} B \xrightarrow{k_2} C$ reaction can be analytically solved with $A(t=0)=1 \text{ mol/m}^3$, $B(t=0)=0 \text{ mol/m}^3$, and $C(t=0)=0 \text{ mol/m}^3$ initial values.

$$C(t) = 1 - \frac{k_2}{k_2 - k_1} e^{-k_1 t} + \frac{k_1}{k_2 - k_1} e^{-k_2 t}$$

Determine the k_1 and k_2 rate constants by parameter estimation. Proceed the fit on the data. Starting values of the parameters are $k_1=4$ and $k_2=5$.

$t(\text{s})$	$C(t)$ (mol/m^3)
0.5	0.21
1	0.41
2	0.86
3	0.91
4	0.95
5	0.98
6	0.99
7	0.99
8	1.01
9	1.02

6) (3 points) The dissociation of nitrogen-dioxide is due to the equation $2 \text{NO}_2 \rightleftharpoons 2 \text{NO} + \text{O}_2$. If the reaction is performed at constant volume, the equilibrium partial pressures of the components can be described using the initial pressure of nitrogen-dioxide ($p_{\text{NO}_2}^0$) and the degree of dissociation (α):

$$p_{\text{NO}_2} = (1 - \alpha) p_{\text{NO}_2}^0$$

$$p_{\text{NO}} = \alpha p_{\text{NO}_2}^0$$

$$p_{\text{O}_2} = 0.5 \alpha p_{\text{NO}_2}^0$$

Combining these equations into the equation of the equilibrium constant using partial pressures:

$$K_a = K_p \left[\frac{1}{p^\ominus} \right]^{\Delta v} = \frac{0.5 \alpha p_{\text{NO}_2}^0 (\alpha p_{\text{NO}_2}^0)^2}{(1 - \alpha)^2 (p_{\text{NO}_2}^0)^2} \frac{1}{p^\ominus} \quad (1)$$

where $p^\ominus = 100 \text{ kPa}$ is the standard pressure, Δv is the net change of stoichiometric numbers (here equals to 1). Equation (1) can be solved for α . Calculate the degree of dissociation!! Start with 120, 150 and 180 kPa pure nitrogen-dioxide at 700 K. The equilibrium constant is: $K_a = 0,18$. Compare the results to each other.

7) (2 points) Where is the absolute minima of $4x^3 - 6x^2 - 24x$, if $x \in [0; 10]$?

PART II

Basics of Octave

Set working directory and log file

```
> pwd
> cd myuserid/myplace (Linux) cd c:\a\oktatas\szamkem (Windows)
> diary logbook.txt
> ls
> quit
```

Initial steps

```
> x=2
> y=3;
> z=x+y**3
> x=cos(z)
> disp(" something to the screen")
> function y=f(x,z) y=x+z*3-sqrt(x);endfunction;
> f(3,4)
> x=2
> z=6
> f(x,z)
> y=exp(x)*log(x)-log10(x)+sqrt(x)-ceil(x)+floor(x)-fmod(x,2)+tan(x)
> round(2.4)
> round(2.6)
> sign(3.2)
> sign(-3.2)
> factor(12312)
> lookfor factor
> factorial(12)
> help primes
> primes(100)
```

> Row and column vectors

```
> v=[2;34;2]
> v=[2,34,2]
> u=rand(1,3)
> size(u)
```



```
> v=resize(v,1,4)
> u=resize(u,1,4)
> v*u'
> dot(v,u)
> sum(v)
> prod(v)
> sumsq(v)
> sumsq(v.*u)    # . = multiplication by elements
> v
> sort(v) # sorting
```

Complex numbers

```
> z=0.1+2i
> real(z)
> imag(z)
> abs(z)
```

Coordinate transformations

```
> [theta,r]=cart2pol(1,0)
theta = 0
r = 1
> cart2pol(1,0)
ans = 0
> [x,y]=pol2cart(0,1)
x = 1
y = 0
> [x,y,z]=sph2cart(0,0,1)
x = 1
y = 0
z = 0
```

> Graphics (after starting 'startx' on Linux)

```
> plot(u,v,"4+")    #4th colour, + sign
> print -djpg proba.jpg #saves graph to a JPG file
```

Data storage – matrices - statistics

n rows: each row means one object

m columns: each column means one variable

$$D = \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix} \quad n \text{ row, } m \text{ columns}$$

Centering of data: subtraction of column means

$$c_{ij} = d_{ij} - \frac{\sum_{i=1}^n d_{ij}}{n} = d_{ij} - \bar{d}_j$$

Scaling: dividing with the standard deviation of the columns

$$c_{ij} = \frac{d_{ij}}{s_j} = \frac{d_{ij}}{\sqrt{\frac{\sum_{i=1}^n (d_{ij} - \bar{d}_j)^2}{n-1}}}$$

Standardization (studentization, z-scoring) In the case of normal distribution most of the results in -3;+3

$$c_{ij} = \frac{d_{ij} - \bar{d}_j}{s_j}$$

```
> D=rand(10,5)
> D(:,2)=cos(D(:,1))
> D(:,3)=D(:,1)*2+rand()
> D(:,4)=exp(D(:,1))
> D(:,5)=D(:,2)+D(:,4)+rand()*0.1
> mean(D)
> median(D)
> meansq(D)
> std(D)
> var(D)
> sortrows(D,2)
> statistics(D)
> help statistics
> center(D)
> A=studentize(D)
> mean(A) #standardizált átlaga 0
> std(A) #standardizált szórása 1
```

Covariance – common moving of two set of variables (sign and magnitude)

$$s_{xy}^2 = \frac{\sum_{i=1}^n (x_i - E_x)(y_i - E_y)}{n-1}$$

Correlation coefficient, result in [-1,1], connection to linear relationship:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - E_x)(y_i - E_y)}{\sqrt{\sum_{i=1}^n (x_i - E_x)^2 \sum_{i=1}^n (y_i - E_y)^2}} = \frac{s_{xy}}{s_x s_y}$$

> cov(D) #in symmetric matrix

> cor(D)

> cov(A) #standardized covariance = correlation of the original

Statistical tests

F-test

> p=var_test(D(:,1),D(:,2))

directly the p-value of H_0 :

if $\alpha \leq p$, H_0 is accepted

if $p < \alpha$, H_0 is rejected (H_a is accepted).

One sample t-test

> p=t_test(D(:,1),0.5)

> p=t_test(D(:,1),1)

> p=t_test(D(:,1),2,"<>")

Two-sided case:

if $\alpha \leq p$ H_0 is accepted

if $p < \alpha$, H_0 is rejected (H_a is accepted)

$30 < N$, normal distribution can be used:

One-sided case:

> p=t_test(D(:,1),0.78,"<")

> p=t_test(D(:,1),0.9,">")

Two-sample t-test:

> p=t_test_2(D(:,1),D(:,2),"<>")

Analysis of variance

> anova(D)

Matrix manipulation

> A=[1,2;3,4]

> B=randn(2,2)

> A*B

> C=randn(2,3)

> C*A

error: operator *: nonconformant arguments (op1 is 2x3, op2 is 2x2)

```
> A*C
> C'*A
> det(A)
> inv(A)
> A*inv(A)
> eig(A)
> [vA,eA]=eig(A)
> A==B
> v=vec(A)
> u=v'
> B=eye(2)          #Diagonal Matrix
> trace(A)
```

Solution of non-linear equations, search for minima and maxima

```
> function u=f(r) u=4*1*(3**12/r**12-3**6/r**6); endfunction; # Lennard-Jones potential
> for i=1:10 v(i)=i*0.5+2.3;u(i)=f(v(i));endfor;
> plot(v,u)
> fsolve(@f,2.4)
> fsolve(@f,3.2)
> fsolve(@f,4)
> fzero(@f,[2,4])
> fminunc(@f,3)
> fminbnd(@f,2,4)
```

Solution of a set of linear equations

```
> A=resize(A,3,3)
> A=[2,4,6;2,3,1;-1,0,5]
> b=[8;7;-2]
> inv(A)*b
> B=resize(B,3,3)    #to practice Cramer-rule
> B=A
> B(:,1)=b
> det(B)/det(A)
> B=A
> B(:,2)=b
> det(B)/det(A)
```

Solution of set of linear equations – overdetermined case

```
> A=resize(A,4,3)
> A(4,:)=2,3,2]
> b=resize(b,4,1)
> b(4)=4
> inv(A'*A)*A'*b
```

> Adding constant to the predictor matrix

```
> A=resize(A,5,4)
> A(:,4)=1
> A(5,:)=[-1,5,-1,1]
> b=resize(b,5,1)
> b(5)=0
> inv(A'*A)*A'*b
```

Distribution of a function of a random variable

$y=y(x)$ and $x=x(y)$ are one to one unique functions, x and y are random variables. What is the distribution of y , if we know the distribution of x ?

$$f_y(Y)|dy| = f_x(x(Y))|dx|, \text{ by rearrangement: } f_y(Y) = f_x(x(Y)) \left| \frac{dx}{dy} \right|$$

Therefore, the density function $f(y)$ can be determined, if the link between the random variables is known, for every Y , where dx/dy exist and y is continuous.

Task: Determine $f_y(y)$, if $y = \sqrt{x}$ and x can be described with an exponential distribution with $\lambda=1$ ($x \geq 0$). Please also tabulate the density functions with x or y grid size of 0.1 and plot them.

```
for (i=1:60) v(i)=0.1*i; endfor;
for (i=1:60) fx(i)=exp(-v(i)); endfor;
for (i=1:60) fy(i)=2*v(i)*exp(-v(i)*v(i)); endfor;
plot(v,fx)
plot(v,fy)
intx=sum(fx)*0.1
inty=sum(fy)*0.1
```

Task: Determine $f_y(y)$, if $y=x^2$ and x is standard normally distributed. Please tabulate the density functions with x or y grid size of 0.2 and plot them. Is the area under the curve equal to 1? If not, why? Correct it by taking care on the domains of x and y !

Error propagation (by Gauss)

$y=f(x_1,x_2,\dots,x_n)=f(\underline{x})$ is a multivariate function. We do not know the exact $x_1\dots x_n$ variables, we estimate them only with $\xi_1\dots \xi_n$ random numbers. We can calculate the estimated means $E_1\dots E_n$, the estimated standard deviations ($s_1\dots s_n$), their squares - variances (s_i^2) and the covariances (s_{ij}^2). What is the variance of y ?

Using approximations (Taylor-expansion, omitting terms over the linear ones) it can be derived, that

$$s_y^2 = \sum_{i=1}^n \left(\frac{\partial y}{\partial x_i} \right)^2 s_i^2 + 2 \sum_{i < j} \sum_{j=1}^n \left(\frac{\partial y}{\partial x_i} \right) \left(\frac{\partial y}{\partial x_j} \right) s_{ij}^2 = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial y}{\partial x_i} \right) \left(\frac{\partial y}{\partial x_j} \right) s_{ij}^2, \text{ where the partial derivatives are calculated at the}$$

$E_1\dots E_n$ values of the $x_1\dots x_n$ random variables.

Task: Derive for $y = a_0 + \sum_{i=1}^n a_i x_i$ the equation of s_y^2 ! In the case of linear combination $E_y = a_0 + \sum_{i=1}^n a_i E_i$.

Task: How can you simplify the equation, if $\xi_1\dots \xi_n$ are independent random numbers?

Task: Generate a 10x3 X matrix, where at least one column is a function of the others with some numerical error (n=3 columns (=variables), m= 10 rows (=e.g. measurement)). Estimate the variance of y using the above equation. Calculate y_i also for each row of X, and compare the variance obtained directly from these numbers and from the equation! Use the following functions:

a) $y = x_1 + 2 * x_2 - 3 * x_3$

D=randn(10,3)

D(:,3)=D(:,1)+2*D(:,2)+0.1*randn()

CD=cov(D)

y=D(:,1)+2*D(:,2)-3*D(:,3)

PD=[1,2,-3;2,4,-6;-3,-6,9]

var(y)

sum(sum(CD.*PD))

b) $y = x_1 * x_1 + \cos(x_2) * x_3$

c) $y = x_1 * x_1 + \cos(x_2) + x_3$

Check for the linear case, that the expected value can be calculated correctly with the equation shown two exercises earlier!

Advance task: Give the confidence interval for E_y using one of the functions of the previous tasks. The degrees of freedom for the t-distribution can be approximated as:

$$V_y = \frac{s_y^4}{\sum_{i=1}^n \left(\frac{\partial y}{\partial x_i} \right)^4 s_i^4 \frac{1}{m-1}}, \text{ where } m \text{ is the number of measurements (rows of } X)$$

Compare the confidence interval with that one calculated directly from the y_i values.

Integration, polynomials, interpolation

> Integration

```
> u=rand(1,20);
> for i=1:20 v(i)=i*0.1; endfor;
> u
> v
> trapz(v,u)
> quad('cos',0,pi()/2)
```

> Polynomials

```
> c=[2,3,4,5,1]; # coefficients (last one is the constant term)
> a=[1,2,3,4];
> conv(a,c) # product
> polyder(a) # derivation
> q=polyder(a,c) # derivation of the product
> [q,r]=polyder(a,c) # rational functions
> roots(a) #roots
> for i=1:20 u(i)=v(i)+cos(v(i)*0.3); endfor;
> p=polyfit(v,u,5) # fit of polynomials
> pint=polyint(p) # integration of polynomials
> polyval(pint,pi()/2)-polyval(pint,0) # evaluate the polynomial
```

> Interpolation

```
> interp1(v,u,1.05)
> interp1(v,u,1.05,'linear')
> interp1(v,u,1.05,'cubic')
> interp1(v,u,3.05,'extrap')
```

Task

Polynomial fit on measured data, integration

The task is the determination of the time integral of data are recorded in a measurement. The recording of the data in the intermediate time interval is not possible, but the measured quantity can be fairly approximated by cubic polynomial.

Fit a cubic polynomial on the data and calculate its integral for $t \in [0;10]$. Compare the value to another one obtained by simple trapezoidal integration of the original data point. In the latter case use linear extrapolation to $t=0$.

Data:

t	$y(t)$
0.5	-4.5
1	-6.8
1.5	-8.9
2	-12.5
2.5	-15.9
7.5	17.0
8	35.6
8.5	57.6
9	84.9
9.5	117.7
10	155.9

Fourier-transformation, frequency and noise filtering

Jean-Baptiste Joseph Fourier (1768—1830)

The function contains the same information but respect to another variable

Fourier-series expansion

$$f(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos \frac{2\pi n t}{T} + \sum_{n=1}^{\infty} b_n \sin \frac{2\pi n t}{T}$$

$$\frac{a_n}{2} = \frac{1}{T} \int_{t_0}^{t_0+T} f(t) \cos \left(\frac{2\pi n t}{T} \right) dt \quad \frac{b_n}{2} = \frac{1}{T} \int_{t_0}^{t_0+T} f(t) \sin \left(\frac{2\pi n t}{T} \right) dt \quad a_0 = \frac{1}{T} \int_{t_0}^{t_0+T} f(t) dt$$

Fourier-expansion of a periodical function using the exponential form:

$$e^{it} = \cos(t) + i \sin(t) \quad f(t) = \sum_{n=-\infty}^{\infty} c_n e^{i\omega t}$$

$$\omega = \frac{2\pi n}{T} \quad \nu = \frac{n}{T} \quad c_n = \bar{c}_{-n} = \frac{a_n - ib_n}{2}, n \neq 0$$

Fourier-transformation – for continuous non-periodical functions

Fourier-transformation:

inverse Fourier-transformation:

$$F(\nu) = \int_{-\infty}^{\infty} f(t) e^{-i2\pi\nu t} dt$$

$$f(t) = \int_{-\infty}^{\infty} F(\nu) e^{i2\pi\nu t} d\nu$$

Discrete Fourier-transformation

N samples with Δt time grid

$$\text{notations: } \Delta \nu = \frac{1}{N\Delta t}$$

$$2\pi\nu t = \frac{2\pi n k}{N}$$

$$F(\nu) = F(k\Delta\nu) = F_k$$

$$f(t) = f(n\Delta t) = f_n$$

Fourier-transformation:

inverse Fourier-transformation:

$$F_k \equiv \frac{1}{N} \sum_{n=0}^{N-1} f_n e^{-\frac{i2\pi n k}{N}}$$

$$f_n \equiv \sum_{k=0}^{N-1} F_k e^{\frac{i2\pi n k}{N}}$$

Fast Fourier-transformation (FFT): special order of the mathematical calculations to reduce the calculation time, if we have $N=2^k$ data. If we have less data, it is feasible to fill the rest of the data with zeroes up to $N=2^k$.

Tasks:

Transformation of 1 sine function:

```
> n=256
> dt=1/n
> for i=1:n t(i)=(i-1)*dt; endfor;
> f=sin(t*15*2*pi);
> plot(t,f)
> dnu=1/n/dt
> for i=1:n nu(i)=(i-1)*dnu; endfor;
> F=fft(f);
> Fa=abs(F);
> Fr=real(F);
> plot(nu,Fa)
> plot(nu,Fr)
> F(15)
> F(16)
> F(17)
> finv=ifft(F);
> plot(t,finv)
```

Sum of 3 sines, filtering respect one of them (frequency filtering):

```
> f=sin(t*15*2*pi)+sin(t*5*2*pi)+sin(t*30*2*pi);
> plot(t,f)
```

```

> F=fft(f);
> Fa=abs(F);
> plot(nu,Fa)
> F(16)
> F(242)
> F(16)=0+0i;
> F(242)=0+0i;
> finv=ifft(F);
> plot(t,finv)
> plot(t,f)

```

Sum of 3 sines + noise, noise filtering (amplitude filtering):

```

> f=sin(t*15*2*pi)+sin(t*5*2*pi)+sin(t*30*2*pi);
> fz=f;
> fz=f+2*randn(1,256);
> plot(t,fz)
> F=fft(fz);
> Fa=abs(F);
> plot(nu,Fa)
> for i=1:n if(Fa(i)<80) F(i)=0+0i; endif; endfor;
> finv=ifft(F);
> plot(t,finv)
> plot(t,f)

```

Transformation of Gaussians

```

> dx=0.1;
> for i=1:n x(i)=(i-1)*dx; endfor;
> s=1 #repeat from here with s=0.5 and s=3!
> for i=1:n g(i)=1/sqrt(2*pi()*s*s)*exp(-x(i)*x(i)/2/s/s); endfor;
> plot(x,g)
> dnu=1/dx/n;
> for i=1:n nu(i)=(i-1)*dnu; endfor;
> G=fft(g);
> Ga=abs(G);
> plot(nu,G);

```

Modelling

Examples on modelling in chemistry:

- Modelling of intramolecular and atomic structure – electronic structure/orbitals are interesting: quantum mechanics and quantum chemistry
- Interaction of atoms and molecules are focused, but the intramolecular structure is treated in an averaged way: classical mechanics, statistical mechanics
- Quantitative structure-activity relationship (QSAR): e.g., computer aided drug design
- Mesoscopic modelling: fluid dynamics, crystallization, formation of mesoscopic structures
- Macroscopic modelling: hydrodynamics, reaction kinetics, biological models

Simulation: A process is investigated on a more or less realistic model –numerical experiment.

Molecular mechanics, Molecular dynamics, Monte Carlo simulations: The atoms and the molecules interact mostly with classical mechanical potentials. The intramolecular interactions are connected to classical mechanical potential functions on bond lengths, bond angles and dihedral angles. The intermolecular terms are electrostatic ones (partial charges, dipole interactions) and dispersion ones (e.g. Lennard-Jones). Mostly condensed phase systems are investigated and the system includes around a few thousand particles. The interaction parameters are determined on the basis of experiments and quantum chemical calculations. Molecular mechanics: searches for energy minimum of large systems, e.g. biomolecules, results minimum energy geometry. Molecular dynamics: finite temperature simulations due to the Newtonian dynamics, the real systems are followed for 10^{-12} - 10^{-7} s (in simulation time several weeks). Structural, energetic and dynamical properties are available. Monte Carlo simulation: extensive use of random numbers to generate different particle arrangements in agreement with statistical mechanical theories to get data on energetic and structural features.

Quantum chemistry: the electronic structure of atoms and molecules are investigated. Many properties can be obtained, e.g. orbitals, partial charges, optimized molecular geometries, spectroscopic data and formation of chemical bonds. There are two main factors in the quality of the calculations: the method and the bases set. The calculation time extremely increases, when accurate methods and large systems are investigated.

Task: Use of the molecular mechanics, molecular dynamics and quantum chemical modules of the Chem Office software on small systems, e.g. for dipeptides in the case of classical mechanical calculations and water for quantum chemical calculations (e.g. Hartree-Fock calculation on minimal bases set to optimize geometry and calculate spectra).

Examples on didactical science simulations: Collections of Sándor Nagy (<http://www.nagysandor.eu>) and Dennis Rapaport.

Sample test for Part II

1) (1 point) Give the Cartesian coordinates of the following vector given in polar coordinates $\mathbf{v}=(\varphi=1,23; r=5,1)$.

2) (1 point) Calculate the covariance matrix of the transpose of the given data matrix!

$$\begin{pmatrix} 1 & 3 & 3 & 4 \\ 2 & 3 & 5 & 6 \\ 1 & 3 & 5 & 9 \\ 2 & 4 & 9 & 9 \end{pmatrix}$$

3) (2 points) Calculate the sum of the squares of the positive odd numbers smaller than 1000. Give the Octave commands and the result.

4) (3 points) Determine the variance of y using the Gauss error-propagation form, if $y=x_1*x_2-\sin(x_2)$. Compare the result to the one calculated directly on the y_i values.

x_1	x_2
1.84	1.55
2.77	3.96
1.97	6.00
2.59	2.11
2.73	2.55
2.08	3.59
0.12	0.60
2.53	3.88
3.06	3.44
1.45	2.81

5) (2 points) Calculate the inverse, the eigenvalues and the eigenvectors of the matrix!

$$\begin{pmatrix} 1 & 2 & 2 & 4 \\ 5 & 3 & 5 & 6 \\ 1 & 3 & 6 & 7 \\ 2 & 7 & 8 & 9 \end{pmatrix}$$

6e) (2 points) Fit a third-order polynomial on x and y and calculate its integral for the interval $0 \leq x \leq 2$!

x	y
0	1.88
0.2	1.65
0.4	1.88
0.6	2.25
0.8	2.17
1	2.35
1.2	3.24
1.4	3.48
1.6	3.49
1.8	4.06
2	4.13

7) (2 points) Evaluate the polynomial for $x=2.125$

$$P(x) = (2x^7 - x^6 - x^3 - x^2 - 12x - 6) * (7x^4 - 8x^2 + 2x + 1)$$

8) (3 points) Fourier-transform the following data (32 data pairs). Estimate p_1-p_2 , if the original function was calculated as $f(t)=\sin(p_1 2\pi t)+\sin(p_2 2\pi t)+2$. Be careful, there are small non-zero elements in the transformed function (noise) at many frequencies due to the truncation error. You may check your results by generating new data with the obtained parameters.

t	$f(t)$	t	$f(t)$	t	$f(t)$	t	$f(t)$
0	2.000	0.25	2.000	0.5	2.000	0.75	2.000
0.03125	3.631	0.28125	1.783	0.53125	3.631	0.78125	1.783
0.0625	3.707	0.3125	2.293	0.5625	3.707	0.8125	2.293
0.09375	2.324	0.34375	3.090	0.59375	2.324	0.84375	3.090
0.125	1.000	0.375	3.000	0.625	1.000	0.875	3.000
0.15625	0.910	0.40625	1.676	0.65625	0.910	0.90625	1.676
0.1875	1.707	0.4375	0.293	0.6875	1.707	0.9375	0.293
0.21875	2.217	0.46875	0.369	0.71875	2.217	0.96875	0.369

References:

Further reading:

S. Dowdy and S. Wearden: Statistics for research, Wiley 1982, New York

Source and original idea of exercises:

[1] Deutsch T., Vajda S., Valkó P., Riedel M: Kémiai számítástechnika füzetek, ELTE TTK Kémiai Kibernetikai Laboratórium, 1976-1987