1 Linear Algebra

The concepts and methods of linear algebra are essential to almost everything related to imaging and visualization. We will start this chapter with the basic concepts of linear spaces and linear mappings, moving on to linear equation systems and various ways of solving and characterizing these linear systems. The main application example of this chapter will be in the area of computed tomography, but more application examples will be covered in the exercises.

1.1 Linear spaces

The most basic concept of linear algebra is that of a linear space. In this course we will restrict ourselves to linear spaces over the field of the real numbers $\mathbb{R}$ for simplicity. Linear spaces can also be defined over other fields (such as the complex numbers $\mathbb{C}$), but for imaging and visualization applications the real numbers are in most cases sufficient.

Definition (Linear space (over $\mathbb{R}$))**.** *Let V be a non-empty set together with the two operations "sum" $+ : V \times V \to V$ and "scalar multiplication" $\cdot : \mathbb{R} \times V \to V$, such that $a + b \in V$ and $\lambda a \in V$ for every $a, b \in V$ and $\lambda \in \mathbb{R}$. V is called a **linear space** if the following rules are fulfilled:*

(1) $a + b = b + a$ for all $a, b \in V$ (commutativity)

(2) $(a + b) + c = a + (b + c)$ for all $a, b, c \in V$ (associativity)

*(3) there exists a **zero element** $0 \in V$ such that $a + 0 = a$ for all $a \in V$*

*(4) for every $a \in V$ there exists an **inverse element** $-a \in V$ such that $a + (-a) = 0$*

(5) $1a = a$ for all $a \in V$ ($1 \in \mathbb{R}$ real number)

(6) $\lambda(\mu a) = (\lambda\mu)a$ for all $\lambda, \mu \in \mathbb{R}$, $a \in V$

(7) $\lambda(a + b) = \lambda a + \lambda b$ for all $\lambda \in \mathbb{R}$, $a, b \in V$ (distributivity)

(8) $(\lambda + \mu)a = \lambda a + \mu a$ for all $\lambda, \mu \in \mathbb{R}$, $a \in V$ (distributivity).

*An element $a \in V$ is called **vector**. We denote $a + (-b)$ in short as $a - b$.*

Conceptually, linear spaces describe a structure on a set V by defining the two operations, sum and scalar multiplication, which act together in a nice and expected manner.

One of the prime and very often used example of linear spaces is the typical n-dimensional space ($n \in \mathbb{N}$):

$$\mathbb{R}^n := \underbrace{\mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}}_{n \text{ times}} := \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : \; x_1, \ldots, x_n \in \mathbb{R} \right\}$$

with the operations sum

$$+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n, \qquad \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \mapsto \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

and scalar multiplication

$$\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n, \qquad \lambda \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

In particular, for $n = 2$ this corresponds to the intuitive two-dimensional space $\mathbb{R}^2$, and for $n = 3$ this corresponds to the intuitive three-dimensional space, which are often used in imaging and visualization.

A more general example, which actually encompasses the previous example as a special case, is the following: Let I be any non-empty set, that is $I \neq \emptyset$. Then define

$$\mathbb{R}^I := \big\{ f : \quad f : I \to \mathbb{R} \text{ mapping} \big\}.$$

$\mathbb{R}^I$ is a linear space together with the “point-wise” operations

$$\begin{aligned} (f+g)(j) &:= f(j) + g(j) \qquad \forall j \in I, \\ (\lambda f)(j) &:= \lambda f(j) \qquad \forall j \in I, \end{aligned}$$

where $f, g \in \mathbb{R}^I$, $\lambda \in \mathbb{R}$.

For particular choices of the set I, we get for example these special cases:

- Set $I = \{1, \ldots, n\}$ for $n \in \mathbb{N}$. Then $\mathbb{R}^I = \mathbb{R}^n$, as in our previous example.
- Set $I = \big\{(i,j) : \ 1 \leq i \leq n, \ 1 \leq j \leq m, \ i,j \in \mathbb{N}\big\}$ for $n, m \in \mathbb{N}$. Then we have $\mathbb{R}^I = \mathbb{R}^{m \times n}$, the linear space of $m \times n$ matrices.
- Set $I = [a,b] \subset \mathbb{R}$. Then $C([a,b])$, the space of real-valued continuous functions on the compact interval $[a,b]$, is a subset of $\mathbb{R}^{[a,b]}$:

  $$C([a,b]) := \big\{ f : [a,b] \to \mathbb{R} : f \text{ continuous} \big\} \subset \mathbb{R}^{[a,b]}.$$

  With the “point-wise” operations as above, $C([a,b])$ is a linear space itself as well (see exercises).

1.2 Lengths, distances and angles

Now that we have introduced vectors in linear spaces, we want to do something with these vectors. Everyday tasks like measuring lengths, distances or angles are very useful to have for vectors in linear spaces as well. In this section we will introduce the standard approaches on how to do this.

First we introduce the length of a vector:

Definition (Norm)**.** *Let V be a linear space (over $\mathbb{R}$). A mapping*

$$\|\cdot\| : V \to \mathbb{R}$$

is called ***norm*** *on V if it fulfills*

(1) $\|x\| \geq 0$ *and* $\|x\| = 0 \Leftrightarrow x = 0$

(2) $\|\lambda x\| = |\lambda| \; \|x\|$

(3) $\|x + y\| \leq \|x\| + \|y\|$

for $x, y \in V$ and $\lambda \in \mathbb{R}$. The first property is called positive definiteness, the second one homogeneity and the third one is called the triangle inequality.

The main examples of a norm are the so-called ***p*-norms** on the linear spaces $\mathbb{R}^n$ ($n \in \mathbb{N}$): Let $x = (x_i) \in \mathbb{R}^n$ and $p \geq 1$, then the p-norm of x is defined as

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} .$$

For $p = 2$ this corresponds to real-world length measurements, we also call the 2-norm the **Euclidean norm**. The length of the two-dimensional vector $a = \binom{1}{0} \in \mathbb{R}^2$, for example, computes as $\|a\|_2 = \sqrt{1^2 + 0^2} = 1$, while $b = \binom{1}{1} \in \mathbb{R}^2$ has length $\|b\|_2 = \sqrt{1^2 + 1^2} = \sqrt{2}$. Of course there are also many other norms on many other linear spaces.

Using the length, i.e. the norm of a vector, we can also introduce a notion of distance between two vectors by computing the length of the difference between the two vectors:

Definition (Distance)**.** *Let V be a linear space with a norm $\|\cdot\|$. Then for $x, y \in V$ the* ***distance*** *of x and y is defined as*

$$d(x, y) := \|y - x\|.$$

Following the previous example in $\mathbb{R}^2$ with the 2-norm, we can compute the distance between vectors $c = \binom{3}{2} \in \mathbb{R}^2$ and $d = \binom{1}{3} \in \mathbb{R}^2$ as $d(c, d) = \| \binom{1-3}{3-2} \|_2 = \sqrt{2^2 + 1^2} = \sqrt{5}$. As the 2-norm is called Euclidean norm, the corresponding distance is called **Euclidean distance**.

Distances defined like this are also a *metric*, which is a central concept of the next chapter (Analysis), but more on that later.

Finally, it is also very useful to have a notion of angle between two vectors. While angle is an intuitive concept in two dimensions ($\mathbb{R}^2$), it is a bit more involved to extend this to higher dimensions or arbitrary linear spaces. The main mathematical tool to do this is the so-called "scalar product":

Definition (Scalar product)**.** *Let V be a linear space (over $\mathbb{R}$). A mapping*

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$$

is called ***scalar product*** *(or* ***inner product*** *or* ***dot product****) on V if it fulfills*

(1) $\langle x+x', y\rangle = \langle x, y\rangle + \langle x', y\rangle$ *and* $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$

(2) $\langle x, y\rangle = \langle y, x\rangle$

(3) $\langle x, x\rangle \geq 0$ *and* $\langle x, x\rangle = 0 \Leftrightarrow x = 0$

for all $x, x', y \in V$ *and* $\lambda \in \mathbb{R}$*. The first property is called linearity, the second symmetry and the third one positive definiteness.*

The standard example of a scalar product is the *standard inner product* on $\mathbb{R}^n$, which is defined for vectors $x = (x_i), y = (y_i) \in \mathbb{R}^n$ as follows:

$$\langle x, y\rangle := \sum_{i=1}^{n} x_i y_i.$$

An equivalent short notation is $\langle x, y\rangle = x^T y$.

Numerical example: using the same vectors $c = \left(\begin{smallmatrix}3\\2\end{smallmatrix}\right)$ and $d = \left(\begin{smallmatrix}1\\3\end{smallmatrix}\right)$ in $\mathbb{R}^2$ as before: $\langle c, d\rangle = 3 \cdot 1 + 2 \cdot 3 = 9$.

Other scalar products can be defined on $\mathbb{R}^n$ (more on that later), or on other linear spaces. One such example is the linear space $C([a, b])$ from above, here

$$\langle f, g\rangle := \int_a^b f(t)g(t)\ dt$$

for $f, g \in C([a, b])$, defines a scalar product. There are some caveats with linear spaces of functions, the scalar product via an integral of the product of functions does not always work. For example, in general it does **not** work for $\mathbb{R}^{[a,b]}$ due to undefined integrals.

If a linear space V has a scalar product $\langle\cdot, \cdot\rangle$, you can automatically define a norm on V,

$$\|x\| := \sqrt{\langle x, x\rangle}$$

for $x \in V$. We call this norm the *norm induced by the scalar product.*

Recalling the p-norms on $\mathbb{R}^n$ from before, for $p = 2$ the norm is obviously induced by the standard inner product. For $p \neq 2$ the p-norm is not induced by any scalar product (as can be shown via the parallelogram law).

Going back to a general linear space V with scalar product $\langle\cdot, \cdot\rangle$: An important tool using the scalar product is the **Cauchy-Schwarz Inequality**:

$$\langle x, y\rangle^2 \leq \langle x, x\rangle\ \langle y, y\rangle \qquad \forall x, y \in V.$$

We usually do not cover proofs in this course, however the proof of this inequality is quite instructive and hence posed as an exercise (see exercise sheet 1).

We use the Cauchy-Schwarz inequality now in order to introduce the notion of angle between vectors:

Definition (Angle)**.** *Let* V *be a linear space with a scalar product* $\langle\cdot, \cdot\rangle$ *and induced norm* $\|\cdot\|$*. Let* $x, y \in V$*, where* $x, y \neq 0$*, then we have*

$$\langle x, y\rangle = \|x\|\ \|y\| \cos\varphi,$$

where φ *is the* ***angle*** *between* x *and* y*.*

In particular, $\langle x, y\rangle = 0$ implies $\varphi = \frac{\pi}{2} = 90°$, i.e. x, y are orthogonal or in short notation: $x \perp y$.

This angle is a well-defined quantity, as is proven in the following: For $x, y \in V$ with $x, y \neq 0$ we have due to the Cauchy-Schwarz inequality that $\langle x, y\rangle^2 \leq \|x\|^2\ \|y\|^2$, and thus by applying the square root $-\|x\|\ \|y\| \leq \langle x, y\rangle \leq \|x\|\ \|y\|$, which yields

$$-1 \leq \frac{\langle x, y\rangle}{\|x\|\ \|y\|} \leq 1.$$

$\cos : [0, \pi] \to \mathbb{R}$ is strongly monotonously decreasing from 1 to -1. Thus there exists exactly one ($!\exists$) $\varphi \in [0, \pi]$ with

$$\varphi = \cos^{-1}\left(\frac{\langle x, y\rangle}{\|x\|\ \|y\|}\right).$$

Numerical examples: For our previous example vectors $c, d \in \mathbb{R}^2$ with $c = \binom{3}{2}$, $d = \binom{1}{3}$, we have $\frac{\langle c,d\rangle}{\|c\|\ \|d\|} = \frac{9}{\sqrt{130}}$, and thus $\varphi = \cos^{-1}\left(\frac{9}{\sqrt{130}}\right) = 0.661\ldots \approx 37.9°$. (You are welcome to double-check this by a drawing on paper and actually measuring the angle between the two vectors.)

As this notion of angle is now defined in general for any linear space with scalar product and induced norm, we can also (for fun) apply it for example to the linear space $C([0, 1])$ with the previously introduced scalar product $\langle f, g\rangle := \int_0^1 f(t)g(t)\ dt$ for $f, g \in C([0, 1])$. The angle between the two example functions $f(t) = t$ and $g(t) = t^2$ can then be computed as $\varphi \approx 14.5°$ (whatever that actually means).

1.3 Coordinate systems and bases

Many applications involve different coordinate systems, for example in augmented reality. Before we explicitly introduce such coordinate systems, we first need to introduce subspaces as particular parts of linear spaces.

Definition (Subspace)**.** *Let V be a linear space. A nonempty set $U \subset V$ is called **subspace** of V if*

(1) $x, y \in U \quad\Longrightarrow\quad x - y \in U,$

(2) $x \in U,\ \lambda \in \mathbb{R} \quad\Longrightarrow\quad \lambda x \in U.$

U is then a linear space itself. To denote that U is a subspace of V, the short notation $U \leq V$ is used.

Checking these two requirements is sufficient. The zero element is automatically in U, as $x - x = 0 \in U$ for $x \in U$ as per (1). The same is true for the inverse element as $-1 \cdot x = -x \in U$ for $x \in U$ due to (2).

Examples:

- Let V be a linear space with zero element 0_V. Then $\{0_V\}$ is a subspace of V, in fact it is the smallest possible subspace.

- As a more concrete example we have $\mathbb{R}^2 \leq \mathbb{R}^3$, with the embedding $\mathbb{R}^2 = \left\{ \left(\begin{smallmatrix} x \\ y \\ 0 \end{smallmatrix} \right) : \; x, y \in \mathbb{R} \right\}$.
- From our previous examples we have $C([a,b]) \leq \mathbb{R}^{[a,b]}$ (see also exercises).

Subspaces have some useful properties. Let V be a linear space. Then:

- Let U_i, $i \in I$ with I some index set, be a series of subspaces of V, then their intersection $\bigcap_{i \in I} U_i$ is also a subspace of V.
- Let $M \subset V$, then there exists a smallest subspace of V, which we denote $\langle M \rangle$, that contains M. This $\langle M \rangle$ is called the **(linear) span** of M in V. We have

  $$\langle M \rangle = \bigcap_{M \subset U \leq V} U,$$

  i.e. $\langle M \rangle$ is the intersection of all subspaces U of V that contain M, and we can also show that

  $$\langle M \rangle = \left\{ \sum_{x \in M} \lambda_x x : \; \lambda_x \in \mathbb{R}, \text{ almost all } \lambda_x = 0 \right\},$$

  ("almost all $\lambda_x = 0$" means that all except finitely many λ_x are 0), i.e. $\langle M \rangle$ is the set of all finite linear combinations of elements of M.

Using the linear span we can now define generating sets:

Definition (Generating set)**.** *Let V be a linear space. A set of vectors $(b_i)_{i \in I}$ in V is called **generating set** of V if*

$$\big\langle (b_i)_{i \in I} \big\rangle = V.$$

In other words, if the linear span of a set of vectors $(b_i)_{i \in I}$ is already the entire linear space V, then the (b_i) "generate" V.

Numerical examples: Consider $\mathbb{R}^3$ and its canonical (or standard) basis, i.e. the vectors $e_1 = \left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix} \right)$, $e_2 = \left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix} \right)$, $e_3 = \left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix} \right)$. Then $\{e_1, e_2, e_3\}$ generates $\mathbb{R}^3$, as well as $\{e_1 + e_2, e_2 + e_3, e_3\}$. The set $\{e_1, e_3\}$, however, does not generate $\mathbb{R}^3$.

In order to define generalized coordinate systems, one ingredient is missing:

Definition (Linearly independent)**.** *Let V be a linear space. A finite set of vectors $a_1, \ldots, a_n \in V$ is called **linearly independent** if*

$$\lambda_1 a_1 + \ldots + \lambda_n a_n = 0 \qquad \Longrightarrow \qquad \lambda_1 = \ldots = \lambda_n = 0.$$

A set of vectors $(a_i)_{i \in I}$ in V is called linearly independent, if all finite subsets of $(a_i)_{i \in I}$ are linearly independent.

Numerical example: The canonical basis vectors of $\mathbb{R}^3$ from above, e_1, e_2, e_3, are linearly independent, as

$$\lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3 = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

implies $\lambda_1 = \lambda_2 = \lambda_3 = 0$.

Now we can define a generalized coordinate system of linear spaces, which is called "basis":

Definition (Basis)**.** *Let V be a linear space. A set of vectors $(b_i)_{i\in I}$ in V is called **basis** of V, if it is a generating set of V that is linearly independent.*

Numerical examples: Let us again consider the linear space $\mathbb{R}^3$. Then $\{e_1, e_2, e_3\}$ is a basis of $\mathbb{R}^3$, as it both generates $\mathbb{R}^3$ and is linearly independent, as shown above. The set $\{e_1, e_1 + e_2, e_1+e_2+e_3\}$ is also a basis of $\mathbb{R}^3$. As a counterexample, the set $\{e_1, e_2\}$ is not a basis of $\mathbb{R}^3$, as it does not generate $\mathbb{R}^3$, even though it is linearly independent. Another counterexample is the set $\{e_1, e_2, e_3, e_1 + e_2\}$, it is not a basis of $\mathbb{R}^3$ as it is not linearly independent, even though it generates $\mathbb{R}^3$.

There are some notable properties of bases, namely:

- Every linear space V has a basis.
- Let $b_1, \dots, b_n$ be a basis of the linear space V. Then every vector $a \in V$ can be written as a **unique** linear combination $a = \sum_{i=1}^n \lambda_i b_i$ with $\lambda_i \in \mathbb{R}$.
- Let V be a linear space. Then we have

$$\begin{aligned} b_1, \dots, b_n \in V \text{ basis} \quad &\iff \quad b_1, \dots, b_n \in V \text{ minimal generating set of } V \\ &\iff \quad b_1, \dots, b_n \in V \text{ maximal linearly independent set in } V. \end{aligned}$$

- Let V be a linear space with basis $b_1, \dots, b_n$. Then every basis has exactly n elements. This invariant is called **dimension** of V, in short we write $\dim V := n$.

  Our frequently used example $\mathbb{R}^n$ is indeed n-dimensional, i.e. $\dim \mathbb{R}^n = n$.

  For a subspace $U \leq V$ of a linear space V, we always have $\dim U \leq \dim V$ (which explains one motivation of the short-hand notation $U \leq V$).

Using bases that are well suited to our application problem is one of the major tools at our disposal. This is particularly true in image processing, where images can be represented in many fashions.

One way to represent an image is a function $f : X \to Y$, where X is the *domain*, for example 2–dimensional space $\mathbb{R}^2$ for 2–dimensional images, and Y is the *range*, which could be grey values (represented for example by $\mathbb{R}$) or RGB color values (represented for example by $\mathbb{R}^3$). Such a function f is also a vector in the linear space Y^X, and since every linear space has a basis, let's say $\{\psi_i\}$, this f can now be represented as a linear combination of these basis vectors,

$$f = \sum_i \alpha_i \psi_i,$$

where $\alpha_i \in \mathbb{R}$. The ψ_i could for example be a family of wavelets.

In a computer, a discrete representation is usually more appropriate. Instead of representing, for example, a 2-dimensional grey value image as a continuous function $f : X \to Y$ with $X = \mathbb{R}^2$ and $Y = \mathbb{R}$, we can sample X at discrete points $X' := \{1, \dots, M\} \times \{1, \dots, N\}$ and quantize Y into discrete values $Y' := \{1, \dots, L\}$. Then $f' : X' \to Y'$ can be represented on a computer, for example as a $M \times N$ matrix. In fact, even f' is a vector in the linear space $\mathbb{R}^{M\times N}$ and could be represented in different fashions depending on bases of the linear space.

1.4 Linear mappings and matrices

Now that we have a structure on a set (linear spaces), with various tools to describe elements of the set (norm, distance, scalar product) as well as coordinate systems, let us look at functions that play "nice" with the structure on the set.

Definition (Linear mapping)**.** *Let V, V' be linear spaces. A mapping $f : V \to V'$ is called* ***linear*** *(or a morphism) if*

$$\begin{aligned} f(a+b) &= f(a) + f(b) \\ f(\lambda a) &= \lambda f(a) \end{aligned}$$

for all $a, b \in V$, $\lambda \in \mathbb{R}$. A linear mapping $f : V \to V'$ is called ***isomorphism*** *if f is bijective.*

Numerical examples:

- $0 : V \to V$, $0(v) = 0_V$ for $v \in V$ (the *zero mapping*) is linear.
- $\mathrm{id} : V \to V$, $\mathrm{id}(v) = v$ for $v \in V$ (the *identity*) is linear.
- $p_i : \mathbb{R}^n \to \mathbb{R}$, $p_i(v) = v_i$ for $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$ (the *projection* onto the i–th component of v) is linear.
- $t_a : V \to V$, $t_a(v) = v + a$ for $a \in V$, $v \in V$ (the *translation* by a) is **not** linear!

There is a very important theorem that allows us to reduce all finite dimensional dealings to $\mathbb{R}^n$:

Theorem. *Every ($\mathbb{R}$–)linear space V of dimension $n \in \mathbb{N}$ is isomorphic to $\mathbb{R}^n$.*

As it is instructive, here is a sketch of the proof: Let $a_1, \ldots, a_n$ be a basis of V, such that for $x \in V$ we have $x = \sum_{i=1}^n \lambda_i a_i$ with coefficients $(\lambda_i)_{i=1}^n \subset \mathbb{R}$, then

$$\varphi : V \to \mathbb{R}^n, \qquad \varphi(x) := (\lambda_i)_{i=1}^n$$

is the desired isomorphism. (We leave out the missing steps of this proof.)

Conveniently, linear mappings between finite dimensional linear spaces (which are equivalent to $\mathbb{R}^n$ with appropriate n as above) have a very useful representation:

Definition (Matrix)**.** *A vector of $\mathbb{R}^{m \times n}$ is generally written as a* ***matrix*** *$A \in \mathbb{R}^{m \times n}$,*

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \qquad a_{ji} \in \mathbb{R}.$$

Matrix operations. We define the usual operations on matrices briefly. Let $A, B \in \mathbb{R}^{m \times n}$ with $A = (a_{ji})$, $B = (b_{ji})$ and $\lambda \in \mathbb{R}$. Then we can define the following operations:

- Addition:
$$A + B := \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

- Scalar multiplication:
$$\lambda A := \begin{pmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{pmatrix}$$

- Matrix multiplication with $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times r}$:
$$A \cdot B := \begin{pmatrix} \mathrm{row}_1(A)\mathrm{col}_1(B) & \cdots & \mathrm{row}_1(A)\mathrm{col}_r(B) \\ \vdots & & \vdots \\ \mathrm{row}_m(A)\mathrm{col}_1(B) & \cdots & \mathrm{row}_m(A)\mathrm{col}_r(B) \end{pmatrix}$$

  or we can write it as
$$A \cdot B = \begin{pmatrix} A \cdot \mathrm{col}_1(B) & \cdots & A \cdot \mathrm{col}_r(B) \end{pmatrix}.$$

These matrix operations have many useful properties. Let $A, A_1, A_2 \in \mathbb{R}^{m \times n}$, $B, B_1, B_2 \in \mathbb{R}^{n \times r}$, $C \in \mathbb{R}^{r \times s}$. Then we have:

- $(A_1 + A_2)B = A_1B + A_2B$ and $A(B_1 + B_2) = AB_1 + AB_2$
- $\alpha(AB) = (\alpha A)B = A(\alpha B)$ for $\alpha \in \mathbb{R}$
- $A(BC) = (AB)C$
- $I_mA = AI_n = A$, where $I_n = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}$ denotes the $n \times n$ *identity matrix*
- however in general $AB \neq BA$ for $A, B \in \mathbb{R}^{n \times n}$!

Another important operation is the **transpose of a matrix.** Let $A \in \mathbb{R}^{m \times n}$ with $A = (a_{ji})$. Then
$$A^T := \begin{pmatrix} \mathrm{col}_1(A) \\ \vdots \\ \mathrm{col}_n(A) \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{mn} \end{pmatrix} \in \mathbb{R}^{n \times m}$$
is called the **transpose** of matrix A. Some useful properties are:

- $(A + B)^T = A^T + B^T$ for $A, B \in \mathbb{R}^{m \times n}$
- $(\alpha A)^T = \alpha A^T$ for $\alpha \in \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$
- $(A^T)^T = A$ for $A \in \mathbb{R}^{m \times n}$
- $(AB)^T = B^TA^T$ for $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times r}$

As already noted earlier, the transpose can be used to write the scalar product for two vectors $x, y \in \mathbb{R}^n$ as

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i = x^T y = \big(x_1, \ldots, x_n\big) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Now that matrices are introduced, the link between linear mappings in finite dimensional linear spaces and matrices is the following theorem:

Theorem. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. Then by defining the matrix*

$$F = \Big(f(e_1), \ldots, f(e_n)\Big)$$

for a basis $e_1, \ldots, e_n$ of $\mathbb{R}^n$, we have

$$f(x) = Fx.$$

Conversely, to every matrix $A \in \mathbb{R}^{m \times n}$ there corresponds a linear mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ with $f(x) = Ax$.

In other words, every linear mapping between finite dimensional spaces can be represented by an appropriately sized matrix, and vice versa. The proof of this theorem is very simple, so we show it in full:

Proof: For $x = \left(\begin{smallmatrix} x_1 \\ \vdots \\ x_n \end{smallmatrix}\right) \in \mathbb{R}^n$ we have

$$\begin{aligned} f(x) = f(x_1 e_1 + \ldots + x_n e_n) &\overset{\text{linear}}{=} x_1 f(e_1) + \ldots + x_n f(e_n) \\ &\overset{\text{matr. mult.}}{=} \Big(f(e_1), \ldots, f(e_n)\Big) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = Fx. \end{aligned}$$

$\square$

Numerical example: $f : \mathbb{R}^2 \to \mathbb{R}^3$, $f\big(\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)\big) = \left(\begin{smallmatrix} x \\ x+y \\ 2y \end{smallmatrix}\right)$ is a linear mapping, as we have

$$F = \Big(f(e_1), f(e_2)\Big) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 2 \end{pmatrix},$$

using the standard bases of $\mathbb{R}^2$ and $\mathbb{R}^3$. We also have

$$f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ x+y \\ 2y \end{pmatrix}.$$