1 Fundamentals of Linear Algebra

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 $i = 1, \dots, 3$?

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** $\mathbb{F}$.

Operations

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

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

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

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**.*

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

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

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*

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

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

(1) By rows: *Used for computations*

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

$\rightarrow$ This refers to the way of computing the matrix-vector product according to "**row** $\cdot$ **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.

(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, $A := [a_1, \dots, a_n] \in \mathbb{F}^{m \times n}$.*

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.,*

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

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.

1.2 Span and Image – Linear Independence and Kernel

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 vectors*** $a_1, \dots, a_n \in \mathbb{F}^m$ *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 matrix*** $A := [a_1, \dots, a_n] \in \mathbb{F}^{m \times n}$ *is defined by*

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

Let us properly define these concepts:

Definition 1.21 (*Linear independence and kernel*)

i) *Vectors $a_1, \dots, a_r \in \mathbb{F}^m$ 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., $\exists_1 x_1, \dots, x_r \in \mathbb{F} \colon b = x_1 a_1 + \ldots + x_r a_r$.*

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!

1.3 Subspaces of $\mathbb{F}^n$ – Basis and Dimension

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

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*.*

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*

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

1.4 Inverse Matrices

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

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.

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

Remark:

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

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

→ 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 $x, y \in \mathbb{R}^n$,

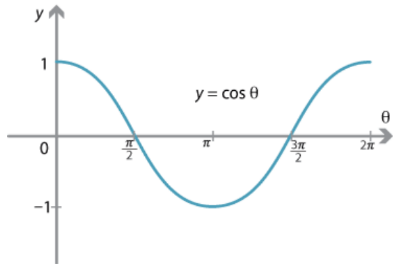
$$\left| x^T y \right| \leq \|x\|_2 \cdot \|y\|_2.$$

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

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

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:

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

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:

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

$$Q^H Q = I_n.$$

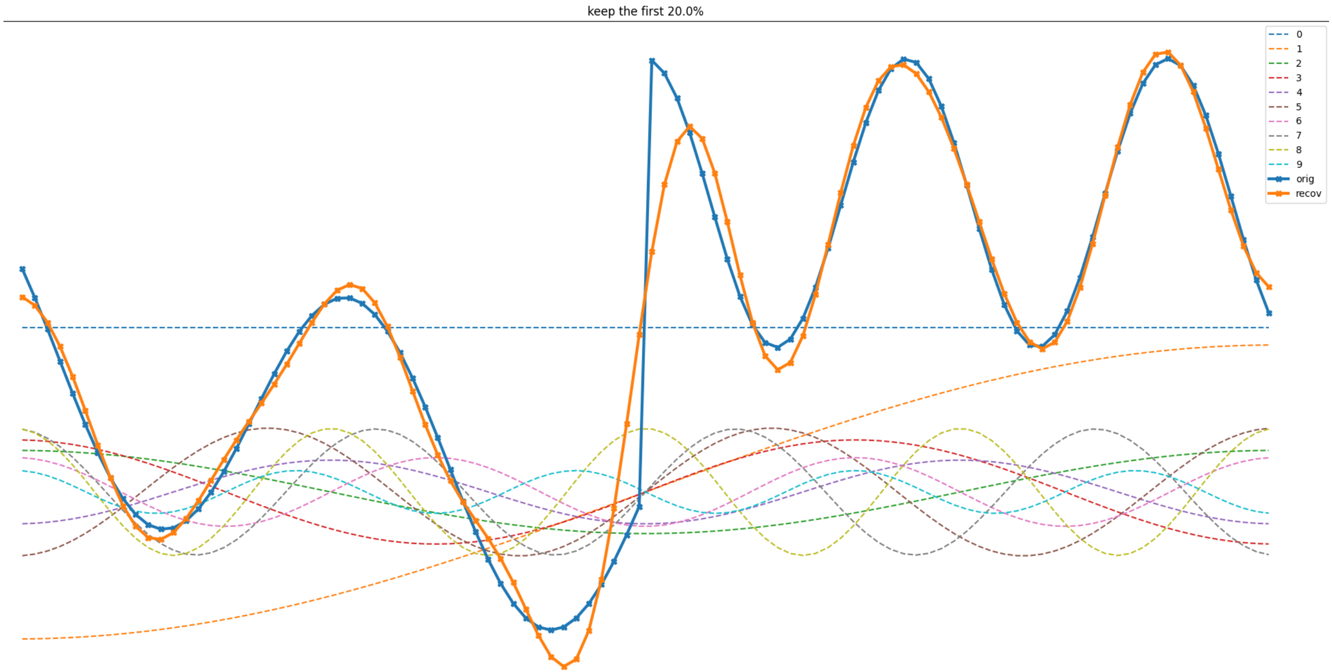
Since this is a central property, we make this a definition:

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**.*

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



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 $\mathbb{F}^{n\times n}$

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 $i = 1$)

- $n = 3$: *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) $\forall A, A' \in \mathbb{F}^{n\times n}: \quad \det(A \cdot A') = \det(A) \cdot \det(A')$

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

1.8.2 Existence and Uniqueness Analysis

Summary

Aim:

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

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

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

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:

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

Example 1.42 Let us consider

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

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

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 $\mathbb{R}^n$.*

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*

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

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 $A^\top \approx$ 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*

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

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., $\operatorname{rank}(A) = n$, so that R is invertible (in particular $\operatorname{rank}(R) = n$ 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

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

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