Exercise 10.1. Find the optimal polynomial p of the second degree, i.e.

$$p(x) = a_0 + a_1 x + a_2 x^2,$$

which minimizes the distance to the following points in the least squares sense:

x	-1	0	1	2
y	0	1	0	2

Sketch the points and the optimal p in a Cartesian coordinate system.

Suggested Solution. The problem is equivalent to minimizing

$$\mathcal{E}(x) = \|Ax - b\|_2^2$$

on $\mathbb{R}^n$, where

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 2 \end{pmatrix}, \quad x = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \in \mathbb{R}^3.$$

With regard to the gaussian normal equations we calculate

$$A^T A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{pmatrix}$$

and

$$A^T b = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 8 \end{pmatrix}.$$

The solution of the Gaussian normal equations

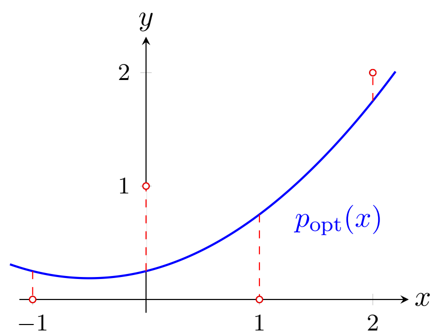
$$A^T A = b$$

finally results in

$$x = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Thus the optimal polynomial is

$$p_{\text{opt}}(x) = \frac{1}{4} + \frac{1}{4}x + \frac{1}{4}x^2.$$



Exercise 10.2. For $n \in \mathbb{N}$, let the following arbitrary pairs of points of $\mathbb{R}^2$ be given:

x	x_1	x_2	$\cdots$	x_n
y	y_1	y_2	$\cdots$	y_n

Here, the x_i should not all be the same. Let the arithmetic mean values of the x_i and y_i be denoted by $\bar{x}$ and $\bar{y}$, i.e.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Show that

$$\beta = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad \text{and} \quad \alpha = \bar{y} - \beta\bar{x}$$

are the optimal parameters of the regression line $y = \alpha + \beta x$ in the least squares sense.

Suggested Solution. The problem is equivalent to minimizing

$$\mathcal{E}(z) := \|Az - y\|_2^2$$

over $\mathbb{R}^2$, where

$$A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad z = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

Writing $n\bar{x} = \sum_{j=1}^{n} x_j$ and $|x|^2 := \sum_{j=1}^{n} x_j^2$, for $x, y \in \mathbb{R}^n$, we obtain for the normal equations

$$A^T A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & |x|^2 \end{pmatrix}$$

and

$$A^T y = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} n\bar{y} \\ \langle x, y\rangle \end{pmatrix}$$

We need to solve

$$\begin{aligned} A^T A z = A^T y \quad &\Leftrightarrow \quad \left(\begin{array}{cc|c} n & n\bar{x} & n\bar{y} \\ n\bar{x} & |x|^2 & \langle x, y\rangle \end{array}\right) \\ &\Leftrightarrow \quad \left(\begin{array}{cc|c} 1 & \bar{x} & \bar{y} \\ 0 & |x|^2 - n\bar{x}^2 & \langle x, y\rangle - n\bar{x}\bar{y} \end{array}\right). \end{aligned}$$

This yields the optimal parameter

$$\beta = \frac{\langle x, y\rangle - n\bar{x}\bar{y}}{|x|^2 - n\bar{x}^2} \quad \text{and} \quad \alpha = \bar{y} - \beta\bar{x}.$$

To achieve the shape for β from the exercise we reshape the expression a little more:

$$\beta = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\langle x, y\rangle - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}}{|x|^2 - 2n\bar{x}^2 + n\bar{x}^2} = \frac{\langle x, y\rangle - n\bar{x}\bar{y}}{|x|^2 - n\bar{x}^2}.$$

Exercise 10.3. Consider the model function

$$f(x) = a_0 + a_1 x + a_2 e^x + a_3 e^{-x}.$$

This is to be optimally fitted to the data points

x	-2	-1	0	1	2
y	e^{-2}	e^{-1}	1	e^{-1}	e^{-2}

in least squares sense.

(a) Determine a matrix A and a vector y, s.t. this is equivalent to minimizing

$$\mathcal{E}(z) = \|Az - y\|_2^2$$

over $\mathbb{R}^2$.

(b) Calculate the optimal model parameters by solving the Gaussian normal equations with `Octave`.

(c) Graph the data points and the best model function (e.g. with `Octave`).

Remark: For part (c) you can use the `Octave` built-in function `plot` (see e.g. Ferreira - MATLAB Codes...)

Suggested Solution.

(a) Here, it is

$$\begin{pmatrix} 1 & -2 & e^{-2} & e^2 \\ 1 & -1 & e^{-1} & e \\ 1 & 0 & 1 & 1 \\ 1 & 1 & e & e^{-1} \\ 1 & 2 & e^2 & e^{-2} \end{pmatrix}, \quad b = \begin{pmatrix} e^{-2} \\ e^{-1} \\ 1 \\ e^{-1} \\ e^{-2} \end{pmatrix}.$$

(b) and **(c)**

```
A = [1,-2,exp(-2),exp(2);
1,-1,exp(-1),exp(1);
1,0,1,1;
1,1,exp(1),exp(-1);
1,2,exp(2),exp(-2)]
AtA = A'*A
y=[exp(-2);exp(-1);1;exp(-1);exp(-2)];
Aty = A'*y
s=AtA\Aty
datax = [-2,-1,0,1,2];
datay = b';
x=-2.5:0.01:2.5;
y=s(1)+s(2)*x+s(3)*exp(x)+s(4)*exp(-x);
plot(x,y,datenx,dateny,"*")
```

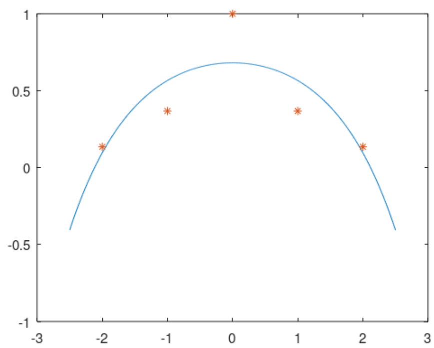


Figure 1: optimal model function with data points

Exercise 10.4. Let $A \in \mathbb{R}^{n\times n}$ be symmetric and $b \in \mathbb{R}^n$. Consider the following mappings

$$\begin{aligned} f\colon \mathbb{R}^n &\to \mathbb{R}, \quad x \mapsto \langle x, Ax\rangle, \\ g\colon \mathbb{R}^n &\to \mathbb{R}, \quad x \mapsto \langle x, b\rangle. \end{aligned}$$

Show that it then holds for the derivative of f or g:

$$\nabla f = 2Ax \quad \text{and} \quad \nabla g = b.$$

Suggested Solution. Using $\partial_k := \frac{\partial}{\partial x_k}$ we obtain using the product rule

$$\begin{aligned} \partial_k f(x) &= \partial_k \sum_{ij} a_{ij} x_i x_j \\ &= \partial_k \sum_i x_i \sum_j a_{ij} x_j \\ &= \sum_i \partial_k x_i \sum_j a_{ij} x_j + \sum_i x_i \sum_j a_{ij} \partial_k x_j \\ &= \sum_j a_{kj} x_j + \sum_i x_i \underbrace{a_{ik}}_{=a_{ki}} \\ &= 2(Ax)_k. \end{aligned}$$

Here, $(x)_k$ denotes the k^{th} component of the vector x. Finally, we obtain

$$\partial_k g(x) = \partial_k \sum_i x_i b_i = b_k.$$

Remark: Using these results, we could have derived the normal equations in another way (cf. Remark (a) on page 44 of the lecture notes): With

$$\begin{aligned} \mathcal{E}(x) &= \langle Ax - b, Ax - b\rangle \\ &= \langle Ax, Ax\rangle - 2\langle Ax, b\rangle + \|b\|_2^2 \\ &= \langle x, A^T Ax\rangle - 2\langle x, A^T b\rangle + \|b\|_2^2 \end{aligned}$$

it follows

$$\nabla \mathcal{E}(x) = 2A^T Ax - 2A^T b.$$

The exercises can be discussed in the Tuesday lecture next week.