Guidelines

- Write your name and matriculation number on each sheet of paper.
- Only clearly readable exercise-elaborations are evaluated.
- Results have to be provided together with an evident way of calculation.
- Keep textual answers short and concise. Lengthy or vague statements won't gain points.

Exercise 5.1 (0.5 points)
Given $f(x) = x^3$ with $x \in [-1,1]$ (continuous interval), the least squares approximation seeks a polynomial $p(x) \in \mathcal{P}$ of order 2 that minimizes the error in the l_2-norm, which is equivalent to

$$\min_{p\in\mathcal{P}} \|e\|_2^2 = \min_{p\in\mathcal{P}} \left\| f(x) - p(x) \right\|_2^2.$$

1. Assume that $p(x) = c_0 + c_1 x + c_2 x^2$ is built from the basis functions $\{1, x, x^2\}$. Find the least squares coefficients c_0, c_1 and c_2 using basic calculus.
2. With the same basis functions $\{1, x, x^2\}$, find $p(x)$ using the orthogonality principle.
3. Given an arbitrary basis $\phi_k(x)$, $k = 0,1,2$. Then any $p(x) \in \mathcal{P}$ can be written as
$$p(x) = \sum_{k=0}^{2} c_k \phi_k(x).$$
Formulate the least squares problem
$$\min_{c_k} \left\| f(x) - \sum_{k=0}^{2} c_k \phi_k(x) \right\|_2^2$$
in matrix-vector form $A\underline{c} = \underline{b}$ where $\underline{c} = [c_0, c_1, c_2]^T$. What is A and $\underline{b}$ with respect to $\phi_k(x)$ and $f(x)$?
4. Now assume that $\phi_k(x)$, $k = 0,1,2$, is an orthonormal basis. How do you find the least squares coefficients $\underline{c} = [c_0, c_1, c_2]^T$?

Terms: least squares, orthogonality principle, Gramian.

Exercise 5.2 (0.5 points)
Consider the following data set obtained via measurement:

$$\underline{x} = [1,2,3,4,5,6,7,8]^T \ , \ \underline{\widetilde{y}} = [3,1,1,1,2,3,6,13]^T \, .$$

Unfortunately, the student who conducted the measurement rounded $\underline{\widetilde{y}}$ to integer values. Let us try to figure out the real-valued numbers in $\underline{\widetilde{y}}$ using regression techniques. *Use Matlab or similar tools for computations and visualization, analytical statements have to be conducted by hand.*

1. From previous experiments, we know that the data obeys $\widetilde{y} = x^a \mathrm{e}^{bx}$. Find a suitable transformation (function) so that you can write this relation in matrix vector form $\underline{y} = A\underline{c} + \underline{e}$, where $\underline{c} = [a,b]^T$ stores the coefficients and $\underline{e}$ is the error vector. Determine the coefficients a and b that approximate y in the least squares sense; explicitly state the least squares solution for $\underline{c}$. *(Note that afer transformation, $\widetilde{y}$ becomes y.)*

2. Assume that an oracle reveals the original coefficients as $a = -3$ and $b = 1.1$. Compute the mean squared error between your least squares approximation and the original as $\mathrm{MSE} = \frac{1}{N}\|\underline{\widetilde{e}}\|_2^2$, N being the number of elements in $\underline{\widetilde{e}}$.
Hint: $\underline{\widetilde{e}} = A\underline{c}_{orig} - A\underline{c}_{LS}$.
A valid alternative is to consider $\underline{\widetilde{e}} = x^{a_{orig}}\mathrm{e}^{b_{orig}x} - x^{a_{LS}}\mathrm{e}^{b_{LS}x}$.

   We can rewrite the least squares error formulation as $\underline{e} = B\underline{y}$ using the least squares coefficient vector $\underline{c}$. Determine B; what can you say about this matrix/what properties does it have?
Hint: for this part, start from $\underline{e} = \underline{y} - A\underline{c}_{LS} = ... = B\underline{y}$.

3. The student remembers that the data points at $x = \{1,3,8\}$ were pretty close to the integer number and required "almost no rounding". Introduce a weighting matrix into the least squares formulation that emphasizes these measurements in the coefficient computation. Give these measurements 100 times more weight and compute the $\mathrm{MSE} = \frac{1}{N}\|\underline{\widetilde{e}}\|_2^2$ with the new coefficients; by what factor did the MSE improve?

4. Plot the data points and all regressions in a single figure in order to visualize the results. *(Plot function $\widetilde{y} = x^a\mathrm{e}^{bx}$ of original and all regeressions vs. $x \in [1,8]$. Also Plot the given data points.)*

Terms: least squares, regression, weighted least squares.

Exercise 5.3 (0.5 points)
In this example, the differential equation

$$x\frac{d^2}{dx^2}y(x) + (1-x)\frac{d}{dx}y(x) + my(x) = q(x)$$

with $q(x) = 1 + x(10x+1)$ is investigated. As the canonical solution is unknown a priori, we first try to approximate it by orthogonal polynomials p_n:

$$y(x) = a_0 p_0(x) + a_1 p_1(x) + a_2 p_2(x) + \dots$$

1. Use Tschebyscheff polynomials with highest order $n = 3$ and calculate the coefficients $a_0, \dots, a_3$ for the homogeneous (i.e. $q(x) = 0$) and inhomogeneous differential equation ($a_0 = 1$, $m = 3$).

2. Not really satisfied with this solution, we ask for help in the yellow area of Freihaus. There, we get the hint that the differential equation is related to Schrödinger's equation to describe the radial part of the hydrogen atom, and we should use a power series:

$$y(x) = \sum_{n=0}^{\infty} a_n x^n.$$

   (a) Calculate the coefficients $a_0, \dots, a_3$ of $y(x) = \sum_{n=0}^{3} a_n x^n$ for the homogeneous differential equation ($q(x) = 0$, $y(x=0) = 1$).

   (b) Show that the calculated coefficients match with $a_n = (-1)^n \binom{m}{n} \frac{a_0}{n!}$ for $m = 3$. These are the coefficients of the so called *Laguerre polynomial* $L_m(x)$ of order $m = 3$. The differential equation is therefore called the Laguerre-differential equation with the canonical solutions:

$$L_m(x) = \frac{e^x}{m!}\frac{d^m}{dx^m}(x^m e^{-x})$$

   (c) With the weighting function $w(x) = \exp(-x)$ Laguerre polynomials form an orthogonal family of polynomials on the interval $[0,\infty)$. Show $\langle L_0, L_2 \rangle_{w(x)} = 0$.

Terms: least squares, Tschebyscheff polynomial, Laguerre polynomial.

MATLAB-Exercise 5.1 (1 point)
Consider the quadratic function:

$$f(\underline{x}) = \underline{x}^T R \underline{x} - 2\underline{b}^T \underline{x}$$

with

$$R = \begin{bmatrix} 105 & 95 \\ 95 & 105 \end{bmatrix} \qquad \underline{b} = \begin{bmatrix} 200 \\ 200 \end{bmatrix} \qquad \text{and} \qquad \underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

From this function, the minimum is to be found with help of the gradient algorithm (steepest descent). First generally calculate the gradient of the function f

$$\nabla f(\underline{x}) = \frac{\partial f}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}$$

and indicate the minimum $\underline{x}_0$ of $f(\underline{x})$.
Now implement the gradient algorithm in Matlab

$$\underline{x}^{(i+1)} = \underline{x}^{(i)} - \mu \nabla f(\underline{x}^{(i)}).$$

Examine in particular the convergence for different step sizes of μ. From which step size μ does the algorithm become unstable?
Terms: Gradient, steepest decent, stability.