1. Fundamentals of Linear Algebra
   - Matrices and Vectors
   - Span and Image – Linear Independence and Kernel
   - Subspaces of $\mathbb{F}^n$ – Basis and Dimension
   - Inverse Matrices
   - The Euclidean Norm
   - Orthogonal Vectors and Matrices
   - The Determinant
   - Linear Systems of Equations
   - More on Image and Kernel

November 15, 2022

Recommended reading for this section:

- Lectures 1,2,3 in [3]
- Sections I.1, I.2, I.3, I.5(, I.11) in [2]
- Chapters 1,3(,4,5) in [1]

Literature:

[1] G. Strang.
Introduction to Linear Algebra.
Wellesley-Cambridge Press, 2003.

[2] G. Strang.
Linear Algebra and Learning from Data.
Wellesley-Cambridge Press, 2019.

[3] L.N. Trefethen and D. Bau.
Numerical linear algebra.
SIAM, Soc. for Industrial and Applied Math., Philadelphia, 1997.

1 Fundamentals of Linear Algebra

1.1 Matrices and Vectors

Example 1.1 (*Interpolation*)
Assume we are given the following measurements

z_i	-1	0	1
y_i	0	1	0

We postulate that these measurements can be explained exactly by the (quadratic) model

$$f(z) := f_x(z) := x_1 + x_2 z^2.$$

Question: Can we find parameters $x_1, x_2 \in \mathbb{R}$, so that $f(z_i) = y_i$ for all $i = 1, \ldots, 3$?

We first translate the task into the following system of *linear* equations:

$$\begin{array}{lrcl} i=1: & 1x_1 + 1x_2 & = & 0 \\ i=2: & 1x_1 + 0x_2 & = & 1 \\ i=3: & 1x_1 + 1x_2 & = & 0 \end{array} \quad :\Leftrightarrow \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}}_{b} \quad \Leftrightarrow \quad Ax = b$$

Why linear? Roughly speaking, because the x_i only appear with power 1 and there are no combinations $x_i x_j$.

Also note, the abstracting notation $Ax = b$ provides a rigorous interface to analyze this problem theoretically and also to implement numerical solvers.

[ex:solve and draw points]

Throughout we will consider matrices over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. However, $\mathbb{F}$ could be replaced by any field.

Definition 1.2 (*Matrix*)

Let $m, n \in \mathbb{N}$. Then a rectangular array of numbers in $\mathbb{F}$ with m rows and n columns, written as

$$A = (a_{ij})_{ij} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

is called a $(m \times n)$ **matrix with coefficients in** *$\mathbb{F}$.*

$\mathbb{F}^{m\times n}$: *set of all $(m \times n)$ matrices with coefficients in $\mathbb{F}$*

$\mathbb{F}^n := \mathbb{F}^{n\times 1}$: *matrices with just one column are called (column)* ***"vectors"****;*
elements in $\mathbb{F}^{1\times n}$ are referred to as row vectors

a_{ij}: *the (i, j)-th* **coefficient** *or* **entry**

$(a_{i1}, ..., a_{in})$: *the i-th* **row** *of A (this is a $(1 \times n)$-matrix, , n-dimensional vector)*

$\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$: *the j-th* **column** *of A (this is a $(m \times 1)$-matrix, m-dimensional vector)*

0: *the* **zero matrix** *or* **null matrix***, i.e. all entries 0*

I_n: *the* ***unity*** *or* ***identity matrix*** *in $\mathbb{F}^{n\times n}$, i.e. the Matrix with entries*

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{else} \end{cases} \quad \text{("Kronecker delta"),} \quad \text{i.e.} \quad I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

Operations

We can add matrices of same size and scale the entries of a matrix.

Example 1.3 (*Summing and scaling matrices*)

$$\begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} := \begin{pmatrix} 3 & 0 \\ 1 & -1 \end{pmatrix}$$

$$2 \cdot \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 2 \cdot 2 & 2 \cdot 1 \\ 2 \cdot 0 & 2 \cdot (-1) \end{pmatrix} := \begin{pmatrix} 4 & 2 \\ 0 & -2 \end{pmatrix}$$

Definition 1.4 (*Summing and scaling matrices*) *Let $A, B \in \mathbb{F}^{m \times n}$ be matrices, $m, n \in \mathbb{N}$ and $r \in \mathbb{F}$.*

i) ***Sum of matrices:*** $+ \colon \mathbb{F}^{m \times n} \times \mathbb{F}^{m \times n} \to \mathbb{F}^{m \times n}$

The sum $C := A + B$ of the two matrices A and B is defined to be the matrix $C = (c_{ij})_{ij} \in \mathbb{F}^{m \times n}$ with entries

$$c_{ij} := a_{ij} + b_{ij} \quad \text{for} \quad i = 1, \ldots, m, \; j = 1, \ldots, n.$$

ii) ***Multiplication with scalars:*** $\cdot \colon \mathbb{F} \times \mathbb{F}^{m \times n} \to \mathbb{F}^{m \times n}$

The product of the matrix A with $r \in \mathbb{F}$ is defined to be the scaled matrix

$$r \cdot A := (r \cdot a_{ij})_{ij}.$$

In this context, elements of the field $\mathbb{F}$ are called ***scalars****.*

Question: How do summing and scaling get along in a mixed expression?
We can prove the following rules:

Lemma 1.5 (*Compatibility properties of summing and scaling*) *Let $A, B \in \mathbb{F}^{m\times n}$ and $r, s \in \mathbb{F}$. Then*

$$\begin{array}{lr} i) & (r\cdot s)\cdot A = r\cdot(s\cdot A) \\ ii) & (r+s)\cdot A = r\cdot A + s\cdot A \\ & r\cdot(A+B) = r\cdot A + r\cdot B \\ iii) & 1\cdot A = A \end{array}$$

Proof. Follow immediately from Definition 1.4 in terms of matrix coefficients and the field properties of $\mathbb{F}$. For example:

i) $(r\cdot s)A = [(r\cdot s)a_{ij}]_{ij} \overset{\text{(associativity in}\mathbb{F})}{=} [r\cdot(s\cdot a_{ij})]_{ij} = r\cdot(s\cdot A)$

ii) $(r+s)A = [(r+s)a_{ij}]_{ij} \overset{\text{(distributivity in}\mathbb{F})}{=} [ra_{ij} + sa_{ij}]_{ij} = r\cdot A + s\cdot A$

□

Remark 1.6 One can show that $(\mathbb{F}^{m\times n}, +)$ is a so–called abelian group. Then the compatibility properties from Lemma 1.5 imply that $(\mathbb{F}^{m\times n}, +, \cdot)$ is a so-called vector space (over $\mathbb{F}$).

From Lemma 1.5 we can deduce further properties of $(\mathbb{F}^{m\times n}, +, \cdot)$.

Corollary 1.7 *Let $A \in \mathbb{F}^{m\times n}$ and $r \in \mathbb{F}$. Then*

$$\begin{array}{lrl} i) & 0\cdot A = 0 & \\ ii) & r\cdot 0 = 0 & \\ iii) & r\cdot A = 0 & \Rightarrow r = 0 \ \vee\ A = 0 \\ iv) & (-1)\cdot A = -A & \qquad \textit{(i.e., } (-1)\cdot A \textit{ is the additive inverse of } A\textit{)} \end{array}$$

Proof. Follows from Lemma 1.5 and the field properties of $\mathbb{F}$. For example:

iv) To show: $(-1)\cdot A + A = 0$. We find

$$(-1)\cdot A + A \overset{\text{(L.1.5 iii))}}{=} (-1)\cdot A + 1\cdot A \overset{\text{(L.1.5 (ii))}}{=} (-1+1)\cdot A = 0\cdot A \overset{\text{(i)}}{=} 0.$$

Next we provide a notation which enables us to write linear systems of equations in a concise way.
We recall from Example 1.1:

$$\begin{array}{rcl} 1x_1 + 1x_2 & = & 0 \\ 1x_1 + 0x_2 & = & 1 \\ 1x_1 + 1x_2 & = & 0 \end{array} \quad :\Leftrightarrow \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}}_{b} \quad \Leftrightarrow \quad Ax = b$$

Definition 1.8 (*Matrix-Vector Product*) *Let $A \in \mathbb{F}^{m\times n}$ and $x \in \mathbb{F}^n$. Then the **matrix-vector product** $b = Ax \in \mathbb{F}^m$ is defined by*

$$b_i = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n =: \sum_{\ell=1}^{n} a_{i,\ell}x_\ell\,, \qquad \forall i = 1, \ldots, m.$$

A matrix $A \in \mathbb{F}^{m\times n}$ can therefore also be considered as a (linear) mapping

$$f_A : \mathbb{F}^n \to \mathbb{F}^m, \quad x \mapsto Ax.$$

Example 1.9 (*Matrix-Vector Product*)

Let us consider $A = \begin{pmatrix} 1 & 2 \\ 2 & 0 \\ 0 & 1 \end{pmatrix}$ with columns $a_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ and $a_2 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$. Then

$$\begin{array}{l} i=1 \\ i=2 \\ i=3 \end{array} \quad \begin{pmatrix} 1 & 2 \\ 2 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0.5 \\ 3 \end{pmatrix} = \begin{pmatrix} 1\cdot 0.5 + 2\cdot 3 \\ 2\cdot 0.5 + 0\cdot 3 \\ 0\cdot 0.5 + 1\cdot 3 \end{pmatrix} = \begin{pmatrix} 6.5 \\ 1 \\ 3 \end{pmatrix}.$$

There are two ways of perceiving the matrix-vector product:

(1) By rows: *Used for computations*

$$\begin{pmatrix} 1 & 2 \\ 2 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1x_1 + 2x_2 \\ 2x_1 + 0x_2 \\ 0x_1 + 1x_2 \end{pmatrix} \qquad = \begin{pmatrix} \text{inner products} \\ \text{of the rows} \\ \text{with } (x_1, x_2) \end{pmatrix}$$

→ This refers to the way of computing the matrix-vector product according to "**row** · **column**".

We give this type of product of two vectors a special name:

Definition 1.10 (*Inner product*) *Let $x, y \in \mathbb{F}^n$ be two vectors. Then the (standard)* ***inner product of*** x ***and*** y *is defined by*

$$(x, y)_2 := \overline{x} \cdot y := \sum_{i=1}^{n} \overline{x}_i y_i = \overline{x}_1 y_1 + \cdots + \overline{x}_n y_n,$$

where $\overline{x_i}$ denotes the complex conjugate.

- For real vectors $x, y \in \mathbb{R}^n$ this simplifies to

$$(x, y)_2 = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + \cdots + x_n y_n.$$

- This operations is sometimes also called *scalar* or *dot product*. It is a central operation and we will illuminate some properties later on.

Example 1.11 (*Inner products)*

$$\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} = 1 \cdot 3 + 2 \cdot (-1) + 0 \cdot 1 = 1, \quad \begin{pmatrix} 2+3i \\ i \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -i \\ 0 \end{pmatrix} = (\overline{2+3i}) \cdot 1 + \overline{i} \cdot (-i) + \overline{1} \cdot 0 = 2 - 3i + i^2 = 1 - 3i.$$

(2) By columns: *Used for understanding*

$$\begin{pmatrix} 1 & 2 \\ 2 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \qquad = \begin{pmatrix} \text{linear combination} \\ \text{of the columns} \\ a_1, a_2 \end{pmatrix}$$

Definition 1.12 (*Linear combination*) *Let $a_1, \dots, a_n \in \mathbb{F}^m$, $x \in \mathbb{F}^n$. Then*

$$\sum_{i=1}^{n} x_i a_i = x_1 a_1 + \cdots + x_n a_n = Ax \in \mathbb{F}^m$$

*is called **linear combination** of the vectors $a_1, \dots, a_n$. Here, $A := [a_1, \dots, a_n] \in \mathbb{F}^{m \times n}$.*

Example 1.13 (*Linear combination*)

Let us consider $\mathbb{F} = \mathbb{R}$ and again the vectors $a_1 := \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ and $a_2 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$.

Then examples of linear combinations are:
$a_1 = 1 \cdot a_1 + 0 \cdot a_2$,
$a_2 = 0 \cdot a_1 + 1 \cdot a_2$,
$a_1 + a_2 =$ ...,
....

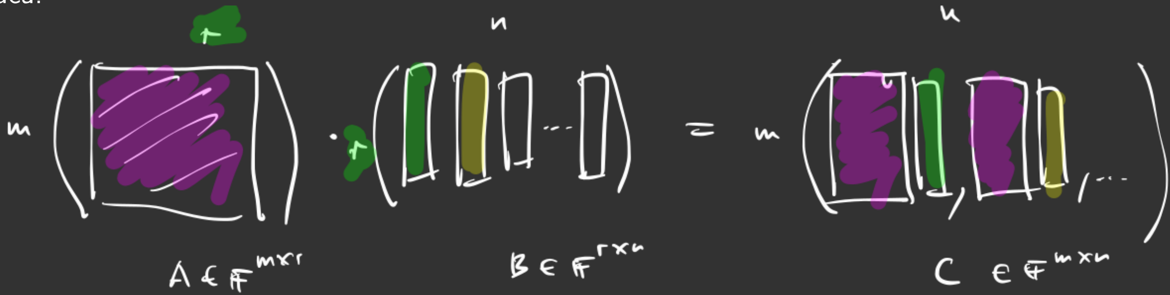
[ex:picture]

The matrix product

We generalize the *matrix-vector* product above to a *matrix-matrix* product by observing that:

"A matrix is just a collection of columns (or vectors)."

Idea:



We make this a rigorous definition:

Definition 1.14 (*Matrix-Matrix Product*) *For matrices $A \in \mathbb{F}^{m\times r}$ and $B \in \mathbb{F}^{r\times n}$, we define the **matrix-matrix product** (or simply matrix product) $C := A \cdot B \in \mathbb{F}^{m\times n}$ as a column wise product, i.e.,*

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ a_{21} & a_{22} & \dots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mr} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{r1} & b_{r2} & \dots & b_{rn} \end{pmatrix}, \text{ i.e. } \boxed{c_{ij} = \sum_{\ell=1}^{r} a_{i\ell} b_{\ell j}} \quad \begin{array}{l} i = 1, \dots, m \\ j = 1, \dots, n \end{array}$$

Note that it is of utmost importance that the matrix dimensions fit in so far as the middle dimension of $A \in \mathbb{F}^{m\times r}, B \in \mathbb{F}^{r\times n}$ (i.e., r) is the same. Otherwise, this product cannot be formulated!

Question: How do matrix sum (see Def. 1.4) and matrix product (see Def. 1.14) get along in a mixed expression?

We can prove the following rules:

Lemma 1.15 (*Compatibility properties of matrix sum and product*) *Let $A, \tilde{A} \in \mathbb{F}^{m\times n}$, $B, \tilde{B} \in \mathbb{F}^{n\times l}$, $C \in \mathbb{F}^{l\times t}$, $r \in \mathbb{F}$. Then*

(i)	(*Associativity*)	$(A \cdot B) \cdot C = A \cdot (B \cdot C)$
(ii)	(*Distributivity 1*)	$(A + \tilde{A})B = AB + \tilde{A}B$
(iii)	(*Distributivity 2*)	$A(B + \tilde{B}) = AB + A\tilde{B}$
(iv)	(*left/right neutral*)	$I_m A = A I_n = A$
(v)		$(r \cdot A) \cdot B = r(A \cdot B) = A(r \cdot B)$
(vi)	(*left/right null*)	$0A = A0 = 0$

Proof. Only a) is not trivial:

$$[(A \cdot B) \cdot C]_{ij} = \sum_{r=1}^{l} \left(\sum_{k=1}^{n} a_{ik} b_{kr} \right) c_{rj} = \sum_{k=1}^{n} a_{ik} \left(\sum_{r=1}^{l} b_{kr} c_{rj} \right) = [A \cdot (B \cdot C)]_{ij}.$$

$\square$

The (conjugate) Transpose Matrix

We finally introduce the operation of transposing matrices (and vectors):

Definition 1.16 (*Conjugate Transpose matrix*)
For a matrix $A := (a_{ij})_{ij} \in \mathbb{F}^{m\times n}$ the ***conjugate (or Hermitian) transpose matrix*** A^H ***of*** A *is defined as*

$$A^H := (\overline{a}_{ji})_{ij} \in \mathbb{F}^{n\times m},$$

where $\overline{a}_{ji}$ denotes the complex conjugate of the coefficient a_{ji}.
For a real matrix $A := (a_{ij})_{ij} \in \mathbb{R}^{m\times n}$, so that $\overline{a}_{ji} = a_{ji}$, this simplifies to

$$A^\top := A^H = (a_{ji})_{ij} \in \mathbb{R}^{n\times m}$$

which we then simply call the ***transpose matrix*** $A^\top$ ***of*** A.

Observe that we have the relation

$$A^H = \overline{A}^T,$$

where $\overline{A}$ is understood as the component-wise complex conjugate.

Example 1.17 (*Conjugate transpose*)

- Transposing a matrix.
- Transposing a vector.
- The inner product can be written as $x^H y$ (or $x^\top y$ for real vectors).
- Adjoint operator: Consider a matrix $A = [a_1, a_2] \in \mathbb{R}^{3\times 2}$ which maps $\mathbb{R}^2 \to \mathbb{R}^3$. Then $A^\top \in \mathbb{R}^{2\times 3}$ maps $\mathbb{R}^3 \to \mathbb{R}^2$ and

$$A^T p = \begin{pmatrix} a_1^\top p \\ a_2^\top p \end{pmatrix}$$

collects all inner products of p with the columns. We will relate the inner product to projections later on.

Exercise:

Question: How do transposing and other matrix operations behave in mixed expressions?

We can prove the following properties

Lemma 1.18 (*Rules for transposing*) *Let $A \in \mathbb{F}^{m\times r}, B \in \mathbb{F}^{r\times m}$ and $r \in \mathbb{F}$. Then*

i) *Transposing twice: $(A^H)^H = A$,*
ii) *Matrix product: $(AB)^H = B^H A^H$,*
iii) *Scalar product: $(rA)^H = \bar{r}A^H$,*
iv) *Matrix sum: $(A+B)^H = A^H + B^H$,*
v) *Inverse (see below): For $A \in GL(n,\mathbb{F})$ we have $(A^H)^{-1} = (A^{-1})^H$.*

Remark: For real matrices (i.e., $\mathbb{F} = \mathbb{R}$ in the lemma above), the same rules hold true for $(\cdot)^\top$.

Proof. Apply definitions and exploit properties of $\mathbb{F}$. $\square$

1.2 Span and Image – Linear Independence and Kernel

Example 1.19 (*Span and Image*)
Span: Let us again consider the two real vectors

$$a_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \; a_2 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^3.$$

Question: What are the vectors that we can represent as linear combination thereof?
There are two operations involved:

$\cdot$: Scaling each vector a_1 and a_2 individually yields infinite lines through these vectors.
$+$: By adding arbitrary vectors from these lines we fill out the infinite plane in-between.

All combinations of these two vectors form an infinite plane in $\mathbb{R}^3$. We say the plane is "spanned" by a_1 and a_2.
The terminology for the set of all linear combinations is therefore accordingly:

$$\mathrm{span}(a_1, a_2) := \{x_1 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} : \; x_1, x_2 \in \mathbb{R}\}.$$

Image: By considering these two vectors as columns of a matrix, more precisely,

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 0 \\ 0 & 1 \end{pmatrix},$$

the analogue notion is given by the so-called *image* of the matrix, which collects all matrix–vector products, i.e.,

$$\mathrm{Im}(A) \; := \{Ax : \; x \in \mathbb{R}^2\} = \{x_1 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} : \; x_1, x_2 \in \mathbb{R}\} \; = \; \mathrm{span}(a_1, a_2).$$

The set of all possible linear combinations or matrix–vector products is given a special name:

Definition 1.20 (*Span and Image*)

i) *The* ***span of** <u>vectors</u>* $a_1, \dots, a_n \in \mathbb{F}^m$ *is defined by*

$$\operatorname{span}(a_1, \dots, a_n) := \left\{ \sum_{i=1}^{n} x_i a_i : x_i \in \mathbb{F} \right\} \subset \mathbb{F}^m.$$

The set $\{a_1, \dots, a_n\}$ *is called* ***generating system*** *of* $\operatorname{span}(a_1, \dots, a_n)$.

ii) *The* ***image (or column space) of a** <u>matrix</u>* $A := [a_1, \dots, a_n] \in \mathbb{F}^{m \times n}$ *is defined by*

$$\operatorname{Im}(A) := \{Ax : x \in \mathbb{F}^n\} = \operatorname{span}(a_1, \dots, a_n) \subset \mathbb{F}^m.$$

With this terminology we find

$$\text{“}Ax = b \text{ is solvable”} \quad \Leftrightarrow \quad b \text{ is spanned by the columns of } A \quad \Leftrightarrow \quad b \in \operatorname{Im}(A).$$

[ex:picture]

Consider the example from above (i.e., $A := [a_1, a_2]$) and some vector $b = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$. By “solving” the system $Ax = b$ we want to find a linear combination of the columns a_1 and a_2 (i.e., scalars x_1 and x_2), so that this combination produces the vector b, i.e.,

$$Ax = b \quad \Leftrightarrow \quad x_1 \cdot \begin{matrix} | \\ a_1 \\ | \end{matrix} + x_2 \cdot \begin{matrix} | \\ a_2 \\ | \end{matrix} = b \quad \Leftrightarrow \quad x_1 \cdot \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + x_2 \cdot \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}.$$

In our example we find that this b is not contained in the span of the columns a_1, a_2 (= the infinite plane) so that our system is **not** solvable. In particular we find $\operatorname{Im}(A) \subsetneq \mathbb{R}^3$ (we say f_A is not “surjective”).

However, what happens if we consider an additional column, e.g.,

$$A_2 := \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}, \quad \text{or} \quad A_3 := \begin{pmatrix} 1 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

What is the difference?
(A_2) *Dependent columns:*
Here we find that the new column is a linear combination of the first two columns. More precisely,

$$1 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} \in \mathrm{Im}(A).$$

Therefore, no additional information added to the span: $A_2 x = b$ is still not solvable. Furthermore, we easily find a linear combination of the three vectors that yield the zero vector:

$$1 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} + (-1) \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We will later pay special attention to all vectors x for which $Ax = 0$ as here $x := (1, 1, -1)^\top \in \ker(A_2)$.
(A_3) *Independent columns:*

$$x_1 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{for } \textbf{all} \ \ x_1, x_2.$$

Thus, in this case we find

$$x_1 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \overset{!}{=} 0 \quad \Rightarrow \quad x_1 = x_2 = x_3 = 0$$

We will later understand: $A_3 x = b$ is definitely solvable.

Let us properly define these concepts:

Definition 1.21 ***(Linear independence and kernel)***

i) *Vectors $a_1, \ldots, a_r \in \mathbb{F}^m$ are called **linearly independent**, if the only combination that gives the zero vector is $0a_1 + \cdots + 0a_r$.*

ii) *The **kernel** of a matrix $A \in \mathbb{F}^{m \times n}$ is defined by*

$$\ker(A) := \{x \in \mathbb{F}^n : Ax = 0\},$$

i.e., the preimage of $\{0\}$ under f_A.

We find the following important equivalent formulation of linear independence:

Lemma 1.22 *For vectors $a_1, ..., a_r \in \mathbb{F}^n$ we have the equivalence:*

$a_1, ..., a_r$ *linearly independent* $\Leftrightarrow$ *every vector $b \in$ span$(a_1, ..., a_r)$ can be **uniquely** linearly combined from the set $\{a_1, ..., a_r\}$, i.e., $\exists_1 x_1, \ldots, x_r \in \mathbb{F}: b = x_1 a_1 + \ldots + x_r a_r$.*

Remark. This result implies the following for solutions of linear systems: Let x solve $Ax = b$. If A has independent columns, then the solution x is unique! On the contrary, if the columns are dependent, we will learn that there are infinitely many solutions!

Proof of Lemma 1.22 . Strategy: We split $\mathcal{A} \Leftrightarrow \mathcal{B}$ into $\mathcal{A} \Rightarrow \mathcal{B}$ and $\mathcal{A} \Leftarrow \mathcal{B}$.

- $\mathcal{A} \Rightarrow \mathcal{B}$: Let $a_1, \cdots a_r$ be linearly independent and $b \in \operatorname{span}(a_1, \cdots, a_r)$. To show unique existence, we assume that there exist two instances and show that they are the same. Therefore, let us assume that there are two sets of coefficients, say x_i and y_i, so that

$$b = \sum_{i=1}^{r} x_i a_i \text{ and } b = \sum_{i=1}^{r} y_i a_i.$$

We know that such coefficients exist, since $b \in \operatorname{span}(a_1, \cdots, a_r)$ by assumption. Now we find

$$0 = b - b = \sum_{i=1}^{r} x_i a_i - \sum_{i=1}^{r} y_i a_i = \sum_{i=1}^{r} (x_i - y_i) a_i$$

$$\Rightarrow x_i - y_i = 0, \text{ since } a_1, \cdots, a_r \text{ are linearly independent}$$

$$\Rightarrow x_i = y_i, \text{ i.e., linear combination is unique.}$$

- $\mathcal{A} \Leftarrow \mathcal{B}$: We apply a proof by contradiction (i.e., $\neg\mathcal{A} \Rightarrow \neg\mathcal{B}$). For this purpose, we assume $a_1, \cdots, a_r$ are linearly dependent ($\neg\mathcal{A}$), so that there exist $x_1, \cdots, x_r \in \mathbb{F}$ with $x_j \neq 0$ for at least one x_j, so that

$$0 = \sum_{i=1}^{r} x_i a_i = \sum_{i=1, i \neq j}^{r} x_i a_i + x_j a_j.$$

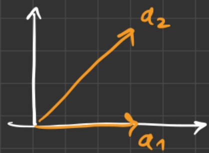
Since $x_j \neq 0$, we can write $a_j = 0a_1 + \ldots + 1a_j + \ldots + 0a_r$ also as another linear combination

$$a_j = \sum_{i=1, i \neq j}^{r} \frac{-x_i}{x_j} a_i = \frac{-x_1}{x_j} a_1 + \ldots + 0a_j + \ldots + \frac{-x_r}{x_j} a_r.$$

Thus the linear dependence ($\neg\mathcal{A}$) applies that there exists a $b \in \operatorname{span}(a_1, ..., a_r)$ (here a_j), which is not uniquely linearly combined from $a_1, \ldots, a_r$ ($\neg\mathcal{B}$).

Summary: Relation between vector and matrix notions

$a_1, \dots, a_n \in \mathbb{F}^m$

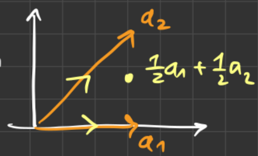


$A = \begin{pmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{pmatrix} \in \mathbb{F}^{m \times n}$

LINEAR COMBINATION

$$\sum_{i=1}^{n} x_i \begin{matrix} | \\ a_i \\ | \end{matrix} = x_1 \begin{matrix} | \\ a_1 \\ | \end{matrix} + \dots + x_n \begin{matrix} | \\ a_n \\ | \end{matrix}$$

$x_i \in \mathbb{F}$



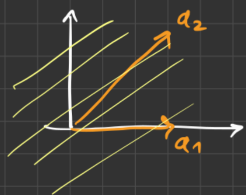
MATRIX - VECTOR PRODUCT

$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n$

$$Ax = x_1 \begin{matrix} | \\ a_1 \\ | \end{matrix} + \dots + x_n \begin{matrix} | \\ a_n \\ | \end{matrix}$$

SPAN OF VECTORS

$$\operatorname{span}(a_1, \dots, a_n) := \left\{ \sum_{i=1}^{n} x_i a_i : x_i \in \mathbb{F} \right\}$$



IMAGE

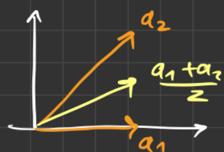
$$\operatorname{Im} A := \{ Ax : x \in \mathbb{F}^n \}$$

LINEARLY INDEPENDENT

$a_1, \dots, a_n$ are called independent

$:\Longleftrightarrow$

$Ax = x_1 a_1 + \dots + x_n a_n = 0 \quad \overset{\Rightarrow}{\underset{\Leftarrow \checkmark}{\Longleftrightarrow}} \quad x_i = 0$



KERNEL

$$\ker A := \{ x \in \mathbb{F}^n : Ax = 0 \}$$

Example 1.23 (*Subspaces*)
Let $\mathbb{F} = \mathbb{R}$.

- $n = 2$: A straight line $L := \{xa \colon x \in \mathbb{R}\} \subset \mathbb{R}^2$ spanned by a fixed $a \in \mathbb{R}^2$. For a linear function $f(x) = mx$ with $m := \frac{a_2}{a_1}$ we find

$$\text{graph}(f) := \{(x, f(x)) : x \in \mathbb{R}\} = \{(x, mx) : x \in \mathbb{R}\} = \text{span}(\begin{pmatrix} 1 \\ \frac{a_2}{a_1} \end{pmatrix}) = \text{span}(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}) \quad (\text{note: } a_1\mathbb{R} = \mathbb{R})$$

- $n = 3$: A plane $P := \{x_1 a_1 + x_2 a_2 \colon x_i \in \mathbb{R}\} \subset \mathbb{R}^3$ spanned by fixed $a_1, a_2 \in \mathbb{R}^3$.
- Not a (linear) subspace: The graph of nonlinear functions, such as x^2 or $\sin(x)$ in $\mathbb{R}^2$.

[ex:pictures]

Definition 1.24 (*Subspace*) *A subset $V \subset \mathbb{F}^n$ is called **(linear) subspace of** $\mathbb{F}^n$ if*

i) *it is nonempty, i.e., $V \neq \emptyset$,*

ii) *and if it is closed under linear combinations, i.e., if*

$$\lambda_1 v_1 + \lambda_2 v_2 \in V \quad \textit{for all} \quad v_1, v_2 \in V,\ \lambda_1, \lambda_2 \in \mathbb{F}.$$

Question: Is it possible to describe a linear subspace of $\mathbb{F}^n$ by a finite number of vectors?

Definition 1.25 (*Basis*) *Let $V \subset \mathbb{F}^n$ be a subspace of $\mathbb{F}^n$. Then a set of vectors $\{v_1, \ldots, v_r\} \subset V$ with $r \leq n$ is called* ***basis of*** V*, if*

i) $v_1, \ldots, v_r$ *are linearly independent,*

ii) *span*$(v_1, \ldots, v_r) = V$.

- Let $\{v_1, \ldots, v_r\} \subset V$ be a basis. Then, in particular, any $v \in V = \text{span}(v_1, \ldots, v_r)$ can be written as

$$v = \sum_{j=1}^{r} \lambda_j v_j$$

for some *uniquely* determined scalars $\lambda_j \in \mathbb{F}$ (see Lemma 1.22).
$\rightarrow$ These scalars are called **coordinates** of v with respect to the basis $\{v_1, \ldots, v_r\}$.

- One can show that
  - there exists a basis (general result based on Zorn's lemma),
  - any basis of a subspace of $\mathbb{F}^n$ has the same length (doable proof), which we call **dimension of V ($\dim(V)$)**.

- With other words:

The maximum number of linearly independent vectors is called dimension and the set of such vectors is called a basis.

Example 1.26
1) Let us consider $\mathbb{F} = \mathbb{R}$ and $V := \mathbb{R}^2 \subset \mathbb{R}^2$.

We first show that V is a subspace of $\mathbb{R}^2$ and then try to find some bases.
Following Definition 1.24 we show:

i) $V \neq \emptyset$: Consider $0 \in V$.

ii) V is closed under linear combinations: Let $v_1, v_2 \in V$, $\lambda_1, \lambda_2 \in \mathbb{R}$, then clearly $\lambda_1 v_1 + \lambda_2 v_2 \in \mathbb{R}^2 = V$.

Now it makes sense to talk about a *basis* for V. We next try to find a set of vectors $v_1, \ldots, v_r \in V$ that satisfies Definition 1.25.

1a) Let us consider the vectors $e_j := (\delta_{ij})_{1 \le i \le n} = (0 \ldots 1 \ldots 0)^\top \in \mathbb{R}^n$, here with $n = 2$. We show that $\{e_1, e_2\}$ is a basis of V by verifying the two conditions in Definition 1.25.

i) We show that e_1, e_2 are linearly independent. From

$$x_1 e_1 + x_2 e_2 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \overset{!}{=} 0$$

we easily conclude that $x_1 = 0 = x_2$, so that e_1, e_2 are indeed linearly independent.

ii) By Definition 1.4 any $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in V = \mathbb{R}^2$ can be written as

$$v = v_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = v_1 e_1 + v_2 e_2.$$

Thus

$$V = \mathbb{R}^2 = \{v_1 e_1 + v_2 e_2 \colon v_1, v_2 \in \mathbb{R}\} = \mathrm{span}(e_1, e_2).$$

In terms of the previous slide, we have that v_1, v_2 are the coordinates of v w.r.t the basis $\{e_1, e_2\}$ and the dimension of V is 2, we write $\dim(V) = 2$.

***Remark**: Analogue results hold true for any $\mathbb{R}^n$ (not just $n = 2$) and the set of vectors $\{e_1, \ldots, e_n\}$ is called the standard basis or unit vectors in $\mathbb{R}^n$.*

1b) Let us find another basis for $\mathbb{R}^2$ and check whether its length is still 2 in accordance with the remarks after Definition 1.25. For instance, let us consider the vectors

$$a_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \; a_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \in V.$$

Again, we verify that the two conditions in Definition 1.25 are satisfied:

i) Now let us use the equivalent formulation of linear independence from Lemma 1.22. For this purpose let $v \in \text{span}(a_1, a_2)$ so that there exist scalars $\lambda_1, \lambda_2 \in \mathbb{R}$ with

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \lambda_1 a_1 + \lambda_2 a_2,$$

which, after inserting the precise numbers for a_1 and a_2, is equivalent to

$$\begin{pmatrix} \lambda_1 \\ 0 \end{pmatrix} + \begin{pmatrix} \lambda_2 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

Using matrix notation we can even write

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

In order to apply Lemma 1.22, we need to show that the scalars λ_1, λ_2 are uniquely determined by this equation. Therefore, let us now solve this upper triangular system (we will later learn about *backward substitution* to do this algorithmically). We observe from the bottom equation that $\lambda_2 = v_2$. Inserting this into the top equation then yields

$$\lambda_1 + \lambda_2 = v_1 \quad \Leftrightarrow \quad \lambda_1 + v_2 = v_1 \quad \Leftrightarrow \quad \lambda_1 = v_1 - v_2.$$

Observe that λ_1 and λ_2 are uniquely determined, i.e., there are no other λ_1 and λ_2 solving the upper equations; thus a_1, a_2 are independent by Lemma 1.22. Also, let us make a quick test:

$$(v_1 - v_2)a_1 + v_2 a_2 = \begin{pmatrix} v_1 - v_2 \\ 0 \end{pmatrix} + \begin{pmatrix} v_2 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

ii) We note that $\text{span}(a_1, a_2) \subset \mathbb{R}^2 = V$ is obviously true and to prove the reverse subset relation we choose the scalars $\lambda_1 = v_1 - v_2, \; \lambda_2 = v_2$ for $v = (v_1, v_2)^\top \in V = \mathbb{R}^2$.

All in all, we have that $\lambda_1 = v_1 - v_2, \; \lambda_2 = v_2$ are the coordinates of $v \in V$ w.r.t the basis $\{a_1, a_2\}$ of V and $\dim(V) = 2$.

Remark: *The notation for vectors introduced in D. 1.2 implicitly assumes that vectors are represented in the standard basis.*

In the exercises we will prove that for any matrix $A \in \mathbb{F}^{m\times n}$, the kernel $\ker(A)$ is a subspace of $\mathbb{F}^n$ and the image $\mathrm{Im}(A)$ is a subspace of $\mathbb{F}^m$. In the context of matrices these are important spaces and we give their dimensions a special name:

Definition 1.27 (*rank and nullity*) *Let $A \in \mathbb{F}^{m\times n}$. Then*

- $\mathrm{rank}(A) := \dim(\mathrm{Im}(A))$ *is called the (column)* ***rank of*** A,
- $\mathrm{nullity}(A) := \dim(\ker(A))$ *is called the* ***nullity of*** A.

Example 1.28

1) Let us consider the matrix $A = [a_1, a_2, a_3] := \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}$.

1a) By Definition 1.20 of the image we have $\mathrm{Im}(A) = \mathrm{span}(a_1, a_2, a_3)$. By observing $a_3 = a_2 + a_1$, i.e., a_3 is a linear combination of a_1 and a_2, we even find that $\mathrm{Im}(A) = \mathrm{span}(a_1, a_2)$. Since the vectors a_1, a_2 have been identified to be linearly independent (see Example 1.26), we find by Definition 1.25 that they form a basis for $\mathrm{Im}(A)$. Thus

$$\mathrm{rank}(A) = \dim(\mathrm{Im}(A)) = 2.$$

1b) What about the nullity? We first need to find a basis of the kernel (we will do this by re-writing it as a span of some independent vectors). For this purpose, let $x \in \ker(A)$, which by Definition 1.21 is equivalent to

$$Ax = 0 \quad \Leftrightarrow \quad x_1 + x_2 + 2x_3 = 0, \;\; x_2 + x_3 = 0.$$

Now from the second equation we obtain $x_2 = -x_3$. Let us also write x_1 as a function of x_3. This is achieved by inserting $x_2 = -x_3$ into the first equation to obtain

$$x_1 + x_2 + 2x_3 = x_1 - x_3 + 2x_3 = x_1 + x_3 = 0 \quad \Leftrightarrow \quad x_1 = -x_3.$$

Thus we find

$$Ax = 0 \quad \Leftrightarrow \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -x_3 \\ -x_3 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}.$$

With other words, we can write

$$\ker(A) = \{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 \colon x = x_3 \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \} = \{ x_3 \cdot \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} : x_3 \in \mathbb{R} \} = \mathsf{span}(\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}).$$

Since $\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \neq 0$ it forms an independent set of length 1 so that by Definitions 1.25 and 1.27 we finally conclude that

$$\mathrm{nullity}(A) = \dim(\ker(A)) = 1.$$

Remark: *We observe that*

$$\mathrm{rank}(A) + \mathrm{nullity}(A) = 3 \quad \textit{(=column dimension).}$$

We will see below that this is generally true – called the dimension formula!

2) Let us consider $A = [a_1, a_2] := \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 0 \end{pmatrix}$.

2a) Since $a_2 = 2 \cdot a_1$, the columns are certainly linearly dependent (e.g., $2 \cdot a_1 + (-1)a_2 = 0 \in \mathbb{R}^2$; a combination that yields zero but with nonzero coefficients). Therefore

$$\mathrm{Im}(A) = \text{span}(a_1, a_2) = \text{span}(a_1),$$

so that

$$\mathrm{rank}(A) = \dim(\mathrm{Im}(A)) = 1.$$

2b) Now let us consider the kernel $\ker(A) = \{x : \; Ax = 0\}$. Following along the lines of the previous slides we get

$$Ax = 0 \quad \Leftrightarrow \quad \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} x_1 + \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} x_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad \begin{pmatrix} x_1 + 2x_2 \\ x_1 + 2x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow \quad x_1 = -2x_2$$

and thus

$$\ker(A) = \{x \in \mathbb{R}^2 : \; x_1 = -2x_2\} = \{\begin{pmatrix} -2x_2 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 : \; x_2 \in \mathbb{R}\} = \{x_2 \begin{pmatrix} -2 \\ 1 \end{pmatrix} \in \mathbb{R}^2 : \; \lambda \in \mathbb{R}\} = \text{span}\left(\begin{pmatrix} -2 \\ 1 \end{pmatrix}\right).$$

So all in all, $\mathrm{nullity}(A) = \dim(\ker(A)) = 1$.

Remark: *Again we observe that*

$$\mathrm{rank}(A) + \mathrm{nullity}(A) = 2 \quad \textit{(=column dimension).}$$

3) Similarly, considering the matrices from above we find $\mathrm{rank}(A_2) = 2$ and $\mathrm{rank}(A_3) = 3$.

Question: Can we find a general relation between the nullity and the rank of a matrix?

Theorem 1.29 (*Dimension Formula/Rank–Nullity Theorem*) *Let $A \in \mathbb{F}^{m\times n}$, then*

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n.$$

- The dimension formula also reads as

$$\dim(\operatorname{Im}(A)) + \dim(\ker(A)) = \dim(\mathbb{F}^n).$$

- **Intuition:** Let us again think of a matrix $A \in \mathbb{F}^{m\times n}$ as a mapping from $\mathbb{F}^n$ to $\mathbb{F}^m$. If the matrix maps some vectors of this n–dimensional space $\mathbb{F}^n$ to 0 – precisely those vectors from the kernel of A – then we can say that this "piece of information" gets lost. What prevails from $\mathbb{F}^n$ makes up the image of A whose dimension is the rank of the matrix by definition. So, the amount of information in $\mathbb{F}^n$ equals the information that gets lost after mapping it by A ($\operatorname{nullity}(A)$) plus the one that prevails ($\operatorname{rank}(A)$).

1.4 Inverse Matrices

Example 1.30 (*Inverses*) Let us consider the following matrix

$$A = [a_1, a_2] = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

which is composed of the vectors considered in Example 1.26 1b).

Recall results Ex.1.26 1b): We have already observed that a_1, a_2 are independent and $\operatorname{span}(a_1, a_2) = \mathbb{R}^2$. With other words, for any $b \in \mathbb{R}^2$, by Lemma 1.22 there exist unique (!) scalars x_1, x_2, so that $Ax = x_1 a_1 + x_2 a_2 = b$. More precisely, for $x_b = \begin{pmatrix} b_1 - b_2 \\ b_2 \end{pmatrix}$ (coordinates of b wrt. the basis $\{a_1, a_2\}$) we found

$$Ax_b = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 - b_2 \\ b_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = b.$$

Clearly, the vector x_b is composed of information from b. Now let us consider the following mapping

$$b \mapsto x_b = \begin{pmatrix} b_1 - b_2 \\ b_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} b_1 + \begin{pmatrix} -1 \\ 1 \end{pmatrix} b_2 = \underbrace{\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}}_{=:A^{-1}} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

We observe that the mapping from the vector b to its coordinates x_b w.r.t. the basis $\{a_1, a_2\}$ can be expressed as a matrix–vector product. The involved matrix A^{-1} is referred to as the *inverse matrix* of A.

In general:

Consider the matrix as a mapping

$$f_A : \mathbb{F}^n \to \mathbb{F}^n,\ x \mapsto Ax.$$

Then by definition the *mapping* f_A is **invertible**, if there exists a mapping $f_A^{-1} : \mathbb{F}^n \to \mathbb{F}^n$ such that for all $x, b \in \mathbb{F}^n$ we have

$$f_A(x) = b \quad \Leftrightarrow \quad x = f_A^{-1}(b).$$

Inserting the definition of f_A this reads as

$$Ax = b \quad \Leftrightarrow \quad x = A^{-1}b.$$

Verifying this condition for all possible x and b would be an ambitious endeavor. Luckily, this condition can be rephrased into conditions solely involving the matrix A. More precisely, by inserting one into the other we obtain

i) $Ax = b \quad \Leftrightarrow \quad AA^{-1}b = b \quad \Leftrightarrow \quad AA^{-1} = I$,

ii) $x = A^{-1}b \quad \Leftrightarrow \quad x = A^{-1}Ax \quad \Leftrightarrow \quad A^{-1}A = I$.

Let us quickly check this for Example 1.30:

$$A^{-1}A = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2 \ \checkmark, \qquad AA^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2 \ \checkmark.$$

Let us make this a definition.

Definition 1.31 (*Inverse matrix*) *A matrix $A \in \mathbb{F}^{n\times n}$ is called* **invertible**, *if there exists a matrix $\tilde{A} \in \mathbb{F}^{n\times n}$ with*

$$A \cdot \tilde{A} = \tilde{A} \cdot A = I_n. \tag{1}$$

In case of existence we find that $\tilde{A}$ is unique (see below) and we denote by $A^{-1} := \tilde{A}$ the **inverse matrix** *of A. The set of all invertible matrices in $\mathbb{F}^{n\times n}$ is denoted by $GL_n(\mathbb{F})$, the so-called* general linear group.

Consider the linear equation

$$Ax = b$$

By setting $x := A^{-1}b$, we find

$$Ax = AA^{-1}b = AA^{-1}b = I_n b = b.$$

If A is invertible, then

$$\text{"solving } Ax = b\text{"} \quad = \quad \text{"applying the inverse matrix } A^{-1}\text{"}$$

(*numerical methods*) (*not accessible in practice*)

From the dimension formula 1.29 for $n = m$, we find "**injectivity** $=$ **surjectivity**"

$$A = \begin{pmatrix} | & & | \\ a_1 & - & a_n \\ | & & | \end{pmatrix} \in \mathbb{F}^{n \times n}, \qquad f_A : \mathbb{F}^n \to \mathbb{F}^n, \qquad x \mapsto Ax$$

mapping f_A	matrix A	vectors (columns) $a_1, \dots, a_n$

$$\boxed{f_A \text{ injective}} \overset{\text{Ex.}}{\Longleftrightarrow} \boxed{\ker A = \{0\}} \Longleftrightarrow \boxed{a_1, \dots, a_n \text{ are independent}}$$

$$\Updownarrow$$

$$\dim(\ker A) = 0$$

$$\Updownarrow \quad \text{[DIMENSION FORMULA]}$$

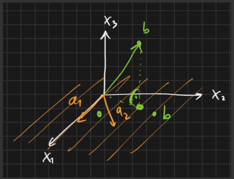
$$\dim(\operatorname{Im} A) = n - \dim(\ker A) = n$$

$$\Updownarrow$$

$$\boxed{f_A \text{ surjective}} \Longleftrightarrow \boxed{\underbrace{\operatorname{Im} A}_{f_A(\mathbb{F}^n)} = \mathbb{F}^n} \Longleftrightarrow \boxed{\operatorname{span}(a_1, \dots, a_n) = \mathbb{F}^n}$$

Also see Lemma 1.22

Remark:
A System $Ax = b$ can be solvable even if A is not squared (and thus not invertible)!



The Difference:

- *invertible* $(m = n)$: For **any** b there is a **unique** x so that $Ax = b$, i.e.,

$$A \text{ is invertible} \quad \Rightarrow \quad \text{we always have that } b \in \mathrm{Im}(A)$$

  This unique x is given by $A^{-1}b$.

- *solvable* ($m \neq n$ allowable): Given a **fixed** b we find **at least one** x so that $Ax = b$, i.e.,

$$Ax = b \text{ is solvable} \quad \Leftrightarrow \quad b \in \mathrm{Im}(A)$$

  We will learn later that in some cases this x is given by $(A^\top A)^{-1}A^\top b$.

Thus:

$$\text{invertible} \quad \Rightarrow \quad \text{solvable}$$

Exercise:

Lemma 1.32 (*Porperties of inverse matrices*) *We find the following properties:*

i) *The inverse A^{-1} is also invertible, with inverse $\left(A^{-1}\right)^{-1} = A$.*
ii) *Since $\mathbb{F}$ is a field, any left-inverse is also a right-inverse and vice-versa.*
iii) *An invertible matrix $A \in \mathbb{F}^{n\times n}$ has exactly one inverse matrix.*
iv) *The product of two invertible matrices, say A and B, is invertible with inverse*

$$(AB)^{-1} = B^{-1}A^{-1}.$$

v) *A diagonal matrix*

$$D = \mathit{diag}(d_1, \dots, d_n) = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} \in \mathbb{F}^{n\times n}$$

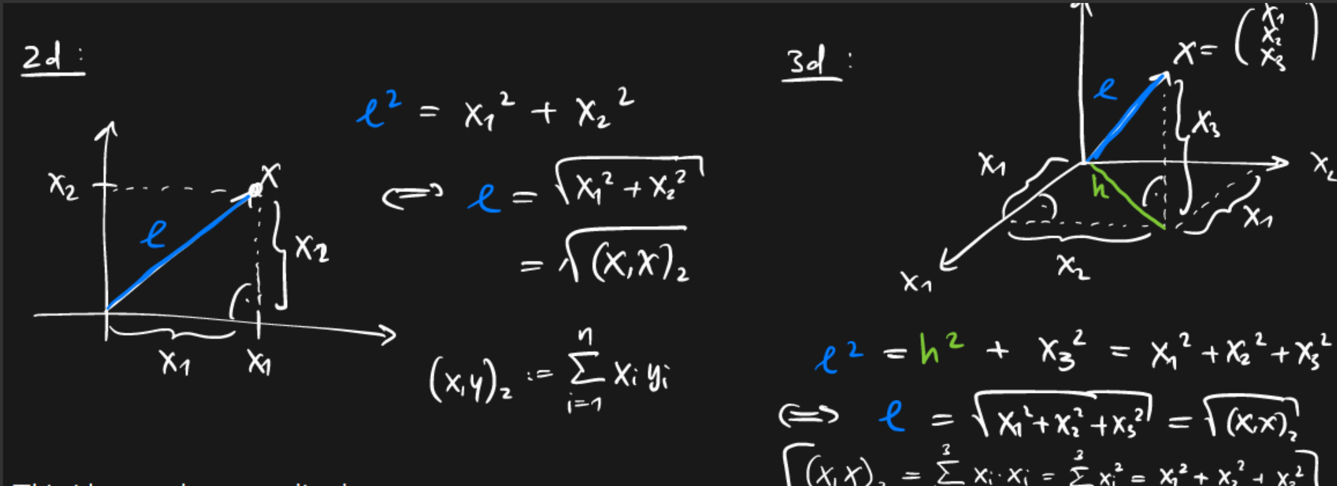
is invertible if and only if $d_i \neq 0$ for all $i = 1, \dots, n$. Its inverse is given by

$$D^{-1} = \mathit{diag}(d_1^{-1}, \dots, d_n^{-1}) = \begin{pmatrix} d_1^{-1} & & \\ & \ddots & \\ & & d_n^{-1} \end{pmatrix} \in \mathbb{F}^{n\times n}.$$

In particular, $(GL_n(\mathbb{F}), \cdot)$ is a group.

1.5 The Euclidean Norm

Let us first consider the 2d and 3d case:



This idea can be generalized to:

Definition 1.33 (*Euclidean Norm*) *The Euclidean norm of a vector $x \in \mathbb{F}^n$ is defined by*

$$\|x\|_2 := \sqrt{\sum_{i=1}^{n} |x_i|^2} = \sqrt{x^H x}$$

where $|a + ib|^2 := a^2 + b^2$ denotes the absolute value of a complex number. For a real vector $x \in \mathbb{R}^n$ this simplifies to $\|x\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^\top x}$.

$\rightarrow$ We will also get to know other "norms" (e.g., Manhattan norm or maximum norm).

Relating the inner product to projections

Let us consider $\mathbb{F} = \mathbb{R}$. As a special case of the so-called **Cauchy Schwarz inequality** one can show that, for any two real vectors $x, y \in \mathbb{R}^n$,

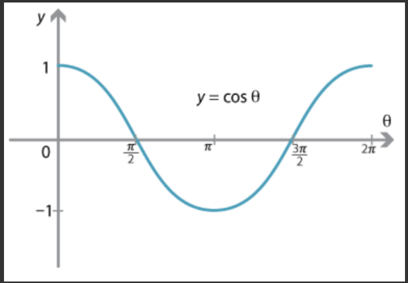
$$\left|x^T y\right| \leq \|x\|_2 \cdot \|y\|_2.$$

This is equivalent to (assumed both vectors are nonzero, otherwise trivial case)

$$-1 \leq \frac{x^T y}{\|x\|_2 \cdot \|y\|_2} = \left(\frac{x}{\|x\|_2}\right)^T \left(\frac{y}{\|y\|_2}\right) \leq 1.$$

Since $\cos\colon (0,\pi) \to (-1,1)$ is bijective, we find an uniquely defined angle $\alpha \in (0,\pi)$, so that

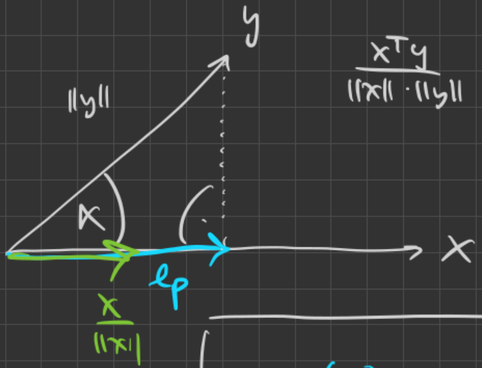
$$\cos(\alpha) = \frac{x^T y}{\|x\|_2 \cdot \|y\|_2} \quad (\in (-1,1)).$$



We also use the notation $\alpha := \sphericalangle(x,y)$, since α can be considered´ the angle between x and y.

Geometric insights from the identity

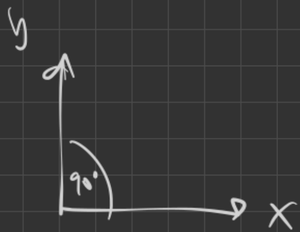
$$\text{cosine} = \frac{\text{adjacent}}{\text{hypotenuse}}.$$



$$\frac{x^T y}{\|x\| \cdot \|y\|} = \cos(\alpha) = \frac{\ell p}{\|y\|} \iff$$

$$\ell p = \frac{x^T y}{\|x\|}$$

$$\boxed{\mathrm{proj}_x(y) = \frac{x}{\|x\|} \cdot \ell p = \left(\frac{x^T y}{\|x\|}\right) \cdot \frac{x}{\|x\|}}$$



1.6 Orthogonal Vectors and Matrices

Let us again consider the relation

$$\cos(\alpha) = \frac{x^T y}{\|x\|_2 \cdot \|y\|_2}, \quad x, y \in \mathbb{R}^n.$$

Now let us assume that the angle $\alpha = \sphericalangle(x, y)$ between the two vectors x, y is 90°, i.e., $\alpha = \pm\frac{\pi}{2}$, meaning that they are *perpendicular*. Then we find

$$0 = \cos\left(\pm\frac{\pi}{2}\right) = \frac{x^T y}{\|x\|_2 \cdot \|y\|_2} \qquad \Leftrightarrow \qquad 0 = x^\top y.$$

In mathematics we call this *orthogonal* and make it a general definition:

Definition 1.34 (*Orthogonal/-normal vectors*)

i) *Two vectors $x, y \in \mathbb{F}^n$ are called **orthogonal** if $(x, y)_2 = x^H y = 0$.*

ii) *Two vectors $x, y \in \mathbb{F}^n$ are called **orthonormal** if they are orthogonal and have length 1 (i.e., $\|x\|_2 = \|y\|_2 = 1$).*

iii) *Vectors $x_1, \ldots, x_r \in \mathbb{F}^n$ are called (mutually) **orthogonal (orthonormal)** if x_i, x_j are **orthogonal (orthonormal)** for all possible pairs $i \neq j \in \{1, \ldots, r\}$.*

One can show that:

$$x, y \ \text{ orthogonal} \quad \Rightarrow \quad x, y \ \text{ linearly independent.} \tag{2}$$

Counter example for backwards implication: $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, y = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Example 1.35

For $\mathbb{F} = \mathbb{R}$ and $n = 2$ consider, e.g.,

- Standard basis vectors.
- Rotation of the standard basis vectors.

Now let us extend this notion to matrices:

For this purpose observe that the matrix-matrix product $Q^H Q$ for $Q \in \mathbb{F}^{n\times n}$ contains all possible inner products of its columns:

$$\underbrace{\begin{pmatrix} - & \overline{q_1}^\top & - \\ - & \overline{q_2}^\top & - \\ & \vdots & \\ - & \overline{q_n}^\top & - \end{pmatrix}}_{Q^H} \underbrace{\begin{pmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{pmatrix}}_{Q} = \begin{pmatrix} q_1^H q_1 & \cdots & q_1^H q_n \\ q_2^H q_1 & \cdots & q_2^H q_n \\ \vdots & \ddots & \vdots \\ q_n^H q_1 & \cdots & q_n^H q_n \end{pmatrix} \quad \left(= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \right)$$

Let us assume that the columns of Q are mutually ortho*normal*, then

$$Q^H Q = I_n.$$

Since this is a central property, we make this a definition:

Definition 1.36 (*Orthogonal/Unitary matrix*) *A matrix $Q \in \mathbb{F}^{n\times n}$ is called* ***unitary****, if*

$$Q^H Q = I_n.$$

For a real matrix $Q \in \mathbb{R}^{n\times n}$ this condition simplifies to $Q^T Q = I_n$, in which case we then call the matrix ***orthogonal****.*

Since orthogonality implies linear independence (see statement (2)) we know that **orthogonal matrices are invertible**. From the defining equation $Q^T Q = I_n$ we can even deduce its inverse

$$Q^{-1} = Q^T$$

and therefore also $QQ^T = I_n$.

$\rightarrow$ This is one (of the many) reasons why the property of orthogonality is very desirable.

Understanding $QQ^\top(\cdot)$ as orthogonal projection

For a vector $q \in \mathbb{R}^n$ of length 1, i.e., $\|q\|_2 = 1$, and a vector $y \in \mathbb{R}^n$ we find

$$\mathrm{proj}_q(y) = (q^\top y) \cdot q.$$

Now let $Q = [q_1, \ldots, q_n] \in \mathbb{R}^{n \times n}$ be an orthogonal matrix (i.e., columns q_i are mutually orthonormal), then

$$y = I \cdot y = QQ^\top y = Q \begin{pmatrix} q_1^\top y \\ \vdots \\ q_n^\top y \end{pmatrix} = \sum_{i=1}^{n} q_i^\top y \cdot q_i = \sum_{i=1}^{n} \mathrm{proj}_{q_i}(y).$$

With other words, in order to obtain the coordinates of y with respect to the *orthonormal* basis $\{q_1, \ldots, q_n\}$ we solely have to project y onto each basis vector q_i.

Example

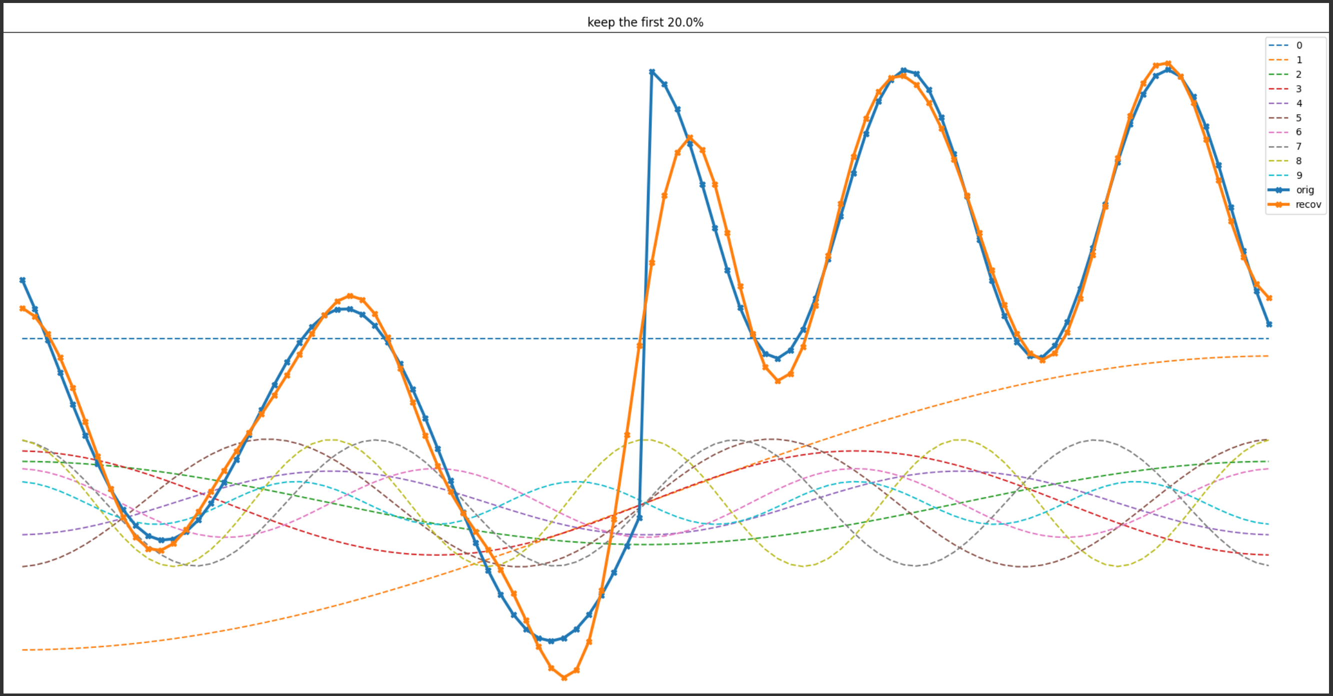
Famous related examples from signal processing include the Discrete Cosine Transform (DCT) and the Discrete Fourier Transform (DFT) which can be written as a matrix–vector product $Q^\top(\cdot)$ with an orthogonal/unitary matrix Q. In this context, the q_i may correspond to discrete periodic functions of different frequency. For a time-discrete signal

$$y = (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$$

one says that the transformed signal

$$Q^\top y = (q_1^\top y, \ldots, q_n^\top y)^\top$$

lives in the frequency space.



1-d DCT compression example (where high frequencies are removed):

$$y = \sum_{i=1}^{n} q_i^\top y \cdot q_i \approx \sum_{i=1}^{m} q_i^\top y \cdot q_i \quad (m < n).$$

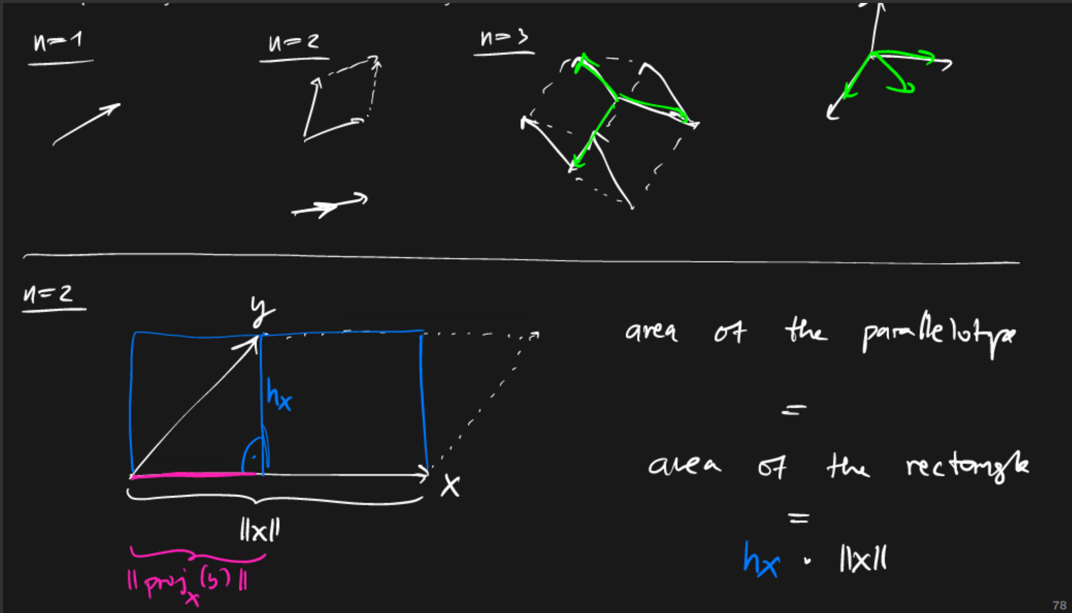
2-d DCT compression example (where high frequencies are removed)

1.7 The Determinant

Aim: For n vectors in $\mathbb{F}^n$ we want to have a *measure of linear independence*
– or equivalently a *volume measure* for the parallelotope spanned by these vectors
– or equivalently a *measure for the invertibility* of a matrix in $\mathbb{F}^{n\times n}$

Why are all these measures the same?

- n linear dependent vectors do not span a volume in $\mathbb{F}^n$.
- Linear independent columns of a quadratic matrix imply invertibility.



Let us derive a formula in the two–dimensional case: Let

$$a := \text{area of the parallelogram} = \text{area of rectangle} = h_x \cdot \|x\|_2.$$

By the Pythagorean identity we obtain:

$$\begin{aligned} h_x^2 + \|\mathrm{proj}_x(y)\|_2^2 &= \|y\|_2^2 \\ \Leftrightarrow h_x^2 &= \|y\|_2^2 - \|\mathrm{proj}_x(y)\|_2^2 = \|y\|_2^2 - \frac{(x^T y)^2}{\|x\|_2^2} \\ &= \frac{\|y\|_2^2 \|x\|_2^2 - (x^T y)^2}{\|x\|_2^2} \\ &= \frac{(y_1^2 + y_2^2)(x_1^2 + x_2^2) - (x_1 y_1 + x_2 y_2)^2}{\|x\|_2^2} \\ &= \dots \\ &= \frac{(x_1 y_2 - x_2 y_1)^2}{\|x\|_2^2}. \end{aligned}$$

Thus

$$h_x = \left| \frac{x_1 y_2 - x_2 y_1}{\|x\|_2} \right| = \frac{|x_1 y_2 - x_2 y_1|}{\|x\|_2},$$

which implies

$$a = |x_1 y_2 - x_2 y_1| =: |\det(A)|.$$

In general, there is the following (recursive) formula, which we use as the definition here:

Definition 1.37 (*Laplace formula*) *Let $A \in \mathbb{F}^{n\times n}$ and let $A_{ij} \in \mathbb{F}^{(n-1)\times(n-1)}$ be the matrix resulting from erasing the i-th row and j-th column. Then the mapping* $\det\colon \mathbb{F}^{n\times n} \to \mathbb{F}$ *defined by*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}), \quad \text{for a fixed but arbitrary } \ i \in \{1,\dots,n\},$$

*is called the **determinant** (of A), where $\det(a) := a$ for $a \in \mathbb{R} = \mathbb{R}^{1\times 1}$.*

One can show: The determinant is a well-defined function, i.e., by the formula above the function $\det(\cdot)$ assigns to each matrix $A \in \mathbb{F}^{n\times n}$ exactly one number in $\mathbb{F}$.

Laplace formula for $n = 2$ and $n = 3$:

- $n = 2$ (we fix $i = 1$)

Here we have

$$\det(A) = \sum_{j=1}^{2} (-1)^{1+j} a_{1j} \det(A_{1j})$$

$$A_{11} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = [a_{22}], \ A_{12} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = [a_{21}], \ A_{21} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = [a_{12}], \ A_{22} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = [a_{11}]$$

So that all in all

$$\det(A) = (-1)^{1+1} a_{11} \det(A_{11}) + (-1)^{1+2} a_{12} \det(A_{12}) = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$$

- $n = 3$: *Sarrus rule* (exercise)

One can show:

Theorem 1.38 (*Determinant properties*) *The determinant satisfies the following computational rules:*

i) $\forall A \in \mathbb{F}^{n\times n}: \quad \det(A) \neq 0 \Leftrightarrow A \in GL(n,\mathbb{F})$ ($\Leftrightarrow$ *columns of A are linearly independent*)

ii) $\forall A \in \mathbb{F}^{n\times n}: \quad \det(A^\top) = \det(A)$

iii) *if $A \in \mathbb{F}^{m\times m}, B \in \mathbb{F}^{m\times n}, C \in \mathbb{F}^{n\times n}$ and*

$$M := \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \in \mathbb{F}^{(m+n)\times(m+n)}$$

then $\det M = \det A \cdot \det C$

iv) $\forall A, A' \in \mathbb{F}^{n\times n}: \quad \det(A \cdot A') = \det(A) \cdot \det(A')$

The central result for us is i).

Question: Are there matrices for which the computation of the determinant is easy?

Yes, as in many other situations it turns out that orthogonal and triangular matrices are easy to treat! More precisely, we find:

Corollary 1.39 (*Triangular matrices*) *Let $U \in \mathbb{F}^{n\times n}$ be upper triangular, i.e.,*

$$U = \begin{pmatrix} u_{11} & x & \cdots & x \\ 0 & u_{22} & & \vdots \\ \vdots & & \ddots & x \\ 0 & \cdots & 0 & u_{nn} \end{pmatrix}.$$

Then

$$\det(U) = u_{11} \cdot u_{22} \cdot \ldots \cdot u_{nn}.$$

In particular, we find

$$U \text{ is invertible} \quad \Leftrightarrow \quad \det(U) \neq 0 \quad \Leftrightarrow \quad \forall i: \; u_{ii} \neq 0$$

Proof. Exercise: For the product formula apply Theorem 1.38 iii) inductively. The second part then easily follows from Theorem 1.38 i). □

Corollary 1.40 (*Orthogonal matrices*) *Let $Q \in \mathbb{R}^{n\times n}$ be an orthogonal matrix, then $|\det(Q)| = 1$.*

Proof. From Cor. 1.39 we find $\det(I) = 1$. Then result follows from Theorem 1.38 ii) and iv). □

1.8 Linear Systems of Equations

Aim:

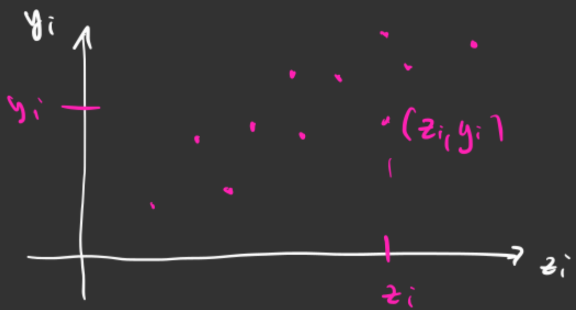
Given $A \in \mathbb{R}^{m \times n}$ ($m \neq n$ possible) and $b \in \mathbb{R}^m$, find $x \in \mathbb{R}^n$ such that

$$Ax = b.$$

1.8.1 Motivation: Curve Fitting

As a motivating example let us consider *curve fitting*.

Assume we are given $m \in \mathbb{N}$ measurements $(z_1, y_1), \ldots, (z_m, y_m) \in \mathbb{R}^2$ (or more generally in any product space, say $Z \times Y$)



Question: Is there a "significant" relation between the z_i and y_i?

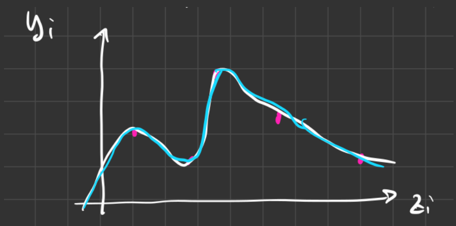
Let us consider the z_i as input (independent/explanatory/exogenous) variable, and the y_i as output (dependent/predicted/response...) variable.

Examples: $z_i =$ (temperature, light intensity), $y_i =$ plant height or $z_i =$ year, $y_i =$ global mean temperature

Mathematically asking: Is there a function f, so that $f(z_i) \cong y_i$ for all $i = 1, \ldots, m$?

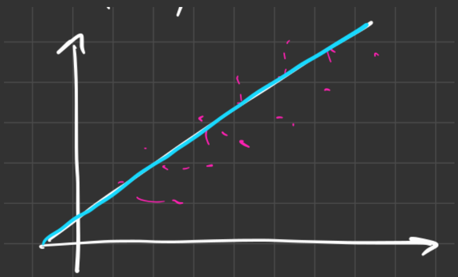
Exact fit: **Interpolation**

$$f(z_i) = y_i \;\forall i = 1, \ldots, m$$



Approximate fit: **Regression/Smoothing**

$$f(z_i) \approx y_i \;\forall i = 1, \ldots, m$$



In order to find such a fit, we need to restrict ourselves to certain classes of functions f. With other words we need to assume a certain "model":

$$z_i \overset{f}{\mapsto} y_i$$

In this course we will consider models of the following kind:

$$f : \mathbb{R} \to \mathbb{R}, \quad f(z) := \sum_{k=1}^{n} x_k f_k(z)$$

$\rightarrow$ More precisely, we assume that the relation between the z_i and the y_i can be modeled by a *linear combination* of $\underbrace{\text{some functions } f_k : \mathbb{R} \to \mathbb{R}}_{\textit{given by assumption}}$ with $\underbrace{\text{some coefficients/parameters } x_k}_{\textit{to be determined } (f(z_i) \cong y_i)}$

(More generally $f, f_k \colon \mathbb{R}^k \to \mathbb{R}$. Important here is the fact that our model f is linear combined from the f_k.)

Example 1.41 (*Polynomial Interpolation/Regression*) One often considers a polynomial model:

$$f_k(z) := z^{k-1}, \quad \text{so that} \quad f(z) = x_1 + x_2 z + x_3 z^2 + \cdots + x_n z^{n-1}$$

For example, if $n = 2$ then $f(z) = x_1 + x_2 z$ (an affine linear model).

How does this translate into a linear system "$Ax \cong b$"?

For all measurements $(z_1, y_1), \dots, (z_m, y_m)$ we require:

$$\sum_{k=1}^{n} x_k f_k(z_i) = f(z_i) \cong y_i \quad \text{for all } i = 1, \dots, m$$

Writing these equations row by row for each i-th measurement gives:

$$\begin{array}{ccc} i = 1: & x_1 f_1(z_1) + x_2 f_2(z_2) + \cdots + x_n f_n(z_1) & = y_1 \\ \vdots & \vdots & \vdots \\ i = m: & x_1 f_1(z_m) + x_2 f_2(z_m) + \cdots + x_n f_n(z_m) & = y_m \end{array}$$

Using matrix notation, this system can be written as:

$$\underbrace{\begin{pmatrix} f_1(z_1) & \cdots & f_n(z_1) \\ \vdots & \ddots & \vdots \\ f_1(z_m) & \cdots & f_n(z_m) \end{pmatrix}}_{=:A \in \mathbb{R}^{m \times n}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{=:x \in \mathbb{R}^{n}} \cong \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}}_{=:b \in \mathbb{R}^{m}} \quad \Leftrightarrow \quad Ax \cong b$$

$\rightarrow$ Also revisit Example 1.1.

Exact: **Interpolation**

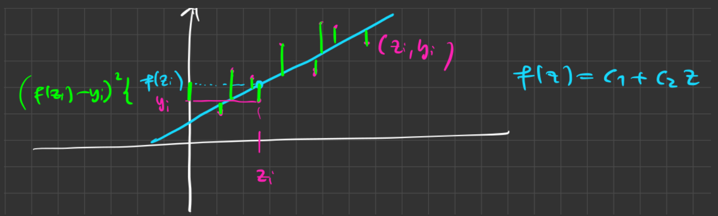
$$Ax = b$$

Approximate: **Regression**

$$Ax \approx b$$

A common approach to address a regression problem is a linear least squares formulation:

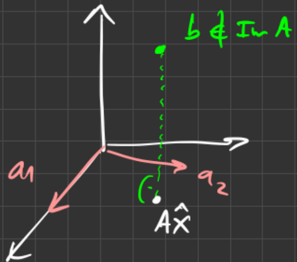
$$\begin{aligned}\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2 &= \sum_{i=1}^{m}(Ax-b)_i^2 \\ &= \sum_{i=1}^{m}(f(z_i)-y_i)^2\end{aligned}$$



1.8.2 Existence and Uniqueness Analysis

Let us consider the following cases:

$$A = \begin{pmatrix} | & | \\ a_1 & a_2 \\ | & | \end{pmatrix} \in \mathbb{R}^{3\times 2} \qquad Ax = b$$



(no solution to $Ax = b$)

$$\hat{x} := \underset{x\in\mathbb{R}^2}{\arg\min} \|Ax-b\|_2^2$$

unique solution x^*, $Ax^* = b$

a solution x, $Ax = b$

Summary

Aim:

$$\textit{Given } A \in \mathbb{R}^{m\times n} \textit{ (} m \neq n \textit{ possible) and } b \in \mathbb{R}^m\textit{, find } x \in \mathbb{R}^n \textit{ such that}$$
$$Ax = b.$$

Here:

$$m \;=\; \sharp \text{ equations} = \sharp \text{ measurements} = \text{length of the column vectors}$$
$$n \;=\; \sharp \text{ unknowns} = \sharp \text{ parameters} = \sharp \text{ columns}$$

$$A = \begin{array}{cc} & n \\ m & \begin{pmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{pmatrix} \end{array}, \quad \text{with image } \mathrm{Im}(A) = \{Ax :\in \mathbb{R}^n\} \;=\; \mathrm{span}(a_1, \ldots, a_n)$$

Let us define the solution set

$$S := \{x \in \mathbb{R}^n : Ax = b\} \;=\; f_A^{-1}(\{b\}),$$

then there are three possible states, namely,

$$|S| = \begin{cases} 0 \;:\; \text{“no solution”, if } \; b \notin \mathrm{Im}(A) \\ 1 \;:\; \text{“unique solution”, if } \; b \in \mathrm{Im}(A) \text{ and independent columns } (\ker(A) = \{0\}) \\ \infty \;:\; \text{“infinitely many solutions”, if } \; b \in \mathrm{Im}(A) \text{ and dependent columns } (\ker(A) \neq \{0\}) \end{cases}$$

For a given b, observe the relations between image and existence as well as kernel and uniqueness. In fact, $b \in$ or $\notin \mathrm{Im}(A)$ decides if solutions **exist** and ker(A) $=$ or $\neq$ $\{0\}$ gives the solutions' **uniqueness**.

Quick clarification:

Let $\ker(A) \neq \{0\}$, then there exists a $w \in \mathbb{R}^n$ so that

$$A(\alpha w) = 0$$

for *all* scalars $\alpha \in \mathbb{R}$.

If $b \in \mathrm{Im}(A)$, then there exists an $x \in \mathbb{R}^n$, so that

$$Ax = b.$$

Adding these two equations gives

$$A(x + \alpha w) = b \quad \forall \alpha \in \mathbb{R}.$$

With other words, we find infinitely many solutions.

Another equivalent argument in view of Lemma 1.22: A nontrivial kernel implies that the columns are linearly dependent and thus vectors in their span are not uniquely combined. Now we see that this already implies infinitely many ways to linearly combine the columns of A to obtain b (assumed b lies in there span).

1.9 More on Image and Kernel

Let us fix $\mathbb{F} = \mathbb{R}$ in this section. In this subsection we derive some more results on the kernel

$$\ker(A) = \{x \in \mathbb{R}^n : Ax = 0\} \subset \mathbb{R}^n$$

and the image

$$\operatorname{Im}(A) = \{Ax : x \in \mathbb{R}^n\} \subset \mathbb{R}^m.$$

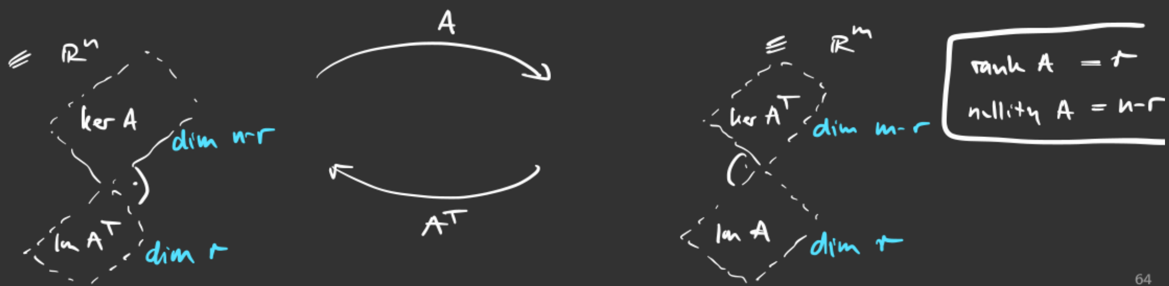
These results prove useful in later sections; in particular when we talk about the singular value decomposition.

The Four Fundamental Subspaces

In the context of a matrix $A \in \mathbb{R}^{m\times n}$ there are four subspaces that stand out:

$$\begin{aligned} \ker(A) &\quad\perp\quad \operatorname{Im}(A^\top) \\ \operatorname{Im}(A) &\quad\perp\quad \ker(A^\top). \end{aligned}$$

The big picture of linear algebra:



Example 1.42 Let us consider

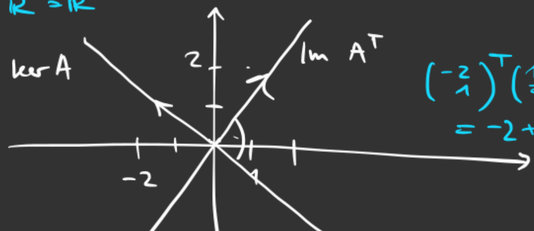
$$A = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}, \qquad A^\top = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix}$$

Then we find

$$\begin{aligned} \mathrm{Im}(A) &= \mathrm{span} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \\ \ker(A) &= \{x \in \mathbb{R}^2 : Ax = 0\} \\ &= \{x \in \mathbb{R}^2 : x_1 \begin{pmatrix} 1 \\ 3 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 6 \end{pmatrix} = 0\} \\ &= \{x \in \mathbb{R}^2 : 1x_1 + 2x_2 = 0\} \\ &= \{x \in \mathbb{R}^2 : x_1 = -2x_2\} \\ &= \mathrm{span} \begin{pmatrix} -2 \\ 1 \end{pmatrix} \end{aligned}$$

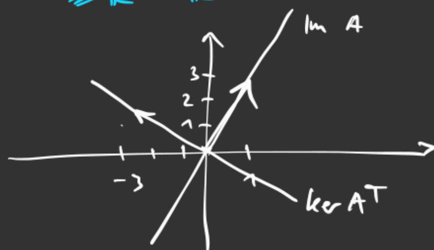
$$\begin{aligned} \mathrm{Im}(A^\top) &= \mathrm{span} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\ \ker(A^\top) &= \{x \in \mathbb{R}^2 : Ax = 0\} \\ &= \{x \in \mathbb{R}^2 : 1x_1 + 3x_2 = 0\} \\ &= \{x \in \mathbb{R}^2 : x_1 = -3x_2\} \\ &= \mathrm{span} \begin{pmatrix} -3 \\ 1 \end{pmatrix} \end{aligned}$$

$\equiv \mathbb{R}^n = \mathbb{R}^2$



$\begin{pmatrix} -2 \\ 1 \end{pmatrix}^\top \begin{pmatrix} 1 \\ 2 \end{pmatrix} = -2 + 2 = 0$

$\equiv \mathbb{R}^m = \mathbb{R}^2$



We need another definition:

Definition 1.43 (*Orthogonal subspaces*) *Let $U, V \subset \mathbb{R}^n$ be two subspaces.*

i) *We call U and V **orthogonal** ($U \perp V$) if $u^\top v = 0$ for all $u \in U, v \in V$.*

ii) *We call*

$$U^\perp := \{x \in \mathbb{R}^n \colon x^\top u = 0 \ \ \forall\, u \in U\}$$

*the **orthogonal complement** of V in $\mathbb{R}^n$.*

Exercise: Show that $(U^\perp)^\perp = U$ and $U \perp U^\perp$.

Example 1.44

i) $n = 2$, $U := \mathrm{span}(\begin{pmatrix} 1 \\ 0 \end{pmatrix}), V := \mathrm{span}(\begin{pmatrix} 0 \\ 1 \end{pmatrix})$. Then

$$\forall u = \begin{pmatrix} u_1 \\ 0 \end{pmatrix} \in U, v = \begin{pmatrix} 0 \\ v_2 \end{pmatrix} \in V : u^\top v = u_1 \cdot 0 + 0 \cdot v_2 = 0.$$

Thus, $U \perp V$.

ii) $n = 3$, $U := \mathrm{span}(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix})$. Thus for any $u \in U$ we have $u = \begin{pmatrix} u_1 \\ u_2 \\ 0 \end{pmatrix}$. Then

$$\begin{aligned} U^\perp &= \{x \in \mathbb{R}^3 \colon x^\top u = 0 \ \ \forall\, u \in U\} = \{x \in \mathbb{R}^3 \colon x_1 u_1 + x_2 u_2 = 0 \ \ \forall\, u_1, u_2 \in \mathbb{R}\} \\ &= \{x \in \mathbb{R}^3 \colon x_1 = x_2 = 0\} \qquad \textit{(choose } u_1 = 1, u_2 = 0 \textit{ and } u_1 = 0, u_2 = 1\textit{)} \\ &= \mathrm{span}(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}). \end{aligned}$$

[ex:draw pictures]

We now prove the orthogonality relation between the four fundamental subspaces:

Lemma 1.45 *Let $A \in \mathbb{R}^{m\times n}$. Then*

$$\mathrm{Im}(A)^{\perp} = \ker(A^{\top}) \quad \textit{and} \quad \ker(A)^{\perp} = \mathrm{Im}(A^{\top}).$$

*In words, $\ker(A^T)$ is the **orthogonal complement** of $\mathrm{Im}(A)$ in $\mathbb{R}^m$ and $\mathrm{Im}(A^T)$ is the orthogonal complement of $\ker(A)$ in $\mathbb{R}^n$.*

Proof. We show the first equation. The orthogonal complement of $\mathrm{Im}(A)$ can be characterized as

$$\mathrm{Im}(A)^{\perp} = \{y \in \mathbb{R}^m : z^{\top} y = 0 \ \ \forall z \in \mathrm{Im}(A)\} = \{y \in \mathbb{R}^m : x^{\top} A^{\top} y = 0 \ \ \forall x \in \mathbb{R}^n\}. \qquad \textit{(simply write } z = Ax\textit{)}$$

Now we show mutual subset relation. First,

$$\begin{aligned} y \in \mathrm{Im}(A)^{\perp} &\Rightarrow \forall x \in \mathbb{R}^n : x^{\top}(A^{\top} y) = 0 \\ &\Rightarrow \text{for the basis vectors } e_1, \ldots, e_n : e_i^{\top}(A^{\top} y) = (A^{\top} y)_i = 0 \\ &\Rightarrow A^{\top} y = 0 \text{ ,i.e., } \ y \in \ker(A^{\top}). \end{aligned}$$

Second,

$$\begin{aligned} y \in \ker(A^{\top}) &\Rightarrow A^{\top} y = 0 \\ &\Rightarrow \forall x \in \mathbb{R}^n : x^{\top}(A^{\top} y) = (Ax)^{\top} y = 0 \\ &\Rightarrow y \in \mathrm{Im}(A)^{\perp}. \end{aligned}$$

The second equation follows from applying the first equation to $C = A^T$ and $(U^{\perp})^{\perp} = U$. □

In terms of the transpose matrix we find two more characterizations of the image and kernel:

Lemma 1.46 *Let $A \in \mathbb{R}^{m\times n}$. Then*

$$i) \quad \ker(A) = \ker(A^\top A) \quad (\textit{and} \quad \ker(A^\top) = \ker(AA^\top)),$$
$$ii) \quad \operatorname{Im}(A) = \operatorname{Im}(AA^\top) \quad (\textit{and} \quad \operatorname{Im}(A^\top) = \operatorname{Im}(A^\top A)).$$

Proof. We only prove i) here. We show this by mutual subset relation:

- "$\ker(A) \subseteq \ker(A^TA)$":

  Let $x \in \ker(A) \overset{\text{Def. } \ker(A)}{\Rightarrow} Ax = 0 \Rightarrow A^TAx = 0 \overset{\text{Def. } \ker(A^TA)}{\Rightarrow} x \in \ker(A^TA)$.

- "$\ker(A^TA) \subseteq \ker(A)$":

  Let $x \in \ker(A^TA) \overset{\text{Def.}}{\Rightarrow} A^TAx = 0 \Rightarrow \underbrace{x^TA^TAx}_{=\|Ax\|_2^2} = 0 \overset{\text{norm } \|\cdot\|_2 \text{ is definite}}{\Rightarrow} Ax = 0 \overset{\text{Def.}}{\Rightarrow} x \in \ker(A)$.

Exercise: To prove ii) one can exploit the orthogonality of the subspaces as derived above.
The results for the transpose follow by applying the results to $C := A^T$. □

Remark
The so-called Gram matrix $A^\top A$ plays a crucial role in many applications and also analysis, for instance

- it plays a key role to derive the singular value decomposition
- it is the system matrix in the normal equation $A^\top Ax = A^\top x$ for solving least squares problems
- in graph theory it appears as graph Laplacian
- if $A \approx \nabla$ (gradient), then $A^\top \approx \operatorname{div}$ (divergence) and $A^\top A \approx \Delta$ (Laplacian)

A generalization of this result is given by the following lemma.

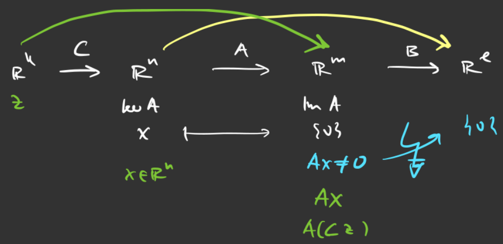
Lemma 1.47 *Let $A \in \mathbb{R}^{m \times n}$. Then*

i) *For a matrix $B \in \mathbb{R}^{\ell \times m}$ with $\ker(B) = \{0\}$ ("injective") we have*

$$\ker(BA) = \ker(A).$$

ii) *For a matrix $C \in \mathbb{R}^{n \times k}$ with $\mathrm{Im}(C) = \mathbb{R}^n$ ("surjective") we have*

$$\mathrm{Im}(AC) = \mathrm{Im}(A).$$



Proof. We show mutual subset relation:

i) "$\ker(BA) \subseteq \ker(A)$":

Let $BAx = B(Ax) = 0$. Since by assumption $\ker(B) = \{0\}$, we have $Ax = 0$.

"$\ker(A) \subseteq \ker(BA)$":

For $Ax = 0$ we also have $BAx = B(Ax) = B0 = 0$.

ii) "$\mathrm{Im}(AC) \subseteq \mathrm{Im}(A)$":

Let $y = ACx = A(Cx)$ for some $x \in \mathbb{R}^k$. Then with $z := Cx \in \mathbb{R}^n$ we see that $y = Az$ is in the image of A.

"$\mathrm{Im}(A) \subseteq \mathrm{Im}(AC)$":

Let $y = Az$ for some $z \in \mathbb{R}^n$. Since $\mathrm{Im}(C) = \mathbb{R}^n$ we find some coefficients $x \in \mathbb{R}^k$ so that $z = Cx$. With $y = ACx$ we see that y is in the image of AC. $\square$

Example

The typical context to apply Lemma 1.47 occurs when we have a decomposition of a matrix A and want to investigate its kernel and its image.

For example, consider the reduced QR decomposition $A = QR$, where $Q \in \mathbb{R}^{m\times n}$ contains orthonormal columns and $R \in \mathbb{R}^{n\times n}$ is upper triangular. Suppose that A has full rank, i.e., $\mathrm{rank}(A) = n$, so that R is invertible (in particular $\mathrm{rank}(R) = n$ and $\ker(R) = \{0\}$). We thus find by Lemma 1.47 i) that

$$\ker(A) = \ker(QR) = \ker(R) = \{0\}$$

and by Lemma 1.47 ii) that

$$\mathrm{Im}(A) = \mathrm{Im}(QR) = \mathrm{Im}(Q).$$

With other words, the n columns in Q are an orthonormal basis for the image $\mathrm{Im}(A)$ of A.