Mathematical modelling

Chapter 1

Linear models, systems of equations

Faculty of Computer and Information Science
University of Ljubljana

2017/2018

Introduction

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

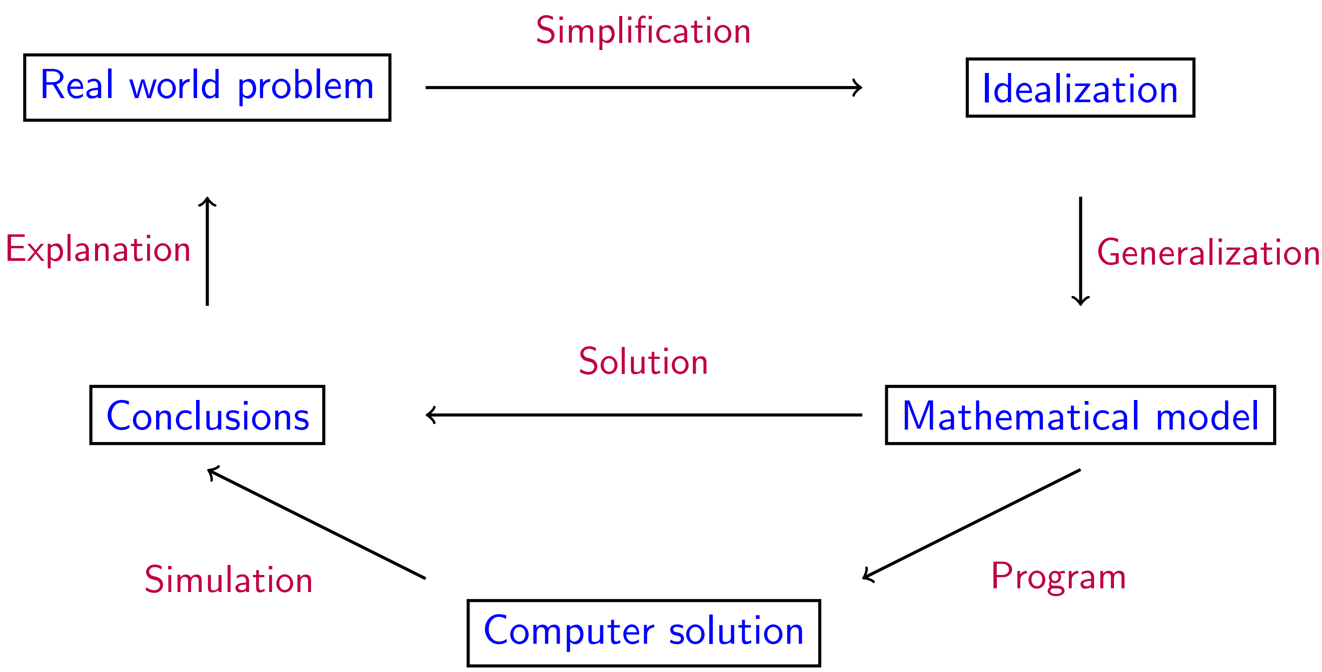
In this course we will introduce some (by far not all) mathematical tools that are used in setting up and solving mathematical models.

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

Contents

- Linear models: sistems of linear equations, matrix inverses, SVD decomposition, PCA
- Gemetric models: vector functions, curves and surfaces
- Dynamical models: differential equations, dynamical systems

Modelling cycle



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 experimantal 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.

Example

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 = $\frac{\text{Absolute error}}{\text{Correct value}}$

$$\delta_x = \frac{\Delta x}{x}$$

Estimating computational errors

Operation	exact value	approximate relative error
$\pm$	$x \pm y$	$\frac{x}{x \pm y}\delta_x + \frac{y}{x \pm y}\delta_y$
$\cdot$	xy	$\delta_x + \delta_y$
$/$	$\frac{x}{y}$	$\delta_x - \delta_y$

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 ε (precision of data and computation) should be known.

Example: quadratic equation

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

Analytic solutions are $x_1 = -p - \sqrt{p^2 + q}$ in $x_2 = -p + \sqrt{p^2 + q}$.

What happens if

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

1. Linear mathematical models

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

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 *linear*, when the function F is a linear function of yje 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. *linear regression*: $x, y \in \mathbb{R}$, $\varphi_1(x) = 1, \varphi_2(x) = x$
2. *polinomial regression*: $x, y \in \mathbb{R}$, $\varphi_1(x) = 1, \ldots, \varphi_p(x) = x^{p-1}$
3. *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. *frequency* or *spectral analysis*: $\varphi_1(x) = 1, \varphi_2(x) = \cos \omega x, \varphi_3(x) = \sin \omega x, \varphi_4(x) = \cos 2\omega x, \ldots$ (there can be infinitely many function $\varphi_i(x)$ in this case)

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

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,

$$\begin{bmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \ldots & \varphi_p(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \ldots & \varphi_p(x_2) \\ \ldots & \ldots & \ldots & \ldots \\ \varphi_1(x_m) & \varphi_2(x_m) & \ldots & \varphi_p(x_m) \end{bmatrix} \begin{bmatrix} a_1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} y_1 \\ y_1 \\ \vdots \\ y_p \end{bmatrix}.$$

1.1 Systems of linear equations and generalized inverses

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 he number of unknowns,
x is the *vector of unknowns* and
b is the *right side vector*.

Existence of solutions:

Let $A = [a_1, \dots, a_n]$, where a_i are vector representing the columns od 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, when it is in the column space of A.

By adding b to the columns of A we obtain the *extended matrix of the system*

$$[A \mid b] = [a_1, \dots, a_n \mid b],$$

Theorem (0)

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

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

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 decribe 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.

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 ...

Za pravokotno matriko A reda $m \times n$, $m \neq n$, inverz sploh ni definiran.

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!

Here is an algorithm for computing a general invere of A.

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 element 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

$$\begin{bmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 0 & 1 & 4 \end{bmatrix}.$$

1.2. Generalized inverses and solvable systems of linear equations

Theorem (1)

If the system $Ax = b$ is solvable (that is, $\operatorname{rang} A = \operatorname{rang} [A \mid b]$), and G is a generalized inverse of A, then $x = Gb$ is a solution of the system.
Moreover, every solution of the system is of the form $x = Gb$, where G is some generalized inverse of A.

Proof. Let $A = [a_1, a_2, \dots a_n]$, where a_i are the column vectors of A, and let G be a generailzed inverse of A.

Now let $x = Gb$. We know that the system is solvable, so b is in the column space of A, that is $b = \sum_i \alpha_i a_i$ for some α_i. Then,

$$Ax = A(Gb) = AG(\sum \alpha_i a_i) = \sum \alpha_i AGa_i.$$

Since $AGA = A$ it follows that $AG(a_i) = a_i$ for every column a_i of A, so

$$\sum \alpha_i AGa_i. = \sum \alpha_i a_i = b.$$

Now let x be a solution and let G be any matrix such that $x = Gb$. Then

$$Ax = A(Gb) = AG(\sum_i x_i a_i) = \sum_i x_i AG(a_i) = \sum_i x_i a_i.$$

This imples that $AGa_i = a_i$ for all i, so $AGA = A$.

Theorem (2)

Let A be a matrix of order $m \times n$ and G its generalized inverse. If the system $Ax = b$ is solvable then for any vector $z \in \mathbb{R}^m$ the vector

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

is a solution of the system. Moreover every solution is of this form for some vector $z \in \mathbb{R}^m$.

Proof: The set $\{(GA - I)z; z \in \mathbb{R}^m\}$ is precisely the null space of the matrix A since, for every z,

$$A(GA - I)z = (AGA - A)z = 0.$$

Example:

Find all solutions of the system

$$Ax = b,$$

where $A = \begin{bmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 0 & 1 & 4 \end{bmatrix}$ in $b = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$.

An easy consequence of Theorem 1 above is that a singular matrix A, such that the system $Ax = b$ is solvable, has infinitely many generalized inverses since, in this case, the system must have infinitely many solutions.

Even more is true: any singular matrix has infinitely many generalized inverses. If we know one, we know them all.

Theorem (3)

Let A be an $m \times n$ matrix and G its generalized inverse. Then every matrix of the form

$$G_A = GAG + W - GAWAG,$$

where W is an arbitrary matrix of dimension $m \times n$, is a generalized inverse of A and all generalized inverses are of this form.

A proof of this theorem can be found in the book http://onlinelibrary.wiley.com/book/10.1002/9781118491782

It is easy to see that if A is invertible, then for any W, $G_A = A^{-1}$

1.3 The Moore-Penroseov generalized inverse

Amog all the generalized inverses of a singuar matrix A, one has especially nice properties.

Definition

The *Moore-Penroseov generalized inverse*, or shortly the *MP inverse* of A is the unique matrix A^+ satifying 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: $(A^+A)^T = A^+A$

Let us show that the MP inverse A^+ of a matrix A is unique.

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

Then,

$$\begin{aligned} AG_1 &= (AG_2A)G_1 && \text{by propery (1)} \\ &= (AG_2)(AG_1) = (AG_2)^T(AG_1)^T && \text{by propery (3)} \\ &= G_2^T(AG_1A)^T = G_2^TA^T && \text{by propery (1)} \\ &= AG_2 && \text{by propery (3)} \end{aligned}$$

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

$$G_1A = G_2A,$$

and so

$$G_1 = G_1AG_1 = G_1AG_2 = G_2AG_2 = G_2.$$

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.

Case 1: A^TA is an invertible matrix

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

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

1. $AGA = A((A^TA)^{-1}A^T)A = (A(A^{-1}(A^T)^{-1}A^T)A = A$
2. $GAG = ((A^TA)^{-1}A^T)A((A^TA)^{-1}A^T) = (A^TA)^{-1}A^T$
3. $(AG)^T = (A(A^TA)^{-1}A^T)^T = A(A^TA)^{-1})^TA^T = A((A^TA)^{-1}A) = AG$
4. $\ldots$

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: Σ 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,*
- *V is an orthogonal $n \times n$ matrix and*
- $\Sigma = \left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ & & & \sigma_r & \\ \hline & & 0 & & 0 \end{array}\right]$ *is a diagonal $m \times n$ matrix*

such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

Then $A^+ = V\Sigma^+ U^T$, where Σ^+ is computed according to Case 3.

1.4 SVD and PCA

Let $A = U\Sigma V^T$, where U, Σ and V are as above.

The elements σ_i are the *singular values* of A, and the columns of U and V are the *left* and *right singular vectors* of A, respectively.

How do we find them?

Assume that we have the SVD $A = U\Sigma V^T$. Then

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

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

are diagonalizations of the symmetric square matrices $A^T A$ of dimension $n \times n$ and AA^T of dimension $m \times m$, respectively.

It follows that:

- The diagonal matrices $\Sigma^T\Sigma$ and $\Sigma\Sigma^T$ have the eigenvalues of A^TA and AA^T on the diagonal. Ordered in decreasing order $\sigma_1^2 \geq \sigma_2^2 \geq \sigma_r^2 \geq 0$ they are equal for both matrices, with possibly a different number of zeroes following.
- The columns of the matrices $U = [u_1, \ldots, u_m]$ and $V = [v_1, \ldots, v_n]$ form orthonormal bases of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, consisting of eigenvectors, the first r corresponding to the nonzero eigenvalues, and rest to $\lambda = 0$.
- For any left singular vector u_i, $i = 1, \ldots, r$, the vector $\tilde{v}_i = A^T u_i$ has the property

$$(A^TA)\tilde{v}_i = A^T(AA^T)u_i = A^T(\lambda_i u_i) = \lambda_i(A^Tu_i) = \lambda_i\tilde{v}_i,$$

  so it is an eigenvalue of (A^TA) corresponding to λ_i. Its normalization $v_i = \dfrac{\tilde{v}_i}{\|\tilde{v}_i\|}$ is a right singular vector.
- Similarly, if v_i is a right singular vector, the left singular vector u_i is the normalizatrion of Av_i.

Algorithm for SVD

- Compute the eigenvalues and an orthonormal basis of eigenvectors of the symmetric matrix A^TA 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 λ_i are the nonzero eigenvalues of the symmetric matrices $(A^TA)^{n\times n}$ and $(AA^T)^{m\times m}$.
- The left singular vectors (i.e. columns of U) are the orthonormal eigenvectors of AA^T.
- The right singular vectors (i.e. columns of V) are the orthonormal eigenvectors of A^TA.
- If u (resp. v) is a left (resp. right) singular vector corresponding to the singular value σ_i, then $v = Au$ (resp. $u = A^Tv$) is a right (resp. left) singular vector corresponding to the same singular value.

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, ..., v_r of $A^T A$,
2. Let Σ 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 corenr 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 te matrix of normalized columns of AV.
5. Compute $A^+ = V \Sigma^+ U^T$.

An application of SVD: *principal coponent anaysis* 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 $m << n$, where n is very large, for example, $X_1, \ldots, X_m$ can be

- vectors of gene experssions 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 $n << m$ 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 centeroid is in the origin

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

If not, we substract μ 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 $(X^TX)^{m\times m}$ 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 σ_i are the singular values of X).
- The *principal directions* in $\mathbb{R}^n$ are the columns $v_1, \ldots, v_r$ are the coresponding eigenvectors, i.e. the column of the matrix V from the SVD of X. The remaining clolumns of V (i.e. the eigenvectors correspondong to 0) form a basis of the null space of X.
- The first column v_1, that is, *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 data compression methd 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 $v_1, \ldots, v_k$ in $\mathbb{R}^n$.

The idea is to approximate

$$X_i = u_{1,i}v_1 + \cdots + u_{k,i}v_k + \cdots + u_{m,i}v_m \cong u_{1,i}v_1 + \cdots + u_{k,i}v_k$$

with the first k most informative direction in $\mathbb{R}^n$ and supress the last $m - k$.

PCA has the following amazing property:

Theorem

Among all possible projections of $p\colon \mathbb{R}^n \to \mathbb{R}^k$ 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.

1.5 The MP inverse and systems of linear equations

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

Theorem (4)

1. *An underdetermined system of linear equations*

$$Ax = b \tag{1}$$

   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 have 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, usind 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 Theorem (1).
2. It follows from Theorem (2) that every solution of the system is of the form $x = A^+b + (A^+A - I)z$. Then, (using the lemma in the fourth equality) for any solution x,

$$\begin{aligned}
\|x\| &= \|A^+b + (A^+A - I)z\|^2 \\
&= (A^+b + (A^+A - I)z) \cdot (A^+b + (A^+A - I)z) \\
&= \|A^+b\|^2 + 2A^+b \cdot (A^+A - I)z + \|(A^+A - I)z\|^2 \\
&= \|A^+b\|^2 + \|(A^+A - I)z\|^2 \\
&\geq \|A^+b\|^2,
\end{aligned}$$

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 $\sqrt{11}$ 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

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

The MP inverse is then

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

and the solution is

$$x^+ = A^+ b = \begin{bmatrix} \frac{6}{\sqrt{11}} \\ \frac{2}{\sqrt{11}} \\ \frac{2}{\sqrt{11}} \end{bmatrix}.$$

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, than a best fit for the solution is a vector x such that the error $||Ax - b||$ or, equivalentnly its square

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

is the smallest possible.

Theorem (5)

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 $\{Ax, x \in \mathbb{R}^n\}$ os A is the orthogonal projection onto it of b.

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

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

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

$$\begin{bmatrix} x_1 & 1 \\ \vdots & \\ x_m & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

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*.