Università della Svizzera italiana | Institute of Computing CI

Numerical Computing 2023

Lecture 6: Introduction to Least Squares

Dr. Edoardo Vecchi

Lecture Overview

- This set of slides contains the material concerning lecture 6, which provides a general introduction to least squares and their application to the solution of some exercises.

- Specifically, we will address the following points:
  - We start from the solution of inconsistent systems of equations, we introduce the general idea behind least squares and present some exercises on linear and quadratic models.
  - We then move to models suitable for periodic data and conclude with exponential and power law models, whose solution is here addressed through linearization.

- Feel free to get in touch with me or the TAs in case you have any doubt or concern.

Table of Contents

Inconsistent Systems of Equations

Consider the following inconsistent system of 3 equations in 2 unknowns:

$$\begin{aligned} 2x_1 - x_2 &= 3, \\ x_1 + 2x_2 &= 7, \\ 2x_1 - x_2 &= 4. \end{aligned}$$

Problem: The left-hand side of the first and third equations are the same, while the right-hand side is different. Thus, we conclude that it does not exist a vector $\mathbf{x} \in \mathbb{R}^2$ which satisfies the three equations simultaneously.

However, we could still be interested in finding a vector $\mathbf{x}$ which is **close enough to be a solution**, since this kind of problem arises frequently in all the cases where the number of equations m is greater than the number of unknowns n.

Solving an Inconsistent System of Equations

In other words, starting from a system of equations in the form:

$$A\mathbf{x} = \mathbf{b}, \tag{1}$$

with matrix of coefficients $A \in \mathbb{R}^{m \times n}$ and right-hand side $\mathbf{b} \in \mathbb{R}^n$, we are interested in finding the vector $\mathbf{x} \in \mathbb{R}^n$ which minimizes the following quantity:

$$\|A\mathbf{x} - \mathbf{b}\|_2^2 = 0.$$

The least-squares solution $\mathbf{x}^*$ can be found by solving the system of *normal equations*:

$$(A^T A)\mathbf{x}^* = A^T\mathbf{b}.$$

Quality of the Approximation: Residual Vector

After computing our least square solution $\mathbf{x}^*$, we can evaluate the quality of our approximation by computing the residual vector $\mathbf{r} \in \mathbb{R}^n$, which is defined as follows:

$$\mathbf{r} = \mathbf{b} - A\mathbf{x}^*.$$

Of course, if the system is **consistent** and $\mathbf{r} = 0$, the solution $\mathbf{x}^*$ corresponds to the exact solution of the given system of equations. If instead—as we expect—we have that $\mathbf{r} \neq 0$, we need to define some metrics to measure the quality of our approximation.

Quality of the Approximation: Metrics

1. The **Euclidean norm** of the residual, which is simply given by:

$$\|\mathbf{r}\|_2 = \sqrt{\sum_{i=1}^{n} r_i^2}.$$

2. The **SE**, or squared error, simply corresponding to:

$$\mathsf{SE} = \|\mathbf{r}\|_2^2 = \sum_{i=1}^{n} r_i^2.$$

3. The **RMSE**, or root-mean-square error, which is given by:

$$\mathsf{RMSE} = \sqrt{\frac{\mathsf{SE}}{m}} = \frac{\|r\|_2}{\sqrt{m}}.$$

Università della Svizzera italiana | Institute of Computing CI

Exercise 1: Inconsistent System and Least Squares

Consider the system of equations mentioned at the beginning of the lecture:

$$
\begin{aligned}
2x_1 - x_2 &= 3 \\
x_1 + 2x_2 &= 7 \\
2x_1 - x_2 &= 4
\end{aligned}
$$

Write it in matrix form as $A\mathbf{x} = \mathbf{b}$ and find the least-squares solution $\mathbf{x}^*$. Evaluate the quality of your approximation by computing the euclidean norm of the residual, the SE and the RMSE.

Exercise 1: Solution I

Step 1: Write the system in matrix form $A\mathbf{x} = \mathbf{b}$

Starting from the system reported above, the matrix form is given by:

$$\underbrace{\begin{bmatrix} 2 & -1 \\ 1 & 2 \\ 2 & -1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}}_{\mathbf{b}},$$

with $A \in \mathbb{R}^{3\times 2}$, $\mathbf{x} \in \mathbb{R}^2$ and $\mathbf{b} \in \mathbb{R}^3$. Then we have $n = 2$ unknowns in $m = 3$ equations (we will need to use this number later to compute the value of the RMSE).

Exercise 1: Solution II

Step 2: Compute the components A^TA and A^Tb of the system of normal equations

The matrix of the system of normal equations is given by:

$$A^TA = \begin{bmatrix} 2 & 1 & 2 \\ -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 2 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix},$$

while the right-hand side is given by:

$$A^Tb = \begin{bmatrix} 2 & 1 & 2 \\ -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} = \begin{bmatrix} 21 \\ 7 \end{bmatrix}.$$

Exercise 1: Solution III

Step 3: Solve the system of normal equations and compute the residual $\mathbf{r} = \mathbf{b} - A\mathbf{x}^*$

The system of normal equations is then given by:

$$\underbrace{\begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}}_{A^T A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 21 \\ 7 \end{bmatrix}}_{A^T \mathbf{b}},$$

The system can be solved, e.g., by Gaussian elimination. The triangular form would be:

$$\left[\begin{array}{cc|c} 9 & -2 & 21 \\ -2 & 6 & 7 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 9 & -2 & 21 \\ 0 & \frac{50}{9} & \frac{35}{3} \end{array}\right].$$

Exercise 1: Solution IV

Step 3: (continued)

The least-squares solution $\mathbf{x}^*$ can be computed from the triangular form above:

$$\mathbf{x}^* = \begin{bmatrix} \frac{14}{5} \\ \frac{21}{10} \end{bmatrix},$$

and the residual is given by the difference:

$$\mathbf{r} = \mathbf{b} - A\mathbf{x}^* = \begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 & -1 \\ 1 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} \frac{14}{5} \\ \frac{21}{10} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}.$$

Exercise 1: Solution V

Step 4: Compute the metrics on the residual

1. The **Euclidean norm** of the residual:

$$\|\mathbf{r}\|_2 = \sqrt{\left(-\frac{1}{2}\right)^2 + 0 + \left(\frac{1}{2}\right)^2} = \sqrt{\frac{1}{2}} = 0.7071.$$

2. The **SE**, or squared error:

$$\mathsf{SE} = \|\mathbf{r}\|_2^2 = 0.5.$$

3. The **RMSE**, or root-mean-square error:

$$\mathsf{RMSE} = \sqrt{\frac{\mathsf{SE}}{m}} = \sqrt{\frac{0.5}{3}} = 0.4082.$$

$\star$ Additional Exercises on Inconsistent Systems of Equations

Find the least squares solution and compute the metrics on the residual for these systems:

$$(a)\;\; A_1 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 2 & 1 \end{bmatrix}, \;\; \mathbf{b_1} = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}.$$

$$(b)\;\; A_2 = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}, \;\; \mathbf{b_2} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

$$(c)\;\; A_3 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}, \;\; \mathbf{b_3} = \begin{bmatrix} 2 \\ 3 \\ 1 \\ 2 \end{bmatrix}.$$

Least Squares Fitting of a Model to Data

- Sometimes we are given a set of data and we are interested in developing and testing a model which summarises the dataset in the best possible way.
- For example, we can have a collection of points in the plane $(x_1, y_1), \ldots, (x_m, y_m)$ and we can be interested in finding their best approximation, given a specific class of models. We could, e.g., consider a linear model as $y = \alpha_1 + \alpha_2 x$, where α_1 and α_2 represents the unknown coefficients that we need to calculate on the basis of our data points.
- Least squares can be considered an example of data compression: in other words, we want to find a model (with the smallest number possible of parameters) which fits the data points in the Euclidean norm. The model obtained can then be used to make predictions and explore scenarios beyond the data points we have collected with our measurements.

How to Fit Data by Least Squares?

Given a data set consisting of m data points $(x_1, y_1), \ldots, (x_m, y_m)$, we do the following:

1. **Model selection:** We pick the class of models which will be used to fit the data. An example could be given by the linear model $y = \alpha_1 + \alpha_2 x$ mentioned before.
2. **Write the system in the form $A\mathbf{x} = \mathbf{b}$:** We force the model to fit the data by inserting each of the m data points inside the model. This will result in m equations with n unknowns (in case of the linear model $y = \alpha_1 + \alpha_2 x$ we have that $n = 2$), that can be easily written down in matrix form as explained at the beginning of the lecture.
3. **Find the least-squares solution:** The least-squares solution can be found by solving the system of normal equations arising from the system $A\mathbf{x} = \mathbf{b}$.
4. **Evaluate the model:** Compute the Euclidean norm of the residual, the SE and the RMSE to evaluate the quality of the approximation given by the chosen model and, eventually, to compare it against another candidate model.

Exercise 2: Fitting Data by Least Squares (Linear Model)

Consider three points in $\mathbb{R}^2$: $(x_1, y_1) = (1, 2)$, $(x_2, y_2) = (-1, 1)$ and $(x_3, y_3) = (1, 3)$. Fit your data with a linear model by using least squares.

Exercise 2: Solution I

Step 1: Model selection

In order to fit the data by least squares, we need to pick a class of models.

Here, the exercise suggests us to use a linear model, which can be generally expressed as:

$$y = \alpha_1 + \alpha_2 x.$$

After choosing the model class, we now have to compute the values of α_1 and α_2 that better fit the data (i.e., the values for which the residual is minimal).

Exercise 2: Solution II

Step 2: Write the system in the form $A\mathbf{x} = \mathbf{b}$

We now force the linear model defined above to fit the data, by inserting the points into the model. Since we have three points, this will result in $m = 3$ equations in $n = 2$ unknowns, namely α_1 and α_2. The three equations considered are given by:

$$\begin{aligned} \alpha_1 + \alpha_2 &= 2, \\ \alpha_1 - \alpha_2 &= 1, \\ \alpha_1 + \alpha_2 &= 3. \end{aligned}$$

We can easily notice that the system computed above is inconsistent, since the first and the third equations have the same left hand side, but different right hand side.

Exercise 2: Solution III

Step 2: (continued)

Starting from the system of equations reported above, the matrix form is given by:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}}_{\mathbf{b}},$$

with $A \in \mathbb{R}^{3\times 2}$, $\boldsymbol{\alpha} \in \mathbb{R}^2$ and $\mathbf{b} \in \mathbb{R}^3$. Please notice that, in the current notation, $\boldsymbol{\alpha}$ replaces $\mathbf{x}$ as the vector of unknowns we want to find.

Exercise 2: Solution IV

Step 3: Find the least-squares solution

In order to find the least-squares solution, we compute the matrix of normal equations:

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix},$$

and the right hand side, which is given by:

$$A^T \mathbf{b} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}.$$

Exercise 2: Solution V

Step 3: (continued)

The system of normal equations is then given by:

$$\underbrace{\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}}_{A^T A} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} 6 \\ 4 \end{bmatrix}}_{A^T \mathbf{b}}.$$

The system above can be solved, e.g., by using once again Gaussian elimination.

$\star$ **Question:** What would the triangular form be in this case?

$$\left[\begin{array}{cc|c} 3 & 1 & 6 \\ 1 & 3 & 4 \end{array}\right] \longrightarrow \quad ?$$

Exercise 2: Solution VI

Step 3: (continued)

The system of normal equations is then given by:

$$\underbrace{\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}}_{A^T A} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} 6 \\ 4 \end{bmatrix}}_{A^T \mathbf{b}}.$$

The system above can be solved, e.g., by using once again Gaussian elimination.

⋆ **Answer:** In this case, the triangular form would be the following:

$$\left[\begin{array}{cc|c} 3 & 1 & 6 \\ 1 & 3 & 4 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 3 & 1 & 6 \\ 0 & \frac{8}{3} & 2 \end{array}\right]$$

Exercise 2: Solution VII

Step 3: (continued)

The least-squares solution $\boldsymbol{\alpha}^*$ can be computed from the triangular form above:

$$\boldsymbol{\alpha}^* = \begin{bmatrix} \frac{7}{4} \\ \frac{3}{4} \end{bmatrix},$$

and the residual is given by the difference:

$$\mathbf{r} = \mathbf{b} - A\boldsymbol{\alpha}^* = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{7}{4} \\ \frac{3}{4} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}.$$

Exercise 2: Solution VIII

Step 4: Evaluate the model

1. The **Euclidean norm** of the residual:

$$\|\mathbf{r}\|_2 = \sqrt{\left(-\frac{1}{2}\right)^2 + 0 + \left(\frac{1}{2}\right)^2} = \sqrt{\frac{1}{2}} = 0.7071.$$

2. The **SE**, or squared error:

$$\mathsf{SE} = \|\mathbf{r}\|_2^2 = 0.5.$$

3. The **RMSE**, or root-mean-square error:

$$\mathsf{RMSE} = \sqrt{\frac{0.5}{3}} = 0.4082.$$

Exercise 3: Fitting Data by LS (Quadratic Model)

Consider again the three points in $\mathbb{R}^2$: $(x_1, y_1) = (1, 2)$, $(x_2, y_2) = (-1, 1)$ and $(x_3, y_3) = (1, 3)$. Fit your data with a quadratic model by using least squares. Consider the metrics used to evaluate the model that we computed in Exercise 2 and compare them with the ones obtained in this case: which of the two models would you choose?

Exercise 3: Solution I

Step 1: Model selection

In order to fit the data by least squares, we need to pick a class of models.

In this case, the exercise suggests us to use a quadratic model, which can be expressed as:

$$y = \alpha_1 + \alpha_2 x + \alpha_3 x^2.$$

After choosing the model class, we now have to compute the values of α_1, α_2 and α_3 that better fit the data (i.e., the values for which the residual is minimal).

Exercise 3: Solution II

Step 2: Write the system in the form $A\mathbf{x} = \mathbf{b}$

We now force the linear model defined above to fit the data, by inserting the points into the model. Since we have three points, this will result in $m = 3$ equations in $n = 3$ unknowns, namely α_1, α_2 and α_3. The three equations considered are given by:

$$\begin{aligned} \alpha_1 + \alpha_2 + \alpha_3 &= 2, \\ \alpha_1 - \alpha_2 + \alpha_3 &= 1, \\ \alpha_1 + \alpha_2 + \alpha_3 &= 3. \end{aligned}$$

We can easily notice that the system computed above is inconsistent, since the first and the third equations have the same left hand side, but different right hand side.

Exercise 3: Solution III

Step 2: (continued)

Starting from the system of equations reported above, the matrix form is given by:

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}}_{\mathbf{b}},$$

with $A \in \mathbb{R}^{3\times 3}$, $\boldsymbol{\alpha} \in \mathbb{R}^3$ and $\mathbf{b} \in \mathbb{R}^3$. Matrix A is a rank 2 matrix, since the first and the third rows are linearly dependent, while vector $\boldsymbol{\alpha}$ has three components. This particular case where $m = n$ will have an impact on the least squares solution.

Exercise 3: Solution IV

Step 3: Find the least squares solution

In order to find the least squares solution, we compute the matrix of normal equations:

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{bmatrix},$$

and the right hand side, which is given by:

$$A^T \mathbf{b} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 6 \end{bmatrix}.$$

Exercise 3: Solution V

Step 3: (continued)

The system of normal equations is then given by:

$$\underbrace{\begin{bmatrix} 3 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{bmatrix}}_{A^T A} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} 6 \\ 4 \\ 6 \end{bmatrix}}_{A^T \mathbf{b}}.$$

Unlike what we had in the previous case, the matrix of the normal equations $A^T A$ is rank deficient (rank 2 matrix) and, thus, not invertible. The solution vector $\boldsymbol{\alpha}^*$ to the system of normal equations will then not be unique.

Exercise 3: Solution VI

Step 3: (continued)

The general solution of the least squares problem $\boldsymbol{\alpha}^*$ can be computed by dismissing the third equation (equal to the first) and by obtaining the conditions on the other components:

$$\boldsymbol{\alpha}^* = \begin{bmatrix} \beta \\ \frac{3}{4} \\ \frac{7}{4} - \beta \end{bmatrix}, \quad \text{with } \beta \in \mathbb{R},$$

and the residual is given by the difference:

$$\mathbf{r} = \mathbf{b} - A\boldsymbol{\alpha}^* = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta \\ \frac{3}{4} \\ \frac{7}{4} - \beta \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}.$$

Università della Svizzera italiana | Institute of Computing CI

Exercise 3: Solution VII

Step 4: Evaluate the model

1. The **Euclidean norm** of the residual:

$$\|\mathbf{r}\|_2 = \sqrt{\left(-\frac{1}{2}\right)^2 + 0 + \left(\frac{1}{2}\right)^2} = \sqrt{\frac{1}{2}} = 0.7071.$$

2. The **SE**, or squared error:

$$\mathsf{SE} = \|\mathbf{r}\|_2^2 = 0.5.$$

3. The **RMSE**, or root-mean-square error:

$$\mathsf{RMSE} = \sqrt{\frac{0.5}{3}} = 0.4082.$$

Exercise 3: Solution VIII

Step 5: Comparison of the two models

As we can notice by comparing the results obtained in Exercise 2 and Exercise 3, we get the same RMSE $= 0.4082$ in both cases. We can then dismiss the quadratic model in favour of the linear model, since by choosing the former we would not make our predictions more accurate, but simply increase the computational cost.

In this particular case (in which we end up with $n = m = 3$ for the quadratic model), considering an higher order polynomial is then not beneficial. This is due to the fact that we do not have enough points (i.e., information on the phenomenon we are trying to model) to improve our model by estimating more coefficients α_i.

Model Selection: Evaluating Other Options

- Linear and polynomials models are just one of the many classes of models available to fit the data when using least squares.
- Other classes of models are however present and depend on considerations on the characteristics of the particular data set under consideration.
- For example, **how would you model the mean temperature data in Switzerland**? We can all agree that data of this type has some kind of periodicity (e.g., we can consider a period of 24 hours or a period of 1 year). This kind of data can then be easily modelled by using a function involving a combination of sines and cosines.
- On the other hand, consider **a population of bacteria** which has just adapted to a new medium and starts replicating. The growth of the population (until the maximum carrying capacity of the environment where they reside is almost reached) can be modelled with an exponential, like $y = \alpha_1 e^{\alpha_2 x}$.

Exercise 4: Periodic Model

Consider the following periodic dataset:

x	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{5}{6}$
y	0	2	0	-1	1	1

and the following model:

$$F_1(x) = \alpha_1 + \alpha_2 \cos(2\pi x) + \alpha_3 \sin(2\pi x).$$

Fit the given data points using $F_1(x)$ and evaluate the model.

Exercise 4: Solution I

Step 1: Model selection

In this case the model to adopt to fit the data is suggested directly by the exercise.

We can notice that the model includes sines and cosines to tackle the periodicity of the data:

$$F_1(x) = \alpha_1 + \alpha_2 \cos(2\pi x) + \alpha_3 \sin(2\pi x)$$

We can see that, in $F_1(x)$, we have $n = 3$ unknown parameters to estimate, while $m = 6$, since we are considering 6 data points.

Exercise 4: Solution II

Step 2: Write the system in the form $A\mathbf{x} = \mathbf{b}$

We force our model to fit the data, thus obtaining the following system of 6 equations:

$$
\begin{aligned}
\alpha_1 + \alpha_2 \cos(0) + \alpha_3 \sin(0) &= 0, \\
\alpha_1 + \alpha_2 \cos\left(\frac{\pi}{3}\right) + \alpha_3 \sin\left(\frac{\pi}{3}\right) &= 2, \\
\alpha_1 + \alpha_2 \cos\left(\frac{2\pi}{3}\right) + \alpha_3 \sin\left(\frac{2\pi}{3}\right) &= 0, \\
\alpha_1 + \alpha_2 \cos(\pi) + \alpha_3 \sin(\pi) &= -1, \\
\alpha_1 + \alpha_2 \cos\left(\frac{4\pi}{3}\right) + \alpha_3 \sin\left(\frac{4\pi}{3}\right) &= 1, \\
\alpha_1 + \alpha_2 \cos\left(\frac{5\pi}{3}\right) + \alpha_3 \sin\left(\frac{5\pi}{3}\right) &= 1.
\end{aligned}
$$

Exercise 4: Solution III

Step 2: (continued)

The matrix form, with $A \in \mathbb{R}^{6\times3}$, $\alpha \in \mathbb{R}^3$ and $\mathbf{b} \in \mathbb{R}^6$, is given by:

$$\underbrace{\begin{bmatrix} 1 & 1 & 0 \\ 1 & \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 1 & -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 1 & \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} 0 \\ 2 \\ 0 \\ -1 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{b}}.$$

Exercise 4: Solution IV

Step 3: Find the least-squares solution

In order to find the least-squares solution, we compute the matrix of normal equations:

$$A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \frac{1}{2} & -\frac{1}{2} & -1 & -\frac{1}{2} & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & 0 & -\frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 1 & -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 1 & \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Exercise 4: Solution V

Step 3: (continued)

The right-hand side, instead, is given by:

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \frac{1}{2} & -\frac{1}{2} & -1 & -\frac{1}{2} & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & 0 & -\frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 0 \\ -1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}.$$

We can then use the two quantities derived above to write the system of normal equations.

Exercise 4: Solution VI

Step 3: (continued)

The system of normal equations is then given by:

$$\underbrace{\begin{bmatrix} 6 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}}_{A^T A} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}}_{A^T \mathbf{b}},$$

and the solution $\boldsymbol{\alpha}^*$, which can be straightforwardly computed, is:

$$\boldsymbol{\alpha}^* = \begin{bmatrix} \frac{1}{2} \\ \frac{2}{3} \\ 0 \end{bmatrix}.$$

Exercise 4: Solution VII

Step 3: (continued)

The residual can be easily computed as the difference:

$$\mathbf{r} = \mathbf{b} - A\boldsymbol{\alpha}^* = \begin{bmatrix} 0 \\ 2 \\ 0 \\ -1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 0 \\ 1 & \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 1 & -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 1 & \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{2}{3} \\ 0 \end{bmatrix} = \begin{bmatrix} -1.1667 \\ 1.1667 \\ -0.1667 \\ -0.8333 \\ 0.8333 \\ 0.1667 \end{bmatrix}.$$

Exercise 4: Solution VIII

Step 4: Evaluate the model

1. The **Euclidean norm** of the residual:

$$\|\mathbf{r}\|_2 = \sqrt{(-1.17)^2 + (1.17)^2 + (-0.17)^2 + (-0.83)^2 + (0.83)^2 + (0.17)^2} = 2.0412.$$

2. The **SE**, or squared error:

$$\mathsf{SE} = \|\mathbf{r}\|_2^2 = 4.1667.$$

3. The **RMSE**, or root-mean-square error:

$$\mathsf{RMSE} = \sqrt{\frac{4.1667}{6}} = 0.8333.$$

⋆ Exercise 5: An Alternative Periodic Model

Consider again the periodic data set of Exercise 4:

x	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{5}{6}$
y	0	2	0	-1	1	1

and the following alternative model:

$$F_2(x) = \alpha_1 + \alpha_2 \cos(2\pi x) + \alpha_3 \sin(2\pi x) + \alpha_4 \cos(4\pi x).$$

Compute the RMSE and compare it with the one of the model computed before:

$$F_1(x) = \frac{1}{2} + \frac{2}{3} \cos(2\pi x).$$

Which model would you choose?

Data Linearization

Sometimes we are dealing with models with nonlinear coefficients, like the **exponential** model:

$$y = \alpha_1 e^{\alpha_2 x},$$

or the **power law** model:

$$y = \alpha_1 x^{\alpha_2}.$$

Both cases could be solved through the mean of nonlinear least squares, but—in this case—we adopt an alternative approach, which involves the linearization of the model itself. If we apply the logarithm on both sides, we obtain the following new models:

$$\begin{aligned} \ln(y) &= k_1 + \alpha_2 x, \quad k_1 = \ln(\alpha_1), \\ \ln(y) &= k_1 + \alpha_2 \ln(x), \quad k_1 = \ln(\alpha_1). \end{aligned}$$

Exercise 6: Data Linearization of Exponential Model

You are given the following dataset on the measurements of a response variable:

x	1950	1955	1960	1965	1970	1975	1980
y	53.05	73.04	98.31	139.78	193.48	260.20	320.39

with x representing the year and y the value at that specific time. Use data linearization to find the least squares best fit to a simple exponential model, expressed as:

$$y = \alpha_1 e^{\alpha_2 x},$$

and compute the RMSE both of the log-linearized model and of the original exponential model.

Exercise 6: Solution I

Step 1: Model selection

In order to fit the data by least squares, we consider the exponential model:

$$y = \alpha_1 e^{\alpha_2 x},$$

which can be linearized by applying the logarithm on both sides:

$$\ln(y) = k_1 + \alpha_2 x, \quad k_1 = \ln(\alpha_1).$$

We will now work with the linearized model to perform the least squares computation and, afterwards, we will map back our results to the exponential model.

Exercise 6: Solution II

Step 2: Write the system in the form $A\mathbf{x} = \mathbf{b}$

We force the model to fit the data and we obtain $m = 7$ equations in $n = 2$ unknowns:

$$
\begin{aligned}
k_1 + 1950\alpha_2 &= \ln(53.05),\\
k_1 + 1955\alpha_2 &= \ln(73.04),\\
k_1 + 1960\alpha_2 &= \ln(98.31),\\
k_1 + 1965\alpha_2 &= \ln(139.78),\\
k_1 + 1970\alpha_2 &= \ln(193.48),\\
k_1 + 1975\alpha_2 &= \ln(260.20),\\
k_1 + 1980\alpha_2 &= \ln(320.39).
\end{aligned}
$$

Exercise 6: Solution III

Step 2: (continued)

Starting from the system of equations reported above, the matrix form is given by:

$$\underbrace{\begin{bmatrix} 1 & 1950 \\ 1 & 1955 \\ 1 & 1960 \\ 1 & 1965 \\ 1 & 1970 \\ 1 & 1975 \\ 1 & 1980 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} k_1 \\ \alpha_2 \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} \ln(53.05) \\ \ln(73.04) \\ \ln(98.31) \\ \ln(139.78) \\ \ln(193.48) \\ \ln(260.20) \\ \ln(320.39) \end{bmatrix}}_{\mathbf{b}} .$$

Exercise 6: Solution IV

Step 3: Find the least squares solution

The solution of the least squares problem $\boldsymbol{\alpha}^*$ computed by solving the normal equations is:

$$\boldsymbol{\alpha}^* = \begin{bmatrix} k_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 3.9896 \\ 0.0615 \end{bmatrix}.$$

The log-linearized model is then given by the following equation:

$$\ln(y) = 3.9896 + 0.0615x,$$

while the original exponential model, considering that $\alpha_1 = e^{k_1} = 54.03$, is given by:

$$y = 54.03e^{0.0615x}.$$

Exercise 6: Solution V

Step 4: Computation of the RMSE in the two models

⋆ **Additional exercise:** Compute the RMSE of the log-linearized model and of the original exponential model. The steps to do it are simply:

1. Compute the residual of the log-linearized model.
2. Compute the residual of the original exponential model.
3. Compute the RMSE for the models in steps (1) and (2).[1]

[1] **Solution:** You should obtain that RMSE ≈ 0.0357 in case (1) and RMSE ≈ 9.56 in case (2).

$\star$ Exercise 7: Linearization of Power Law Model

You are given the following dataset on the measurements of a response variable:

x	1	1	2	3	5
y	2	4	5	6	10

Use data linearization to find the least squares best fit to the power law model, expressed as:

$$y = \alpha_1 x^{\alpha_2},$$

and compute the RMSE of both the log-linearized model and of the original power law model.