Mathematical modelling, part 1

Neža Mramor Kosta

Faculty of Computer and Information Science
University of Ljubljana

2019/2020
Introduction

The task of mathematical modelling is to find and evaluate solutions to real world problems with the use of mathematical concepts and tools.

In this course we will introduce some (by far not all) mathematical tools that are used in setting up and solving mathematical models.

We will (together) also solve specific problems, study examples and work on projects.
Contents

- Introduction
- Linear models: systems of linear equations, matrix inverses, SVD decomposition, PCA
- Nonlinear models: vector functions, linear approximation, solving systems of nonlinear equations
- Geometric models: curves and surfaces
- Dynamical models: differential equations, dynamical systems
Modelling cycle

- Real world problem
- Simplification
- Idealization
- Generalization
- Solution
- Mathematical model
- Conclusions
- Simulation
- Program
- Computer solution
- Explanation
What should we pay attention to?

- **Simplification**: relevant assumptions of the model (distinguish important features from irrelevant)
- **Generalization**: choice of mathematical representations and tools (for example: how to represent an object - as a point, a geometric shape, …)
- **Solution**: as simple as possible and well documented
- **Conclusions**: are the results within the expected range, do they correspond to "facts" and experimental results?

A mathematical model is not universal, it is an approximation of the real world that works only within a certain scale where the assumptions are at least approximately realistic.
An object (ball) with mass $m$ is thrown vertically into the air. What should we pay attention to when modelling its motion?

- The assumptions of the model: relevant forces and parameters (gravitation, friction, wind, ...), how to model the object (a point, a homogeneous or nonhomeogeneous geometric object, angle and rotation in the initial thrust, ...)
- Choice of mathematical model: differential equation, discrete model, ...
- Computation: analitic or numeric, choice of method, ...
- Do the results make sense?
Errors

An important part of modelling is estimating the errors!

Errors are an integral part of every model.

Errors come from: assumptions of the model, imprecise data, mistakes in the model, computational precision, errors in numerical and computational methods, mistakes in the computations, mistakes in the programs, . . .

**Absolute error** = Approximate value - Correct value

\[ \Delta x = \bar{x} - x \]

**Relative error** = \( \frac{\text{Absolute error}}{\text{Correct value}} \)

\[ \delta_x = \frac{\Delta x}{x} \]
Keep in mind:

- Real numbers are given only up to some precision.
- The solution to the equation $f(x) = 0$ is any number $x$ with $|f(x)| < \varepsilon$.
- The tolerance $\varepsilon$ (precision of data and computation) should be known.
Example: quadratic equation

$$x^2 + 2px - q = 0$$

Analytic solutions are

$$x_1 = -p - \sqrt{p^2 + q} \quad \text{and} \quad x_2 = -p + \sqrt{p^2 + q}.$$ 

What happens if

- $p = 10000$, $q = 1$?
- $p = -\frac{\varepsilon}{2} - 1$, $q = -1 - \varepsilon$, $\varepsilon = 10^{-8}$?
1. Linear mathematical models

Given are points \{ (x_1, y_1), \ldots, (x_m, y_m) \}, \; x_i \in \mathbb{R}^n, \; y_i \in \mathbb{R},

the task is to find a function \( F(x, a_1, \ldots, a_p) \) that is a good fit for the data.

The values of the parameters \( a_1, \ldots, a_p \) should be chosen so that the equations

\[
y_i = F(x, a_1, \ldots, a_p), \; i = 1, \ldots, m,
\]

are satisfied, or that the error is as small as possible.

*Least squares method:* the parameters are determined so that the sum of squared errors

\[
\sum_{i=1}^{m} (F(x_i, a_1, \ldots, a_p) - y_i)^2
\]

is as small as possible.
The mathematical model is \textit{linear}, when the function $F$ is a linear function of the parameters:

$$F(x, a_1, \ldots, a_p) = a_1 \varphi_1(x) + \varphi_2(x) + \cdots + a_p \varphi_p(x),$$

where $\varphi_1, \varphi_2, \ldots \varphi_p$ are functions of a specific type.

Examples of linear models:

1. \textit{linear regression}: $x, y \in \mathbb{R}$, $\varphi_1(x) = 1$, $\varphi_2(x) = x$
2. \textit{polinomial regression}: $x, y \in \mathbb{R}$, $\varphi_1(x) = 1, \ldots, \varphi_p(x) = x^{p-1}$
3. \textit{multivariate linear regression}: $x \in \mathbb{R}^n$, $y \in \mathbb{R}$, $\varphi_1(x) = 1$, $\varphi_2(x) = x_1, \ldots$, $\varphi_n(x) = x_n$
4. \textit{frequency} or \textit{spectral analysis}:
   \footnotesize{$\varphi_1(x) = 1$, $\varphi_2(x) = \cos \omega x$, $\varphi_3(x) = \sin \omega x$, $\varphi_4(x) = \cos 2\omega x$, \ldots} (there can be infinitely many functions $\varphi_i(x)$ in this case)

Examples of nonlinear models: $F(x, a, b) = ae^{bx}$ and

$$F(x, a, b, c) = \frac{a + bx}{c + x}.$$
Given the data points \( \{(x_1, y_1), \ldots, (x_m, y_m)\} \), \( x_i \in \mathbb{R}^n \), \( y_i \in \mathbb{R} \) the parameters of a linear model

\[
y = a_1 \varphi_1(x) + a_2 \varphi_2(x) + \cdots + a_p \varphi_p(x)
\]

should satisfy the system of linear equations

\[
y_i = a_1 \varphi_1(x_i) + a_2 \varphi_2(x_i) + \cdots + a_p \varphi_p(x_i), \ i = 1, \ldots, m,
\]

or, in matrix form,

\[
\begin{bmatrix}
\varphi_1(x_1) & \varphi_2(x_1) & \cdots & \varphi_p(x_1) \\
\varphi_1(x_2) & \varphi_2(x_2) & \cdots & \varphi_p(x_2) \\
\vdots & \vdots & \ddots & \vdots \\
\varphi_1(x_m) & \varphi_2(x_m) & \cdots & \varphi_p(x_m)
\end{bmatrix}
\begin{bmatrix}
a_1 \\
a_1 \\
\vdots \\
a_p
\end{bmatrix}
=
\begin{bmatrix}
y_1 \\
y_1 \\
\vdots \\
y_p
\end{bmatrix}.
\]
Given is a system of linear equations

\[ Ax = b. \]

\( A \) is the *matrix of coefficients* of order \( m \times n \) where \( m \) is the number of equations and \( n \) is the number of unknowns, \( x \) is the *vector of unknowns* and \( b \) is the *right side vector*. 
Existence of solutions:

Let \( A = [a_1, \ldots, a_n] \), where \( a_i \) are vectors representing the columns of \( A \).

For any vector \( x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \) the product \( Ax \) is a linear combination

\[
Ax = \sum_{i} x_i a_i.
\]

The system is solvable if and only if the vector \( b \) can be expressed as a linear combination of the columns of \( A \), that is, it is in the column space of \( A \), \( b \in C(A) \).
By adding \( b \) to the columns of \( A \) we obtain the extended matrix of the system

\[
[A \mid b] = [a_1, \ldots, a_n \mid b],
\]

**Theorem (0)**

The system \( Ax = b \) is solvable if and only if the rank of \( A \) equals the rank of the extended matrix \([A \mid b]\):

\[
\text{rank } A = \text{rank } [A \mid b] = r.
\]

The solution is unique if the rank of the two matrices equals the number of unknowns: \( r = n \).

An especially nice case is the following:

If \( A \) is a square matrix \((n = m)\) that has an inverse matrix \( A^{-1} \), the system has a unique solution

\[
x = A^{-1} b.
\]
The following four conditions are equivalent and all describe invertible matrices:

- The matrix $A$ has an inverse.
- The rank of $A$ equals $n$.
- $\det(A) \neq 0$.
- The null space $N(A) = \{x : Ax = 0\}$ is trivial.
- All eigenvalues of $A$ are nonzero.
- For each $b$ the system of equations $Ax = b$ has precisely one solution.

We also say that $A$ is invertible or nonsingular.
A square matrix that does not satisfy the above conditions does not have an inverse.

Example

\[
A = \begin{bmatrix}
1 & 0 & 1 \\
0 & 1 & -1 \\
1 & 1 & 1 \\
\end{bmatrix}, \quad B = \begin{bmatrix}
1 & 0 & 1 \\
0 & 1 & -1 \\
1 & 1 & 0 \\
\end{bmatrix}
\]

\(A\) is invertible and is of rank 3, \(B\) is not invertible and is of rank 2.

Check the other conditions from above . . .

For a rectangular matrix \(A\) of dimension \(m \times n\), \(m \neq n\), its inverse is not defined.
Definition
A *generalized inverse* of a matrix $A$ is a matrix $G$ such that

$$AGA = A.$$ 

If $A$ is invertible, it has a unique generalize inverse, which is equal to $A^{-1}$.

Theorem
Every matrix has a generalized inverse.
Proof: An algorithm for computing a generalized inverse of $A$ goes as follows.

Let $r$ be the rank of $A$.

1. Find any nonsingular submatrix $M$ in $A$ of order $r \times r$,
2. in $A$ substitute
   - elements of the submatrix $M$ for corresponding elements of $(M^{-1})^T$,
   - all other elements with 0,
3. the transpose of the obtained matrix is a generalized inverse $G$.

If $A$ is a square matrix of full rank, the algorithm above returns $A^{-1}$. 
Example

Compute at least one generalized inverse of

\[
\begin{bmatrix}
0 & 0 & 2 & 0 \\
0 & 0 & 1 & 0 \\
2 & 0 & 1 & 4
\end{bmatrix}.
\]
Generalized inverses of a matrix play a similar role as the inverse (when it exists) in solving systems of linear equations.

**Theorem**

*If the system $Ax = b$ is solvable (that is, $b \in \mathcal{C}(A)$) and $G$ is a generalized inverse of $A$, then $x = Gb$ is a solution of the system.*

Moreover, for any vector $z \in \mathbb{R}^m$ the vector 

$$\tilde{x} = Gb + (GA - I)z$$

is a solution of the system and every solution is of this form for some vector $z \in \mathbb{R}^m$. 
Proof. Let \( A = [a_1, a_2, \ldots a_n] \), where \( a_i \) are column vectors of \( A \) and let \( G \) be a generalized inverse of \( A \).

From \( AGA = A \) it follows that \( AG(a_i) = a_i \) for every \( a_i \) and

\[
Ax = A(Gb) = AG(\sum \alpha_i a_i) = \sum \alpha_i AGa_i = \sum \alpha_i a_i = b.
\]

Since for any \( z \in \mathbb{R}^n \)

\[
A(GA - I)z = (AGA - A)z = 0,
\]

it follows that \( \{(GA - I)z; z \in \mathbb{R}^m\} \) is precisely the null space of the matrix \( A \), so every solution is of the form

\[
\tilde{x} = x + (GA - I)z, \quad z \in \mathbb{R}^n.
\]
Example:

Find all solutions of the system

\[ Ax = b, \]

where \( A = \begin{bmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 0 & 1 & 4 \end{bmatrix} \) and \( b = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} \).
1.2 The Moore-Penrose generalized inverse

Among all the generalized inverses of a singular matrix $A$, one has especially nice properties.

Definition

The Moore-Penrose generalized inverse, or shortly the MP inverse of $A$ is the unique matrix $A^+$ satisfying the following four conditions:

1. $A^+$ is a generalized inverse of $A$: $AA^+A = A$,
2. $A$ is a generalized inverse of $A^+$: $A^+AA^+ = A^+$,
3. the square matrix $AA^+$ is symmetric: $(AA^+)^T = AA^+$,
4. also the square matrix $AA^+$ is symmetric: $(A^+A)^T = A^+A$. 
Theorem
The MP inverse $A^+$ of a matrix $A$ is unique.

Proof: Assume that there are two matrices $G_1$ and $G_2$ that satisfy the four conditions above.

Then,

$$
AG_1 = (AG_2 A)G_1 = (AG_2)(AG_1) = (AG_2)^T(AG_1)^T = G_2^T(AG_1A)^T = G_2^T A^T = AG_2
$$

A similar argument involving properties (2) and (4) shows that

$$
G_1 A = G_2 A,
$$

and so

$$
G_1 = G_1 AG_1 = G_1 AG_2 = G_2 AG_2 = G_2.
$$
If $A$ is a square invertible matrix, then its only generalized inverse is $A^+ = A^{-1}$.

Here are two properties of $A^+$ that are easy to check:

1. $(A^+)^+ = A$,
2. $(A^T)^+ = (A^+)^T$.

In the rest of this chapter we will be interested in two obvious questions:

- How do we compute $A^+$ for a general singular matrix?
- Why would we want to compute it?

To answer the first question, we will begin by three special cases.
Construction of the MP inverse of $A$:

Case 1: $A^T A$ is an invertible matrix

In this case $A^+ = (A^T A)^{-1} A^T$.

To see this, we have to show that the matrix $G = (A^T A)^{-1} A^T$ satisfies properties (1) to (4):

1. $AGA = A((A^T A)^{-1} A^T)A = A(A^T A)^{-1}(A^T A) = A$
2. $GAG = ((A^T A)^{-1} A^T)A((A^T A)^{-1} A^T) = (A^T A)^{-1} A^T$
3. $(AG)^T = (A(A^T A)^{-1} A^T)^T = A(A^T A)^{-1})^T A^T = A((A^T A)^{-1} A) = AG$
4. ...
Case 2: \( AA^T \) is an invertible matrix

In this case \( A^T \) satisfies the condition for Case 1, so \( (A^T)^+ = (AA^T)^{-1} A \).

Since \( (A^T)^+ = (A^+)^T \) it follows that

\[
A^+ = ((AA^T)^{-1} A)^T = A^T (AA^T)^{-1}.
\]
Case 3: $\Sigma$ is a diagonal $m \times n$ matrix of the form

$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \\ & & & \sigma_m \end{bmatrix},$$

then its MP inverse is

$$\Sigma^+ = \begin{bmatrix} \sigma_1^+ & & \\ & \sigma_2^+ & \\ & & \ddots \\ & & & \sigma_m^+ \end{bmatrix},$$

where $\sigma_i^+ = \begin{cases} \frac{1}{\sigma_i}, & \sigma_i \neq 0, \\ 0, & \sigma_i = 0. \end{cases}$
Case 4: a general matrix $A$

We will use the *singular value decomposition* or *SVD* of $A$.

**Theorem (Singular value decomposition)**

Every $m \times n$ matrix $A$ can be expressed as a product $A = U\Sigma V^T$

- where $U$ is an orthogonal $m \times m$ matrix with left singular vectors $u_i$ in its columns,
- $V$ is an orthogonal $n \times n$ matrix with right singular vectors $v_i$ in its columns,
- $\Sigma = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix}$ is a diagonal $m \times n$ matrix

with singular values of $A$ $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ on the diagonal.
Computing SVD:

If $A = U\Sigma V^T$, then

$$A^T A = (V\Sigma^T U^T)(U\Sigma V^T) = V\Sigma^T \Sigma V^T$$

$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma \Sigma^T U^T$$

It follows that

- $\Sigma^T \Sigma = \Sigma^2$ is the diagonal matrix with eigenvalues $\lambda_i$ of $A^T A$ or, equivalently, $AA^T$ on its diagonal, so the singular values $\sigma_i$ of $A$ are square roots $\sigma_i = \sqrt{\lambda_i}$, $\lambda_i > 0$,

- $U$ has the corresponding eigenvectors (normalized and pairwise orthogonal) of $AA^T$ as its columns, so the left singular vectors are eigenvectors corresponding to nonzero eigenvalues of $AA^T$,

- $V$ has the corresponding eigenvectors (normalized and pairwise orthogonal) of $A^T A$ as its columns, so the right singular vectors are eigenvectors corresponding to nonzero eigenvalues of $A^T A$. 
In addition, for the left singular vector $u_i$ corresponding to $\sigma_i = \sqrt{\lambda_i}$,

$$(A^T A)(A^T u_i) = A^T (AA^T) u_i = A^T (\lambda_i u_i) = \lambda_i (A^T u_i).$$

It follows that

$\blacktriangleright \ A^T u_i$ is an eigenvector of $(A^T A)$ corresponding to $\lambda_i$, so

$\blacktriangleright \ v_i = \frac{A^T u_i}{\|A^T u_i\|}$ is a right singular vector corresponding to $\sigma_i = \sqrt{\lambda_i}$,

Similarly, for a right singular vector $v_i$ corresponding to $\sigma_i$, the vector $\frac{A v_i}{\|A v_i\|}$ is a left singular vector corresponding to $\sigma_i$. 
Algorithm for SVD

- Compute the eigenvalues and an orthonormal basis consisting of eigenvectors of the symmetric matrix $A^T A$ or $AA^T$ (depending on which is of smaller size),
- The singular values of the matrix $A^{m \times n}$ are equal to $\sigma_i = \sqrt{\lambda_i}$, where $\lambda_i$ are the nonzero eigenvalues of the symmetric matrices $(A^T A)^{n \times n}$ and $(AA^T)^{m \times m}$.
- The left singular vectors (i.e. the first $r$ columns of $U$) are the corresponding orthonormal eigenvectors of $AA^T$.
- The right singular vectors (i.e. the first $r$ columns of $V$) are the corresponding orthonormal eigenvectors of $A^T A$.
- If $u$ (resp. $v$) is a left (resp. right) singular vector corresponding to the singular value $\sigma_i$, then $v = Au$ (resp. $u = A^T v$) is a right (resp. left) singular vector corresponding to the same singular value.
- The remaining columns of $U$ and $V$ consist of an orthonormal basis of the null-space (i.e. the eigenspace of $\lambda = 0$) of $AA^T$ or $A^T A$, respectively.
General algorithm for $A^+$

Assume $m \geq n$ (if not, exchange $A^T A$ for $AA^T$).

1. Compute the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ and the corresponding orthonormal eigenvectors $v_1, v_2, \ldots, v_r$ of $A^T A$.

2. Let $\Sigma$ be the $m \times n$ diagonal matrix with $\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \cdots \geq \sqrt{\lambda_n}$ in the upper left corner of the diagonal.

3. Let $V = \begin{bmatrix} v_1 & v_2 & \ldots & v_r & \ldots & v_n \end{bmatrix}$ (where the eigenvectors $v_{r+1}, \ldots, v_n$ corresponding to the eigenvalue 0 do not need to be computed).

4. Let $U$ be the matrix of normalized columns of $AV$.

5. Compute $A^+ = V \Sigma^+ U^T$. 
An application of SVD: *principal component analysis* or *PCA*

PCA is a very well known and efficient method for data compression, dimension reduction, . . .

Due to its importance in different fields, it has many other names: discrete Karhunen-Loève transform (KLT), Hotelling transform, empirical orthogonal functions (EOF), . . .

Let \( \{X_1, \ldots, X_m\} \) be a sample of vectors from \( \mathbb{R}^n \).

In applications, often \( m \ll n \), where \( n \) is very large, for example, \( X_1, \ldots, X_m \) can be
- vectors of gene expressions in \( m \) tissue samples or
- vectors of grayscale in images
- bag of words vectors, with components corresponding to the numbers of certain words from some dictionary in specific texts, . . . ,

or \( n \ll m \) for example if the data represents a point cloud in a low dimensional space \( \mathbb{R}^n \) (for example in the plane).
We will assume that $m << n$.

Also assume that the data is *centralized*, i.e. the centroid is in the origin

\[ \mu = \frac{1}{m} \sum_{i=1}^{m} X_i = 0 \in \mathbb{R}^n. \]

If not, we subtract $\mu$ from all vectors in the data set.
Let $X = [X_1 \ X_2 \ldots \ X_m]^T$ be the matrix of dimension $m \times n$ with data in the rows.

Let $(X^T X)^{m \times m}$ and $(XX^T)^{n \times n}$ be the covariance matrices of the data.

- The **principal values** of the data set $\{X_1, \ldots, X_r\}$ are the nonzero eigenvalues $\lambda_i = \sigma_i^2$ of the covariance matrices (where $\sigma_i$ are the singular values of $X$).

- The **principal directions** in $\mathbb{R}^n$ are corresponding eigenvectors $v_1, \ldots, v_r$, i.e. the columns of the matrix $V$ from the SVD of $X$. The remaining columns of $V$ (i.e. the eigenvectors corresponding to 0) form a basis of the null space of $X$.

- The first column $v_1$, **the first principal direction**, corresponds to the direction in $\mathbb{R}^n$ with the largest variance in the data $X_i$, that is, the most informative direction for the data set, the second the second most important, ... 

- The **principal directions** in $\mathbb{R}^m$ are the columns $u_1, \ldots, u_r$ of the matrix $U$ and represent the coefficients in the linear decomposition of the vectors $X_1, \ldots, X_m$ along the orthonormal basis $v_1, \ldots v_n$ of $\mathbb{R}^n$. 
PCA provides a linear dimension reduction method based on a projection of the data from the space \( \mathbb{R}^n \) into a lower dimensional subspace spanned by the first few principal vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_k \) in \( \mathbb{R}^n \).

The idea is to approximate

\[
X_i = u_{1,i} \mathbf{v}_1 + \cdots + u_{k,i} \mathbf{v}_k + \cdots + u_{m,i} \mathbf{v}_m \approx u_{1,i} \mathbf{v}_1 + \cdots + u_{k,i} \mathbf{v}_k
\]

with the first \( k \) most informative directions in \( \mathbb{R}^n \) and suppress the last \( m - k \).

PCA has the following amazing property:

**Theorem**

Among all possible projections of \( p: \mathbb{R}^n \to \mathbb{R}^k \) onto a \( k \)-dimensional subspace, PCA provides the best in the sense that the error

\[
\sum_i \|X_i - p(X_i)\|^2
\]

is the smallest possible.
1.3 The MP inverse and systems of linear equations

A system of equations $A^{m\times n}x = b$ that has more variables than constraints, that is if \( n > m \) typically has infinitely many solutions, but it may happen that it has no solutions. Such a system is underdetermined.

Theorem

1. An underdetermined system of linear equations

\[ Ax = b \]  \hspace{1cm} (1)

is solvable if and only if $AA^+ b = b$.

2. If there are infinitely many solutions, the solution $A^+ b$ is the one with the smallest norm:

\[ \|A^+ b\| = \min\{\|x\|; Ax = b\}. \]
In the proof we will need the following

**Lemma**

*The vector $A^+b$ is orthogonal to the null space*

$$N(A) = \{ x \in \mathbb{R}^n; Ax = 0 \} = \{ (A^+A - I)z; z \in \mathbb{R}^m \}$$

of $A$.

Proof of the Lemma: we need to show that the inner product of $A^+b$ with any vector from the nullspace, that is, any vector of the form $(A^+A - I)z$, $z \in \mathbb{R}^n$ is 0. So, using the second and fourth property of the MP inverse,

$$A^+b \cdot (A^+A - I)z = ((A^+A - I)z)^T A^+b = z^T ((A^+A)^T A^+ - A^+)b = 0$$
Proof of the Theorem:

1. The first claim follows directly from the fact that $A^+$ is a generalized inverse of $A$.

2. Since every solution of the system is of the form $x = A^+ b + (A^+ A - I)z$, $z \in \mathbb{R}^n$, then, since by the lemma $b \cdot (A^+ A - I)z = 0$, for any solution $x$,

$$
\|x\| = \|A^+ b + (A^+ A - I)z\|^2 \\
= (A^+ b + (A^+ A - I)z) \cdot (A^+ b + (A^+ A - I)z) \\
= \|A^+ b\|^2 + 2A^+ b \cdot (A^+ A - I)z + \|(A^+ A - I)z\|^2 \\
= \|A^+ b\|^2 + \|(A^+ A - I)z\|^2 \\
\geq \|A^+ b\|^2,
$$

so the smallest values is achieved at $z = 0$ and $x = A^+ b$. 
Example:

Find the point on the plane $3x + y + z = 2$ closest to the origin.

In this case,

$$A^{1\times 3} = \begin{bmatrix} 3 & 1 & 1 \end{bmatrix} \quad \text{and} \quad b = [2].$$

Let us find the SVD of $A$:

The matrix $(AA^T)^{1\times 1} = [11]$ has its only eigenvalue $\lambda = 11$ and eigenvector $u = [1]$, so

$$U^{1\times 1} = [1] \quad \text{and} \quad \Sigma^{1\times 3} = \begin{bmatrix} \sqrt{11} & 0 & 0 \end{bmatrix}.$$
We only need the first column $v_1$ of $V$ (since there is only one nonzero eigenvalue): This is the normalized eigenvector corresponding to the eigenvalue 11 and is obtained by normalizing

$$A^T u = \begin{bmatrix} 3 & 1 & 1 \end{bmatrix}^T [1] = \begin{bmatrix} 3 & 1 & 1 \end{bmatrix}^T,$$

so

$$v_1 = \frac{1}{11} \begin{bmatrix} 3 & 1 & 1 \end{bmatrix}^T.$$

The MP inverse is then

$$A^+ = V \Sigma^+ U^T = \begin{bmatrix} 3 & \cdots & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{11} \\ 0 \\ 0 \\ \frac{1}{11} \\ \frac{1}{11} \end{bmatrix} [1] = \begin{bmatrix} \frac{3}{11} \\ \frac{11}{11} \\ \frac{11}{11} \end{bmatrix},$$

and the solution is

$$x^+ = A^+ b = \begin{bmatrix} \frac{6}{11} \\ \frac{11}{11} \\ \frac{11}{11} \end{bmatrix}.$$
The system $Ax = b$ is **overdetermined** $m > n$, that is, there are more constraints than variables. Such a system typically has no solutions, but it might have one or even infinitely many solutions.

**Least squares approximation problem**: if the system $Ax = b$ has no solutions, then a best fit for the solution is a vector $x$ such that the error $\|Ax - b\|$ or, equivalently its square

$$\|Ax - b\|^2 = \sum_{i=1}^{m} (a_i x - b_i)^2$$

is the smallest possible.

**Theorem**

*If the system $Ax = b$ has no solutions, than $x^+ = A^+ b$ is the solution to the least squares approximation problem, that is*

$$\|Ax^+ - b\| = \min \{\|Ax - b\|, x \in \mathbb{R}^n\}.$$
Proof:
The closest vector to $b$ in the column space $C(A) = \{Ax, x \in \mathbb{R}^n\}$ of $A$ is the orthogonal projection of $b$ onto $C(A)$.

So, we will show that $Ax^+$ is the orthogonal projection of $b$ onto $\{Ax, x \in \mathbb{R}^n\}$, that is, the vector $(Ax^+ - b)$ is orthogonal to any vector $Ax$, $x \in \mathbb{R}^n$.

$$(Ax^+ - b) \cdot Ax = x^T A^T (AA^+ b - b) =$$
$$= x^T (A^T (AA^+)^T b - b) = x^T ((AA^+ A)^T - A^T) b = 0.$$  

The first equality is the inner product of vectors expressed as a product of matrices, the second equality follows from property (3) of the MP inverse, and the third from property (1) of the MP inverse.
Example: linear regression

Given points \( \{(x_1, y_1), \ldots, (x_m, y_m)\} \) in the plane, we are looking for the line \( ax + b = y \) which is the least squares best fit.

If \( m > 2 \), we obtain an overdetermined system

\[
\begin{bmatrix}
  x_1 & 1 \\
  \vdots & \vdots \\
  x_m & 1 \\
\end{bmatrix}
\begin{bmatrix}
a \\
b \\
\end{bmatrix}
\begin{bmatrix}
y_1 \\
\vdots \\
y_m \\
\end{bmatrix}
\]

The solution of the least squares approximation problem is given by

\[
\begin{bmatrix}
a^+ \\
b^+ \\
\end{bmatrix}
= A^+
\begin{bmatrix}
y_1 \\
\vdots \\
y_m \\
\end{bmatrix}
\]

The line \( y = a^+x + b^+ \) in the regression line.
3. Nonlinear models

General formulation

Given is a sample of points \( \{(x_1, y_1), \ldots, (x_m, y_m)\} \), \( x_i \in \mathbb{R}^n, y_i \in \mathbb{R} \).

The mathematical model is \textit{nonlinear} if the function

\[ y = F(x, a_1, \ldots, a_p) \]  \hspace{1cm} (2)

is a nonlinear function of the parameters \( a_i \). This means it cannot be written in the form

\[ y = a_1 f_1(x) + a_2 f_2(x) + \ldots + a_p f_p(x), \]

where each \( f_i : \mathbb{R}^n \rightarrow \mathbb{R} \) is some function.

Plugging each data point into (2) we obtain a system of nonlinear equations

\[ y_1 = F(x_1, a_1, \ldots, a_p), \]

\[ \vdots \]

\[ y_m = F(x_m, a_1, \ldots, a_p), \]

for the parameters \( a_1, \ldots, a_p \in \mathbb{R} \).
Examples

1. **Exponential decay or growth**: $F(x, a, k) = ae^{kx}$, $a$ and $k$ are parameters.

   A quantity $y$ changes at a rate proportional to its current value, which can be described by the differential equation
   \[
   \frac{dy}{dx} = ky.
   \]

   The solution to this equation (use separation of variables) is $y = F(x, a, x)$. 

   ![Graph of exponential functions](image.png)
2. **Gaussian model**: \( F(x, a, b, c) = ae^{-\left(\frac{x-b}{c}\right)^2} \), \( a, b, c \in \mathbb{R} \) parameters.

\( a \) is the value of the maximum obtained at \( x = b \) and \( c \) determines the width of the curve.

It is used in statistics to describe the normal distribution, but also in signal and image processing.

In statistics \( a = \frac{1}{\sigma \sqrt{2\pi}} \), \( b = \mu \), \( c = \sqrt{2\sigma} \), where \( \mu, \sigma \) are the expected value and the standard deviation of a normally distributed random variable.
Examples

3. **Logistic model:** \( F(x, a, b, k) = \frac{a}{1+be^{-kx}}, k > 0 \)

The logistic function was devised as a model of population size by adjusting the exponential model which also considers the saturation of the environment, hence the growth first changes to linear and then stops.

The logistic function \( F(x, a, b, k) \) is a solution of the first order non-linear differential equation

\[
\frac{dy(x)}{dx} = ky(x)\left(1 - \frac{y(x)}{a}\right).
\]
4. In the area around a radiotelescope the use of microwave owens is forbidden, since the radiation interferes with the telescope. We are looking for the location \((a, b)\) of a microwave owen that is causing problems.

The radiation intensity decreases with the distance from the source \(r\) according to \(u(r) = \frac{\alpha}{1 + r}\).

Measured values of the signal at three locations are \(z(0, 0) = 0.27\), \(z(1, 1) = 0.36\) in \(z(0, 2) = 0.3\).

This gives the following system of equations for the parameters \(\alpha, a, b\):

\[
\begin{align*}
\frac{\alpha}{1 + \sqrt{a^2 + b^2}} &= 0.27 \\
\frac{\alpha}{1 + \sqrt{(1 - a)^2 + (1 - b)^2}} &= 0.36 \\
\frac{\alpha}{1 + \sqrt{a^2 + (2 - b)^2}} &= 0.3
\end{align*}
\]
An equivalent, more convenient formulation of the nonlinear system

- Input: the data points \( \{(x_1, y_1), \ldots, (x_m, y_m)\} \), \( x_i \in \mathbb{R}^n \), \( y_i \in \mathbb{R} \).
- Fitting function: \( F(x, a_1, \ldots, a_p) \) with unknown parameters \( a_1, \ldots, a_p \).
- Equivalent formulation of the system

\[
F(x_i, a_1, \ldots, a_p) = y_i \quad \text{for} \quad i = 1, \ldots, m,
\]

which will be more suitable for solving with numerical algorithms is:

1. Define the functions \( f_i : \mathbb{R}^p \to \mathbb{R} \) for \( i = i, \ldots, m \) by

\[
f_i(a_1, \ldots, a_p) = y_i - F(x_i, a_1, \ldots, a_p).
\]

2. Solve the system

\[
f_1(a_1, \ldots, a_p) = 0,
\]

\[
\vdots
\]

\[
f_m(a_1, \ldots, a_p) = 0.
\]
In a more compact way we can also write the system from 2. by defining a vector function

\[ f : \mathbb{R}^p \rightarrow \mathbb{R}^m, \]

\[ f(a_1, \ldots, a_p) = (f_1(a_1, \ldots, a_p), \ldots, f_m(a_1 \ldots, a_p)), \]

and search for the tuples \((a_1, \ldots, a_p)\) which are its zeroes, i.e.,

\[ f(a_1, \ldots, a_p) = (0, \ldots, 0). \]  \hspace{1cm} (3)

Solving such a system of nonlinear equations is a difficult problem in general, even for \( p = m = 1 \), e.g., there does not even exist an analytic formula to determine zeroes of the polynomial of degree 5 or more.

**Our plan:** We will learn some numerical algorithms to find approximate solutions of the system (3).
3.1 Vector functions of a vector variable

Formal definition

Let $D$ be any subset of $\mathbb{R}^n$. A vector function $f : D \to \mathbb{R}^m$ is a rule, which assigning to every vector $x \in D$ some vector $f(x) \in \mathbb{R}^m$, i.e.,

$$f : \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mapsto \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} = \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{bmatrix}$$

Equivalently,

$$f \equiv (f_1, \ldots, f_m)$$

is a vector function if all its component functions $f_i : \mathbb{R}^n \to \mathbb{R}$ are functions in the usual sense.
Examples

1. A **linear vector function** \( f : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is such that all the component functions \( f_i \) are linear, i.e., for each \( i \) we have

\[
f_i(x_1, \ldots, x_n) = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n.
\] (4)

Forming a matrix

\[
A = \begin{bmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  a_{21} & a_{22} & \cdots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix},
\]

\( f \) is given by the rule \( f(x) = Ax \), i.e., it is a usual linear transformation.

2. Adding constants \( b_i \) to the left side of (4) we get the definition of an **affine linear vector function**, i.e.,

\[
f_i(x_1, \ldots, x_n) = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n + b_i,
\]

and \( f = Ax + b \), where

\[
b = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}^T.
\]
3. Some nonlinear vector functions are

\[ f : \mathbb{R}^3 \rightarrow \mathbb{R}^2, \quad f : \begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} x^2 + y^2 + z^2 - 1 \\ x + y + z \end{bmatrix}, \]

\[ g : \mathbb{R}^2 \rightarrow \mathbb{R}^3, \quad g : \begin{bmatrix} z \\ w \end{bmatrix} \mapsto \begin{bmatrix} zw \\ \cos z + w^2 - 2 \\ e^{2z} \end{bmatrix}, \]

\[ h : \mathbb{R} \rightarrow \mathbb{R}^2, \quad h : t \mapsto \begin{bmatrix} t + 3 \\ e^{-3t} \end{bmatrix}. \]
Derivative

The derivative of a vector function $f$ in the point $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ is given by the Jacobian matrix:

$$J(a) = Df(a) = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1}(a) & \ldots & \frac{\partial f_1}{\partial x_n}(a) \\
\vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1}(a) & \ldots & \frac{\partial f_m}{\partial x_n}(a)
\end{bmatrix}.$$

▶ If $n = m = 1$, $f$ is a usual function from $D \subset \mathbb{R}$ to $\mathbb{R}$ and $Df(x) = f'(x)$ is the usual derivative.
For general $n$ and $m = 1$, $f$ is a function of $n$ variables and $Df(x) = \text{grad } f(x)$ is its gradient.

For general $m$ and $n$, $Df(x) = \begin{bmatrix} \text{grad } f_1 \\ \vdots \\ \text{grad } f_m \end{bmatrix}$ is a vector of gradients of component functions.
Examples

1. If $f : \mathbb{R}^n \to \mathbb{R}^m$ is an affine linear function $x \mapsto Ax + b$, then $Df(x) = A$.

2. If $f : \mathbb{R}^3 \to \mathbb{R}^2$ is given by

   $$f : \begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} x^2 + y^2 + z^2 - 1 \\ x + y + z \end{bmatrix},$$

then

   $$Df(x) = \begin{bmatrix} 2x & 2y & 2z \\ 1 & 1 & 1 \end{bmatrix}$$
A **linear approximation** of $f$ at the point $a$ is the linear function that has the same value and the same derivative as $f$ at $a$:

$$L_a(x) = f(a) + Df(a)(x - a).$$

$n = m = 1$:

$$L_a(x) = f(a) + f'(a)(x - a)$$

The graph $y = L_a(x)$ is the tangent to the graph $y = f(x)$ at the point $a$: 
\( n = 2, \ m = 1, \) i.e., \( f(x, y) \) is a function of two variables:

\[
L_{(a,b)}(x, y) = f(a, b) + \text{grad } f(a, b) \begin{bmatrix} x - a \\ y - b \end{bmatrix}.
\]

The graph \( z = L_{(a,b)}(x, y) \) is the tangent plane to the surface \( z = f(x, y) \) at the point \((a, b)\).
The linear approximation to the function

\[
f : \begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} x^2 + y^2 + z^2 - 1 \\ x + y + z \end{bmatrix}
\]

at \( a = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \)

is the affine linear function

\[
L_a(x, y, z) = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 & -2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x - 1 \\ y + 1 \\ z - 1 \end{bmatrix} = \\
\begin{bmatrix} 2 + 2(x - 1) - 2(y + 1) + 2(z - 1) \\ 1 + (x - 1) + (y + 1) + (z - 2) \end{bmatrix}
\]

\[
= \begin{bmatrix} 2 & -2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} -4 \\ 0 \end{bmatrix}.
\]
Given a vector function $f : \mathbb{R}^3 \to \mathbb{R}^2$, every point $(x_0, y_0, z_0)$ lies in the intersection of level surfaces

$$f_1(x, y, z) = c_1 \quad \text{and} \quad f_2(x, y, z) = c_2,$$

where $c_1 = f_1(x_0, y_0, z_0)$ and $c_2 = f_2(x_0, y_0, z_0)$.

The intersection of two surfaces in $\mathbb{R}^3$ determines an implicit curve in $\mathbb{R}^3$.

If they are nonzero, the vectors $\text{grad } f_1(x_0, y_0, z_0)$ and $\text{grad } f_2(x_0, y_0, z_0)$ are normal vectors of the two level surfaces, and

$$\text{grad } f_1(x_0, y_0, z_0) \times \text{grad } f_2(x_0, y_0, z_0)$$

is tangential to the implicit curve.
Example

For the function

\[
 f : \begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} x^2 + y^2 + z^2 - 1 \\ x + y + z \end{bmatrix}
\]

the implicit curve through \((1, -1, 1)\) is given by the system

\[
 x^2 + y^2 + z^2 - 1 = 2 \quad \text{and} \quad x + y + z = 1.
\]

Geometrically, it is the intersections of a sphere with radius \(\sqrt{3}\) and a plane with the normal \((1, 1, 1)\), i.e., the circle. The tangent vector to the circle in the point \((1, -1, 1)\) is

\[
 \nabla f_1(1, -1, 1) \times \nabla f_2(1, -1, 1) = \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -4 \\ 0 \\ 4 \end{bmatrix}.
\]
3.2 Solving systems of nonlinear equations

\[ f: D \rightarrow \mathbb{R}^m, \ D \subset \mathbb{R}^n \]

We are looking for solutions of

\[
\begin{bmatrix}
    f_1(x) \\
    \vdots \\
    f_m(x)
\end{bmatrix}
= \begin{bmatrix}
    f_1(x_1, \ldots, x_n) \\
    \vdots \\
    f_m(x_1, \ldots, x_n)
\end{bmatrix}
= \begin{bmatrix}
    0 \\
    \vdots \\
    0
\end{bmatrix}
\]

In many cases an analytic solution does not even exist.

A number of \textit{numerical methods} for approximate solutions is available. We will look at one, based on linear approximations.
Newton’s method for \( n = m = 1 \)

We want to solve an equation \( f(x) = 0, \ x \in \mathbb{R} \).

**Newton’s or tangent method:**

We construct a recursive sequence with

- \( x_0 \) initial term
- \( x_{k+1} \) solution of

\[
L_{x_k}(x) = f(x_k) + f'(x_k)(x - x_k) = 0, \ \text{so} \ x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}
\]
Newton’s method for \( n = m = 1 \) - continued

The sequence \( x_i \) converges to a solution \( \alpha, f(\alpha) = 0 \), if:

1. \( 0 \neq |f'(x)| < 1 \) for all \( x \in I \), where \( I \) is some interval containing \( \alpha \),
2. \( x_0 \in I \).

Under these assumptions the convergence is \textit{quadratic}:

if \( \varepsilon_i = |x_i - \alpha| \) then \( \varepsilon_{i+1} \leq M\varepsilon_i^2 \),

where \( M \) is a constant bounded by \( \max_{x \in I} |f''(x)|/|f'(x)| \).
Newton’s method for $n = m > 1$

Newton’s method generalizes to systems of $n$ nonlinear equations in $n$ unknowns:

- $x_0$ – initial approximation,
- $x_{k+1}$ – solution of

$$L_{x_k}(x) = f(x_k) + Df(x_k)(x - x_k) = 0,$$

so $x_{k+1} = x_k - Df(x_k)^{-1}f(x_k)$.

In practice inverses are difficult to calculate (require to many operations) and the linear system for $x_{k+1}$

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

is solved at each step (using $LU$ decomposition of $DF(x_k)$).

The sequence converges to a solution $\alpha$ if for some $r > 0$ the matrix $Df(x)$ is nonsingular for all $x$, $\|x - \alpha\| < r$, and $\|x_0 - \alpha\| < r$. 
Application of Newton’s method for $n = m > 1$ to optimization

**Newton optimization method:**

We would like to find the extrema of the function $F : \mathbb{R}^n \to \mathbb{R}$.

Since the extrema are *critical points*, the candidates are zeroes of the gradient, i.e.,

$$f(x) = \text{grad } F(x) = \begin{bmatrix} F_{x_1}(x) \\ \vdots \\ F_{x_n}(x) \end{bmatrix} = 0.$$

This is a system of $n$ equations for $n$ variables, the Jacobian of the vector function $f$ is the Hessian of $F$:

$$Df(x) = H(x) = \begin{bmatrix} F_{x_1x_1} & \cdots & F_{x_1x_n} \\ \vdots & \ddots & \vdots \\ F_{x_nx_1} & \cdots & F_{x_nx_n} \end{bmatrix}.$$

If the sequence of iterates

$$x_0, \quad x_{k+1} = x_k - H^{-1}(x_k)\text{grad } F(x_k)$$

converges, the limit is a critical point of $F$, i.e. a candidate for the minimum (or maximum).
Newton’s method for \( m > n > 0 \)

We have an overdetermined system \( f(x) = (0, \ldots, 0) \) of \( m \) nonlinear equations for \( n \) unknowns.

The system \( f(x) = (0, \ldots, 0) \) generally does not have a solution.

We are looking for a best fit to a solution, that is, for \( \alpha \) such that the distance of \( f(\alpha) \) from the origin is the smallest possible:

\[
\|f(\alpha)\|^2 = \min \{\|f(x)\|^2\}.
\]

The \textit{Gauss-Newton method} is a generalization of the Newton’s method where instead of the inverse of the Jacobian its MP inverse is used at each step:

\begin{itemize}
  \item \( x_0 \) initial approximation
  \item \( x_{k+1} = x_k - Df(x_k)^+ f(x_k) \),
\end{itemize}

where \( Df(x_k)^+ \) is the MP inverse of \( Df(x_k) \). If the matrix

\[
(Df(x_k)^T Df(x_k))
\]

is nonsingular at each step \( k \) then

\[
x_{k+1} = x_k - (Df(x_k)^T Df(x_k))^{-1} Df(x_k)^T f(x_k).
\]
Newton’s method for $m > n > 0$ - continued

At each step $x_{k+1}$ is the least squares approximation to the solution of the overdetermined linear system $L_{x_k}(x) = 0$, that is,

$$\|L_{x_k}(x_{k+1})\|^2 = \min\{\|L_{x_k}(x)\|^2, x \in \mathbb{R}^n\}.$$

Convergence is not guaranteed, but:

- if the sequence $x_k$ converges, the limit $x = \lim_k x_k$ is a local (but not necessarily global) minimum of $\|f(x)\|^2$.

It follows that the Gauss-Newton method is an algorithm for the local minimum of $\|f(x)\|^2$. 