Math 308, Linear Algebra with Applications

Natalie Naehrig

Winter 2016

Contents

Chapter 1

Linear Equations

1.1 Notation

Before we start digging into the theory of linear algebra, we need to introduce and fix some notation.

1.1.1 Notation
We let
-$\mathbb{N}$ be the set of natural numbers, i.e. the set $\{1, 2, 3, \ldots\}$
Note that we do not assume that 0 is an element of $\mathbb{N}$.
-$\mathbb{Z}$ be the set of integers, i.e. the set $\{0, 1, -1, 2, -2, 3, -3 \ldots\}$
-$\mathbb{R}$ be the set of real numbers
-$\phi := \{\}$, the empty set.
-$S_1 \subseteq S_2$ means that whenever $s \in S_1$, then it is also an element of S_2.
-$S_1 \subsetneq S_2$ means that $S_1 \subseteq S_2$, and there is at least one element $s \in S_2$, which is not an element of S_1.

1.2 Systems of Linear Equations

Let's start with some applications in linear algebra, so that we understand what all the structures, that we will learn later, are good for.
Consider the system

$$\begin{array}{rcrcr} 2x_1 & + & 3x_2 & = & 12 \\ x_1 & - & x_2 & = & 1 \end{array}$$

Solving this system means finding the set of all pairs (x_1, x_2), which show a true statement when plugged in. We can find solutions in many different ways.
Let us first approach the system algebraically, using *elimination*. The advantage of elimination is that it can easily be generalized to systems of any size. Moreover, by the end of this chapter, we will be able to precisely describe the way of finding solutions algorithmically.
We start with eliminating x_1 in the second row. To this end, compare the *coefficients* of x_1 of the first and second row. To eliminate x_1 in the second line, we need to multiply the second line by -2 and add this to the first line:

$$\begin{array}{rcrcr|l} 2x_1 & + & 3x_2 & = & 12 & \\ x_1 & - & x_2 & = & 1 & \times(-2) \end{array}$$

$$\begin{array}{rcrcr|} 2x_1 & + & 3x_2 & = & 12 \\ -2x_1 & + & 2x_2 & = & -2 \end{array}$$

$$\begin{array}{rcrcr|} 2x_1 & + & 3x_2 & = & 12 \\ & & 5x_2 & = & 10 \end{array}$$

We can now solve for x_2 in the second row $(x_2 = 2)$ and plug this result into the first row:

$$\begin{array}{rcrcr|l} 2x_1 & + & 3 \times 2 & = & 12 & -6 \\ & & x_2 & = & 2 & \end{array}$$

We then get

$$\begin{array}{rrcr|l} 2x_1 & & = & 6 & :2 \\ & x_2 & = & 2 & \end{array}$$

$$\begin{array}{rrcr|} x_1 & & = & 3 \\ & x_2 & = & 2 \end{array}$$

What does this result tell us? It gives us the solution: $(3, 2)$ is the only solution to the system! Do the test! We plug in $(3,2)$ in the linear system:

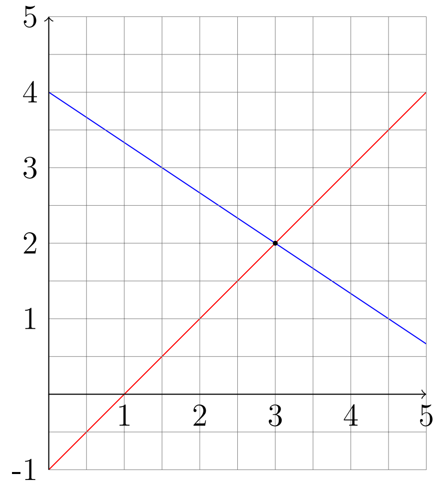
$$\left.\begin{array}{rcrcr} 2(3) & + & 3(2) & = & 12 \\ 3 & - & 2 & = & 1 \end{array}\right|$$

$$\left.\begin{array}{rcr} 12 & = & 12 \\ 1 & = & 1 \end{array}\right|$$

There is also a geometrical interpretation available: Focussing on the first row of the original system, we have

$$2x_1 + 3x_2 = 12.$$

Do you remember functions of the form $ax_1 + bx_2 = c$? These were lines in a specific form. How do we get the slope and x_2-interception from this form?
Slope: $-(\frac{a}{b})$
x_2-intercept: $\frac{c}{b}$.
In our example, we have a line with slope $-\frac{2}{3}$ and x_2-intercept 4.
The second row of the system corresponds to a line with slope 1 and y-intercept -1.
Finding a solution to the system can be interpreted as finding a point that lies on both lines, i.e. the intersection of those lines.



Let's do the same for the following system:

$$\begin{array}{rcrcr|l} 2x_1 & + & x_2 & = & 5 & \\ 4x_1 & + & 2x_2 & = & 10 & \times(-0.5) \end{array}$$

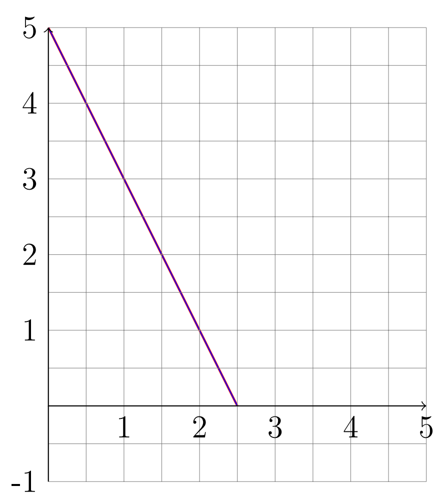
$$\begin{array}{rcrcr|} 2x_1 & + & x_2 & = & 5 \\ 0x_1 & + & 0x_2 & = & 0 \end{array}$$

How do we interpret this result? Obviously, the second row is always true. We just follow the approach we did in the first example, just a bit more general. In the first example, we found a *unique* solution for x_2. That is different here. The second row does not provide a unique solution for x_2 (remember we must never, ever divide by 0). So let us go general: Let $t \in \mathbb{R}$ be any real number, a so called *parameter*. Set $x_2 = t$ and consider the first row:

$$\begin{array}{rcrcr|l|l} 2x_1 & + & t & = & 5 & -t & \times(0.5) \\ & & x_2 & = & t & & \end{array}$$

$$\begin{array}{rcrcr|} x_1 & & & = & 2.5 - t \\ 0x_1 & + & 0x_2 & = & 0 \end{array}$$

The set of solutions is therefore $\{(2.5 - t, t) \mid t \in \mathbb{R}\}$. This has infinitely many elements. Let us have a look at the geometrical interpretation.
Both rows of the original system are forms of the same line, and the solution, i.e. the "intersection of those two lines", is this line.

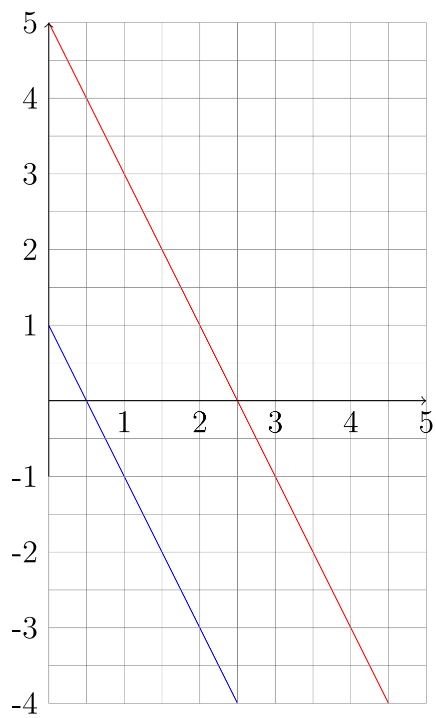


We will now discuss a third example.

$$\begin{array}{rcrcr|l} 2x_1 & + & x_2 & = & 5 & \\ 2x_1 & + & x_2 & = & 1 & \times(-1) \end{array}$$

$$\begin{array}{rcrcr|} 2x_1 & + & x_2 & = & 5 \\ 0x_1 & + & 0x_2 & = & 4 \end{array}$$

The second row reveals a contradiction, so that we conclude that the system has no solution at all. Geometrically, we see two parallel lines, i.e. two lines that do not intersect at all.



Question: Which cases do we expect when dealing with a system of three rows with three variables? What kind of geometrical objects are we dealing with?
Let us continue with formally defining linear equations. This way, we will be able, just from having the definition in its most general setting, to draw conclusions about *any* system of linear equations.

1.2.1 Definition
Let $n \in \mathbb{N}$. Then

$$a_1x_1 + a_2x_2 + \ldots + a_mx_m = b \tag{1.1}$$

with **coefficients** $a_i, b \in \mathbb{R}$ for $1 \leq i \leq m$ and x_i **variables** or **unknowns** for $1 \leq i \leq m$ is called a **linear equation**. A **solution** of this linear equation is an ordered set of m numbers

$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix}$$

(or m-tupel $(s_1, s_2, \ldots, s_m)$) such that replacing x_i by s_i, the equation (1.1) is satisfied. The **solution set** for the equation (1.1) is the set of *all* solutions to the equation.

As we have seen above, we sometimes deal with more than one equation and are interested in solutions that satisfy two or more equations simultaneously. So we need to formulize this as well.

1.2.2 Definition
Let $m, n \in \mathbb{N}$. Then **a system of linear equations/linear system** is a collection of linear equations of the form

$$\begin{array}{ccccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \ldots & + & a_{1,m}x_m & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \ldots & + & a_{2,m}x_m & = & b_2 \\ \vdots & & & & \vdots & & & & \vdots \\ a_{n,1}x_1 & + & a_{n,2}x_2 & + & \ldots & + & a_{n,m}x_m & = & b_n \end{array}$$

with $a_{i,j}, b_i \in \mathbb{R}$ for $1 \leq i \leq n$, $1 \leq j \leq m$. The **solution set of the system** is the set of m-tupels which each satisfy **every** equation of the system.

1.2.3 Remark
Note that any solution of a system with n equations and m variables is a list with m entries (independent of the number of equations n)!!

1.3 Echelon Systems

In this section, we will learn how to algorithmically solve a system of linear equations.

Eliminating Consider the following system:

$$\begin{array}{rcrcrcr|c|c} x_1 & - & 2x_2 & + & 3x_3 & = & -1 & (2) & (-7) \\ -2x_1 & + & 5x_2 & - & 10x_3 & = & 4 & & \\ 7x_1 & - & 17x_2 & + & 34x_3 & = & -16 & & \end{array}$$

$$\begin{array}{rcrcrcr|c} x_1 & - & 2x_2 & + & 3x_3 & = & -1 & \\ & & x_2 & - & 4x_3 & = & 2 & (3) \\ & & -3x_2 & + & 13x_3 & = & -9 & \end{array}$$

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & 3x_3 & = & -1 \\ & & x_2 & - & 4x_3 & = & 2 \\ & & & & x_3 & = & -3 \end{array}$$

Back Substitution Once we have reached this *triangular* form of a system, we can use **back substitution** for finding the set of solutions. To this end, start with the last row of the previous system. We can read off from that line that $x_3 = -3$. So we plug this result in the second last row:

$$x_2 - 4(-3) = 2,$$

which gives us $x_2 = -10$. The solutions for x_3, x_2 can now be plugged in the first line.

$$x_1 = -1 + 2(-10) - 3(-3) = -12.$$

In summary, we get the solution set

$$S = \{ \begin{bmatrix} -12 \\ -10 \\ -3 \end{bmatrix} \}$$

Check the solution Let's go back to the original system, plug in the solution, and verify that we did not make any mistake:

$$\begin{array}{lcccccl} -12 & - & 2(-10) & + & 3(-3) & = & -1 \\ -2(-12) & + & 5(-10) & - & 10(-3) & = & 4 \\ 7(-12) & - & 17(-10) & + & 34(-3) & = & -16, \end{array}$$

which leads to a correct set of equations.

1.3.1 Definition (Consistent/Inconsistent)
If a system (of linear equations) has at least one solution, the system is said to be **consistent**. Else it is called **inconsistent**.

The system above is therefore a consistent system.
How did we actually get to the solution? We managed to bring the system into a specific form, so that we could read of the solution. Because it is a strong mean, we put some terminology into the forms.

1.3.2 Definition
In a system of linear equations, in which a variable x occurs with a non-zero coefficient as the first term in at least one equation, x is called a **leading variable**.

1.3.3 Example
Consider the system:

$$\begin{array}{ccccccccccr} 12x_1 & - & 2x_2 & + & 5x_3 & + & & & -2x_5 & = & -1 \\ & & & & & & 3x_4 & + & 4x_5 & = & 7 \\ & & & & & & 2x_4 & - & x_5 & = & 2 \\ & & & & & & & & 7x_5 & = & -14 \end{array}$$

the variables x_1, x_4, x_5 are leading variables, while x_2, x_3 are not.

1.3.4 Definition
A system of linear equations is in **triangular form**, if it has the form:

$$\begin{array}{ccccccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & a_{1,3}x_3 & + & \dots & + & a_{1,n}x_n & = & b_1 \\ & & a_{2,2}x_2 & + & a_{2,3}x_3 & + & \dots & + & a_{2,n}x_n & = & b_2 \\ & & & & a_{3,3}x_3 & + & \dots & + & a_{3,n}x_n & = & b_3 \\ & & & & \ddots & & & & \vdots & & \vdots \\ & & & & & & & & a_{n,n}x_n & = & b_n \end{array}$$

where the coefficients $a_{i,i}$ for $1 \leq i \leq n$ are nonzero.

The following proposition reveals the meaning of triangular systems in terms of finding the solution of a system.

1.3.5 Proposition
(a) Every variable of a triangular system is the leading variable of exactly one equation.
(b) A triangular system has the same number of equations as variables.
(c) A triangular system has exactly one solution.

Proof: (a) x_i is the leading variable of row number i.
(b) Follows from (a).
(c) Do the back substitution. □

Free variables The notion of a triangular system is very strong. Each variable is the leading variable of *exactly* one equation. When solving systems, we can relax the triangular notion a bit, but still find solutions. We therefore introduce new terminology, after understanding the following example.
Consider the system:

$$\begin{array}{rcrcrcrcr} 2x_1 & - & 4x_2 & + & 2x_3 & + & x_4 & = & 11 \\ & & x_2 & - & x_3 & + & 2x_4 & = & 5 \\ & & & & & & 3x_4 & = & 9 \end{array}$$

Performing back substitution, gives us $x_4 = 3$. But, as x_3 is **not** a leading variable, we do not get a solution for it. We therefore set $x_3 := s$, a **parameter** $s \in \mathbb{R}$. We treat s to be a solution and go forth. After all we get

$$x_2 = -1 + s$$

and

$$x_1 = 2 + s.$$

To emphasize it again, s can be any real number, and for each such choice, we get a different solution for the system. What do we conclude? Yes, this system has infinitely many solutions, given by

$$S = \{ \begin{bmatrix} 2+s \\ -1+s \\ s \\ 3 \end{bmatrix} \mid s \in \mathbb{R}\} = \{ \begin{bmatrix} 2 \\ -1 \\ 0 \\ 3 \end{bmatrix} + s \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \mid s \in \mathbb{R}\}.$$

1.3.6 Definition
A linear system is in **echelon form**, if
(a) Every variable is the leading variable of *at most* one equation.
(b) The system is organized in a descending stair step pattern, so that the index of the leading variables increases from top to bottom.
(c) Every equation has a leading variable.
For each system in echelon form, a variable, that is *not* a leading variable, is called a **free** variable.

1.3.7 Proposition
A system in echelon form is consistent. If it has no free variable, it is in triangular form and has exactly one solution. Otherwise, it has infinitely many solutions.

Proof: If a system is in echelon form, determine its free variables, by determining its variables that are not leading variables. Set the free variables equal to free parameters and perform back substitution. This gives you a solution, and depending on whether there are free variables or not, you get the assertion about the number of solutions. □

1.3.8 Remark
Two comments that are important to know and understand:
- A system in echelon form is consistent and therefore does not have rows of the form

$$0x_1 + 0x_2 + \ldots + 0x_n = c$$

for some nonzero $c \in \mathbb{R}$. That means, **as free variables only have been defined for echelon systems, that we only speak of free variables in consistent systems.**
- The second part of the previous proposition reads like that: If you find free variables, then the system has infinitely many solutions!

Recapitulating what we did to find a solution, we saw that the coefficients of the linear equations changed, but the solution did not (remember, we did the check with the original system!). The reason for this is because we allowed only certain manipulations, which did not change the solution set.

1.3.9 Definition
If two linear systems lead to the same solution set, we say that these systems are **equivalent**. The following operations lead to equivalent systems:

(a) Interchange the position of two equations the system.

(b) Multiply an equation by a **nonzero** constant.

(c) Add a multiple of one equation to another.

Each of these operations are called **elementary row operations.**

1.3.10 Remark (Algorithm for solving a linear system)
- Put system into echelon form:

A Switch rows until variable with least index with non-zero coefficient is first row. This is a leading variable.

B Write down the first row AND leave it from now on.

C Eliminate terms with that variable in all but the first row.

D Repeat step [A] with system 2nd row to last row.

E Step [B] with second row.

F Step [C] with 3rd row until last.

(a) Repeat [A]-[C] likewise with remaining rows.

- If there is a row of the form $0x_1 + 0x_2 + \ldots + 0x_n = 0$, delete that row.
- If there is a row of the form $0x_1 + 0x_2 + \ldots + 0x_n = c$ for $c \neq 0$, stop and state that the system is inconsistent and hence has no solution.
- Once you have system in echelon form identify leading variables.
- Identify then possible free variables. Set those variables equal to parameters t_i.
-Do backward substitution.

1.3.11 Example
Let us apply the above algorithm.

$$\begin{array}{rcrcrcr} & & x_2 & + & 6x_3 & = & -5 \\ 2x_1 & - & x_2 & + & 5x_3 & = & -6 \\ 2x_1 & - & 2x_2 & - & x_3 & = & -1 \end{array} \rightarrow \begin{array}{rcrcrcr} 2x_1 & - & x_2 & + & 5x_3 & = & -6 \\ 2x_1 & - & 2x_2 & - & x_3 & = & -1 \\ & & x_2 & + & 6x_3 & = & -5 \end{array} \rightarrow$$

$$\begin{array}{rcrcrcr} 2x_1 & - & x_2 & + & 5x_3 & = & -6 \\ & & -x_2 & - & 6x_3 & = & 5 \\ & & x_2 & + & 6x_3 & = & -5 \end{array} \rightarrow \begin{array}{rcrcrcr} 2x_1 & - & x_2 & + & 5x_3 & = & -6 \\ & & -x_2 & - & 6x_3 & = & 5 \\ & & 0x_2 & + & 0x_3 & = & 0 \end{array} \rightarrow$$

$$\begin{array}{rcrcrcr} 2x_1 & - & x_2 & + & 5x_3 & = & -6 \\ & & -x_2 & - & 6x_3 & = & 5 \end{array}$$

The leading variables are x_1, x_2, hence the free variable is x_3. We set $x_3 = t$ for a parameter $t \in \mathbb{R}$. Now we perform backward substitution to get

$$\begin{array}{rcrcl} 2x_1 & - & x_2 & = & -6-5t \\ & & -x_2 & = & 5+6t \end{array} \qquad \begin{array}{rcl} 2x_1 & = & -6-5t-5-6t \\ x_2 & = & -5-6t \end{array}$$

$$\begin{array}{rcl} x_1 & = & -11/2-11/2t \\ x_2 & = & -5-6t \end{array}$$

In summary, the solution set is

$$S = \{ \begin{bmatrix} -11/2 \\ -11/2t \\ -5 \\ -6t \end{bmatrix} \mid t \in \mathbb{R}\} = \begin{bmatrix} 11/2 \\ 0 \\ -5 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 11/2 \\ 0 \\ -6 \end{bmatrix} \mid t \in \mathbb{R}\}$$

1.4 Matrices and Gaussian Elimination

From now on we will start to make precise definitions in mathematical style. Doing so, we can always be sure, that conclsuions we draw from assumptions are true.
One might consider the following sections abstract and wonder what it has to do with linear equations. I promise, that in the end, all that follows will an important wole in finding the set of solutions. Let us start with a structure called matrix, appearing all the time in linear algebra.

1.4.1 Definition
Let m, n be natural numbers. We call a rectangular array A with n rows and m columns of the form

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & & & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{bmatrix}$$

with $a_{i,j} \in \mathbb{R}$ for $1 \leq i \leq n$ and $1 \leq j \leq m$ an (n, m)**-matrix** or $(n \times m)$**-matrix**. We also write $(a_{i,j})$ for A. Furthermore, the prefix (n, m) is called the **size** or the **dimension** of A.

1.4.2 Example
Consider the matrix

$$A := \begin{bmatrix} 1 & 7 & 0 \\ 2 & -1 & 2 \\ -3 & 5 & 2 \\ -1 & 0 & -8 \end{bmatrix}$$

Let us get acquainted with the new terminology:
(i) What is the dimension of A? It is a $(4, 3)$-matrix.
(ii) What is $A_{4,2}$, the $(4, 2)$-entry of A? This is 0.

Having this new terminology, we focus again on linear equations. Take a close look at the algorithm we used for finding the solutions. Note that we always only changed the *coefficients* of the equations, rather than the variables (e.g. we never ended with $x_2x_3^2$, from x_2, say). So why bother and carry the variables all the time?? Just focus on the coefficients and proceed as usual. That is the idea behind the next section.

1.4.3 Definition
Let $m, n \in \mathbb{N}$. Consider the system of equations:

$$\begin{array}{ccccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \dots & + & a_{1,m}x_m & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \dots & + & a_{2,m}x_m & = & b_2 \\ \vdots & & & & \vdots & & & & \vdots \\ a_{n,1}x_1 & + & a_{n,2}x_2 & + & \dots & + & a_{n,m}x_m & = & b_n \end{array}$$

with $a_{i,j}, b_j \in \mathbb{R}$ for $1 \leq i \leq n$, $1 \leq j \leq m$.

Then

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & & & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{bmatrix}$$

is called the **coefficient matrix** for the system. Note that it is an (n, m)-matrix.
Furthermore,

$$B = \left[\begin{array}{cccc|c} a_{1,1} & a_{1,2} & \dots & a_{1,m} & b_1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} & b_2 \\ \vdots & & & \vdots & \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} & b_n \end{array}\right]$$

is the **augmented matrix** for the system. This is an $(n, m+1)$-matrix. The augmented matrix B is also denoted by $[A \mid \mathbf{b}]$, where A is the coefficient matrix and

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Let us look at a specific example.

1.4.4 Example
Consider the system

$$\begin{array}{rcrcrcr} 2x_1 & + & 3x_2 & & & = & -1 \\ 4x_1 & & & + & 7x_3 & = & 4 \\ & & 3x_2 & - & x_3 & = & 0 \end{array}$$

The corresponding augmented matrix is

$$\left[\begin{array}{rrr|r} 2 & 3 & 0 & -1 \\ 4 & 0 & 7 & 4 \\ 0 & 3 & -3 & 0 \end{array}\right]$$

Just as we did for linear systems, we can define elementary operations on matrices.

1.4.5 Definition
We define the following **elementary row operations** on matrices.

(a) Interchange two rows.

(b) Multiply a row by a **nonzero** constant.

(c) Add a multiple of a row to another.

Two matrices A, B are said to be **equivalent** if one can be obtained from the other through a series of elementary row operations. In this case, we write $A \sim B$ or $A \to B$ (but we do not write $A = B$!).

Let us apply some elementary row operations to the matrix in Example 1.4.4.

1.4.6 Example
Let the matrix A be as in example 1.4.4.
Interchanging row 2 and row 3, we get

$$\left[\begin{array}{ccc|c} 2 & 3 & 0 & -1 \\ 0 & 3 & -1 & 0 \\ 4 & 0 & 7 & 4 \end{array}\right]$$

Adding -2 times the first row to the third gives:

$$\left[\begin{array}{ccc|c} 2 & 3 & 0 & -1 \\ 0 & 3 & -1 & 0 \\ 0 & -6 & 7 & 6 \end{array}\right]$$

Multiply the second row by 2:

$$\left[\begin{array}{ccc|c} 2 & 3 & 0 & -1 \\ 0 & 6 & -2 & 0 \\ 0 & -6 & 7 & 6 \end{array}\right]$$

And finally adding the second to the third gives:

$$\left[\begin{array}{ccc|c} 2 & 3 & 0 & -1 \\ 0 & 6 & -2 & 0 \\ 0 & 0 & 5 & 6 \end{array}\right]$$

Note that all these matrices, including the original one from Example 1.4.4 are equivalent.
Let us retrieve the associated linear system to this augmented matrix. It looks like:

$$\begin{array}{rcrcrcr} 2x_1 & + & 3x_2 & & & = & -1 \\ & & 6x_2 & - & 2x_3 & = & 0 \\ & & & & 5x_3 & = & 6 \end{array}$$

This system is now in triangular form and we can perform backward substitution to determine the solution set. In particular, we have $x_3 = 6/5$, $x_2 = 2(6/5)/6 = 2/5$, and $x_1 = 0.5(-1 - 3(2/5)) = -11/10$, hence

$$S = \{ \begin{bmatrix} -11/10 \\ 2/5 \\ 6/5 \end{bmatrix} \}$$

We see that the terminology we introduced for linear systems, can be carried over to that of augmented matrices.

1.4.7 Definition
A **leading term** in a row is the leftmost non zero entry of a row. A matrix is in **echelon form**, if

(a) Every leading term is in a column to the left of the of the leading term of the row below it.

(b) All zero rows are at the bottom of the matrix.

A **pivot position** in a matrix in echelon form is one that consists of a leading term. The columns containing a pivot position are called **pivot columns**. This algorithm, which transforms a given matrix into echelon form, is called **Gaussian elimination**.

1.4.8 Remark
(a) Note that the entries in a column that are below a pivot position will only be zeros.
(b) Also note the difference between *systems* in echelon form and *matrices* in echelon form. For matrices we do allow zero rows. The mere stair pattern is a criteria for matrices in echelon form. So it might happen that going back from a matrix in echelon form (which has zero rows) leads to system that is not in echelon form, unless you delete zero rows or state it to be inconsistent.

Remember, that we can find a solution to an associated linear system from a matrix in echelon form by doing backward substitution. But actually, there is a way to transform a matrix in echelon form even further, so that the solution set can be read off from a matrix.

1.4.9 Example
Let us continue the example above. We apply a few more elementary row operations, to obtain a very specific echelon form.
In the example, we ended with

$$\left[\begin{array}{ccc|c} 2 & 3 & 0 & -1 \\ 0 & 6 & -2 & 0 \\ 0 & 0 & 5 & 6 \end{array}\right]$$

Divide each nonzero row by the reciprocal of the pivot, so that we end up with 1 as leading term in each nonzero row. In our example, we have to multiply the first row by 1/2, the second by 1/6 and the third by 1/5. This gives

$$\left[\begin{array}{ccc|c} 1 & 3/2 & 0 & -1/2 \\ 0 & 1 & -2/6 & 0 \\ 0 & 0 & 1 & 6/5 \end{array}\right]$$

Use now elementary row operations to get zeros in the entries *above* each pivot position. In our example, we add 2/6 times the last row to the second row. We get

$$\left[\begin{array}{ccc|c} 1 & 3/2 & 0 & -1/2 \\ 0 & 1 & 0 & 2/6 \\ 0 & 0 & 1 & 6/5 \end{array}\right]$$

Now we add $-3/2$ times the second row to the first row and get

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -11/10 \\ 0 & 1 & 0 & 2/6 \\ 0 & 0 & 1 & 6/5 \end{array}\right]$$

Let us find the associated linear system.

$$\begin{array}{ccccc} x_1 & & & = & -11/10 \\ & x_2 & & = & 2/6 \\ & & x_3 & = & 6/5 \end{array}$$

By finding this particular matrix form, we are able to just read off the solution. So, the solution set is

$$S = \{ \begin{bmatrix} -11/10 \\ 2/5 \\ 6/5 \end{bmatrix} \}.$$

This is, of course, worth a definition.

1.4.10 Definition
A matrix is in **reduced echelon form** if

(a) It is in echelon form.

(b) All pivot positions have a 1.

(c) The only nonzero term in a pivot column is in the pivot position.

The transformation of a matrix to a matrix in reduced echelon form is called **Gauss-Jordan elimination**. Getting to a matrix in echelon form is called the **forward phase**, getting from echelon form to reduced echelon form is called **backward phase**.

1.4.11 Example
This example is to show how to handle systems with free parameters when transforming to reduced echelon form. We start with a matrix in reduced echelon form. (Check that!)

$$\left[\begin{array}{rrr|r} 1 & 0 & -2 & 2 \\ 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

The corresponding system is:

$$\begin{array}{ccccccc} x_1 & - & 2x_2 & & & = & 2 \\ & & x_2 & + & x_3 & = & -2 \end{array}$$

We plug in the free parameter s for x_3 and get $x_2 = -2 - s$ and $x_1 = 2 + 2s$, so that the solution set is

$$S = \{ \begin{bmatrix} 2+2s \\ -2-s \\ s \end{bmatrix}, s \in \mathbb{R} \} = \{ \begin{bmatrix} 2 \\ -2 \\ 0 \end{bmatrix} + s \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, s \in \mathbb{R} \}.$$

We end this chapter with focussing on which kinds of linear systems we encounter.

1.4.12 Definition
(a) A linear equation is **homogeneous**, if it is of the form

$$a_1x_1 + a_2x_2 + \ldots + a_mx_m = 0.$$

Likewise, a system is called **homogeneous**, if it has only homogeneous equations. Otherwise, the system is called **inhomogeneous**.
(b) The vector

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

is the so called **trivial solution** to a homogeneous system.

1.4.13 Theorem (!!!)
A system of linear equations has no solution, exactly one solution, or infinitely many solutions.

Proof: (Proof strategy: Case study) Let A be the transformed matrix, which we obtained from the augmented matrix, such that A is in echelon form, we have one of the following three outcomes:
(a) A has a row of the form $[0, 0, \ldots, 0 \mid c]$ for a nonzero constant c. Then the system must be inconsistent, and hence has no solution.
(b) A has a triangular form, and thus has no free variable. By backward substitution, we find a unique solution.
(c) A corresponds to a system which is in echelon form but not triangular, and thus has one or more free variables. This means, that the system has infinitley many solutions. □

1.4.14 Remark
(a) Note that a homogeneous system is always consistent, because there is always the **trivial solution**, which is $(0, 0, \ldots, 0)$.
(b) Consider an inhomogeneous system. Assume that the corresponding augmented matrix has been reduced to echelon form. If this matrix has a row of the form

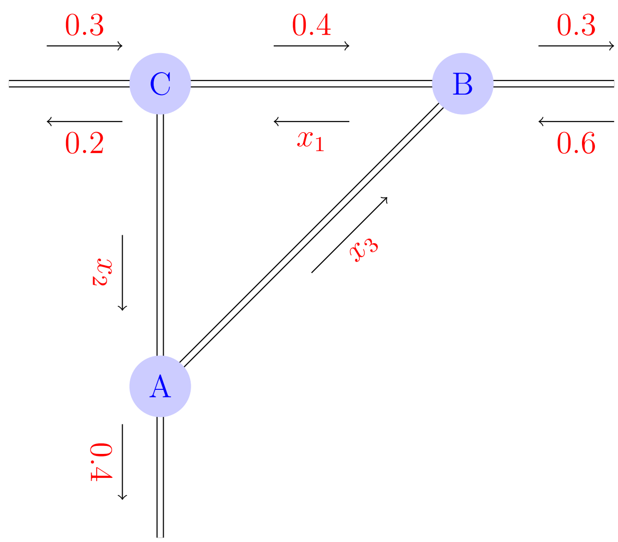
$$[0, 0, 0, \ldots, 0 \mid c]$$

for some *nonzero* constant c, then the system is inconsistent.

Note, that by Remark 1.4.14, we also find at least one solution, which is the trivial one. By Theorem 1.4.13, we therefore have either one or infinitley many solutions for homogenous systems.

Applications We will now discuss different applications, where linear systems are used to tackle the problems.

1.4.15 Example
Consider the following intersections of streets:



Let us sum up node by node, which traffic volumes occur. The incoming traffic needs to be equal to the outgoing, so that we get.
At node A we get $x_2 = 0.4 + x_3$.
At node B we get $x_3 + 0.4 + 0.6 = 0.3 + x_1$.
At node C we get $0.3 + x_1 = x_2 + 0.2 + 0.4$.
We now need to rearrange those linear equations and build up a linear system.

$$\begin{array}{ccccccc} & & x_2 & - & x_3 & = & 0.4 \\ x_1 & & & - & x_3 & = & 0.7 \\ x_1 & - & x_2 & & & = & 0.3 \end{array}$$

The associated augmented matrix looks like:

$$\left[\begin{array}{rrr|r} 0 & 1 & -1 & 0.4 \\ 1 & 0 & -1 & 0.7 \\ 1 & -1 & 0 & 0.3 \end{array}\right]$$

Let us now perform Gauss-Jordan elimination to obtain the reduced echelon form.

$$\left[\begin{array}{rrr|r} 0 & 1 & -1 & 0.4 \\ 1 & 0 & -1 & 0.7 \\ 1 & -1 & 0 & 0.3 \end{array}\right] \sim \left[\begin{array}{rrr|r} 1 & -1 & 0 & 0.3 \\ 0 & 1 & -1 & 0.4 \\ 1 & 0 & -1 & 0.7 \end{array}\right] \sim$$

$$\left[\begin{array}{rrr|r} 1 & -1 & 0 & 0.3 \\ 0 & 1 & -1 & 0.4 \\ 0 & 1 & -1 & 0.4 \end{array}\right] \sim \left[\begin{array}{rrr|r} 1 & -1 & 0 & 0.3 \\ 0 & 1 & -1 & 0.4 \\ 0 & 0 & 0 & 0 \end{array}\right] \sim$$

$$\left[\begin{array}{rrr|r} 1 & 0 & -1 & 0.7 \\ 0 & 1 & -1 & 0.4 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

The corresponding linear system is:

$$\begin{array}{rcrcrcl} x_1 & & & - & x_3 & = & 0.7 \\ & & x_2 & - & x_3 & = & 0.4 \end{array}$$

We set the free variable x_3 equal to the parameter s and thus get after backward substitution:
$x_3 = s$, $x_2 = 0.4 + s$, $x_1 = 0.7 + s$, so the solution set is

$$S = \{ \begin{bmatrix} 0.7 \\ 0.4 \\ 0 \end{bmatrix} + s \in \mathbb{R}^{\geq 0} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \}.$$

Note that s must be greater or equal tpo zero in this situation because it does not make sense in a real world application to have a negative traffic volume. The minimum travel rate from C to A is therefore 0.4.

1.4.16 Example
When propane burns in oxygen, it produces carbon dioxide and water:

$$C_3H_8 + O_2 \to CO_2 + H_2O.$$

We see, that on the left hand side, there is are eight H-atom, whereas on the right, there are two. We need to *balance* the equation, i.e. we are looking for coefficients x_1, x_2, x_3, x_4 in

$$x_1\mathrm{C_3H_8} + x_2\mathrm{O_2} \to x_3\mathrm{CO_2} + x_4\mathrm{H_2O}.$$

Let us go through each kind of atom:
There are $3x_1$ carbon atoms on the left hand side, and x_3 on the left hand side, so that we get

$$3x_1 = x_3.$$

We have $8x_1$ hydrogen atoms on the left, and $2x_4$, hence

$$8x_1 = 2x_4.$$

Further, there are $2x_2$ oxygen atomes on the left hand side, and $2x_3 + x_4$ on the right hand side, thus

$$2x_2 = 2x_3 + x_4.$$

We thus get the following homogeneous system

$$\begin{array}{rcrcrcrcl} 3x_1 & & & - & x_3 & & & = & 0 \\ 8x_1 & & & & & - & 2x_4 & = & 0 \\ & & 2x_2 & - & 2x_3 & - & x_4 & = & 0 \end{array}$$

The associated augmented matrix is

$$\left[\begin{array}{rrrr|r} 3 & 0 & -1 & 0 & 0 \\ 8 & 0 & 0 & -2 & 0 \\ 0 & 2 & -2 & -1 & 0 \end{array}\right] \sim \left[\begin{array}{rrrr|r} 3 & 0 & -1 & 0 & 0 \\ 0 & 0 & 8 & -6 & 0 \\ 0 & 2 & -2 & -1 & 0 \end{array}\right]$$

$$\left[\begin{array}{rrrr|r} 3 & 0 & -1 & 0 & 0 \\ 0 & 2 & -2 & -1 & 0 \\ 0 & 0 & 8 & -6 & 0 \end{array}\right] \sim \left[\begin{array}{rrrr|r} 1 & 0 & -1/3 & 0 & 0 \\ 0 & 1 & -1 & -1/2 & 0 \\ 0 & 0 & 1 & -3/4 & 0 \end{array}\right]$$

$$\left[\begin{array}{ccccc} 1 & 0 & 0 & -1/4 & 0 \\ 0 & 1 & 0 & -5/4 & 0 \\ 0 & 0 & 1 & -3/4 & 0 \end{array}\right]$$

The corresponding system is

$$\begin{array}{rcrcrcrcl} x_1 & & & & & - & 1/4x_4 & = & 0 \\ & & x_2 & & & - & 5/4x_4 & = & 0 \\ & & & & x_3 & - & 3/4x_4 & = & 0 \end{array}$$

We set $x_4 = s$ and get $x_3 = 3/4s$, $x_2 = 5/4s$ and $x_1 = 1/4s$. As we are not looking for the whole solution set, but rather for coefficients, so that the balance is right, we choose $s = 4$ (we want natural numbers as coefficients), so that

$$\mathrm{C_3H_8} + 5\mathrm{O_2} \rightarrow 3\mathrm{CO_2} + 4\mathrm{H_2O}.$$

1.4.17 Example
We are looking for the coefficients of a polynomial of degree 2, on which the points $(0, -1), (1, -1)$, and $(-1, 3)$ lie, i.e. we need to find a, b, c in

$$f(x) = ax^2 + bx + c.$$

Plugging in the x-values of the points, we get

$$\begin{array}{ccccccc} & & & & c & = & -1 \\ a & + & b & + & c & = & -1 \\ a & - & b & + & c & = & 3 \end{array}$$

This gives

$$\left[\begin{array}{ccc|c} 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 3 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1 & 1 & -1 \\ 0 & -2 & 0 & 4 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & -1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -1 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -1 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

The solution to this system is $a = 2, b = -2, c = -1$. This gives us

$$f(x) = 2x^2 - 2x - 1.$$

Chapter 2

Euclidean Space

2.1 Vectors

Have a look at the ingredients of milk with 2% milkfat. It says, that there are $5g$ fat, $13g$ carbs, and $8g$ protein in a cup (240ml). How could we display this information? We arrange the values in a $(3,1)$-matrix in the following way.

$$\begin{bmatrix} 5 \\ 13 \\ 8 \end{bmatrix},$$

where the first row represents fat, the secod carbs, and the third protein. This representation is called a *vector*. If you are interested in the amount of fat, carbs, and protein in 1200ml, say, then all you do is multiply each entry by 5. How can we display that? Remember, that the less unnecessary information you use, the more you keep the clear view. We therefore define

$$5\begin{bmatrix} 5 \\ 13 \\ 8 \end{bmatrix} := \begin{bmatrix} 5\cdot 5 \\ 5\cdot 13 \\ 5\cdot 8 \end{bmatrix} = \begin{bmatrix} 25 \\ 65 \\ 40 \end{bmatrix}$$

Drinking that amount of milk means taking $25g$ fat, $65g$ carbs and $40g$ protein in. This is a new kind of product, namely the product of a *scalar* with a *vector*.

The ingredient vector for whole milk is

$$\begin{bmatrix} 8 \\ 12 \\ 8 \end{bmatrix}.$$

If you drink one cup of 2% milk and one cup of whole milk, you will have ingested $5g + 8g$ fat, $13g + 12g$ carbs and $8g + 8g$ protein. How can we represent that? Just add the vectors:

$$\begin{bmatrix} 5 \\ 12 \\ 8 \end{bmatrix} + \begin{bmatrix} 8 \\ 12 \\ 8 \end{bmatrix} = \begin{bmatrix} 5+8 \\ 12+12 \\ 8+8 \end{bmatrix}.$$

The idea behind this concept will lead to a whole new terminology, even a new and mathematically very rich structure.

2.1.1 Definition
A **vector over $\mathbb{R}$ with n components** is an ordered list of real numbers $u_1, u_2, \ldots, u_n$, displayed as

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

or as

$$\mathbf{u} = [u_1, u_2, \ldots, u_n].$$

The set of all vectors with n entries is denoted by $\mathbb{R}^n$. Each entry u_i is called a **component** of the vector. A vector displayed in vertical form is called a **column vector**, a vector displayed in row form is called a **row vector**.

Remember what we did with vectors in our examples? We wrap a definition around that, because the concepts are very important and easy to generalize to so called vector spaces, which we will encounter later.

2.1.2 Definition
Let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{R}^n$ given by

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Suppose that c is a real number. In the context of vectors, c is called a **scalar**. We then have the following operations:

Equality: $\mathbf{u} = \mathbf{v}$ if and only if $u_1 = v_1, u_2 = v_2, \ldots, u_n = v_n$.

Addition:
$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix}$$

Scalar Multiplication:
$$c\mathbf{u} = c \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} cu_1 \\ cu_2 \\ \vdots \\ cu_n \end{bmatrix}$$

The **zero vector** $\mathbf{0} \in \mathbb{R}^n$ is the vector, whose entries are all 0. Moreover, we define $-\mathbf{u} := (-1)\mathbf{u}$. The set of all vectors of $\mathbb{R}^n$, *together* with addition and scalar multiplication, is called a **Euclidean Space**.

2.1.3 Example
Let

$$\mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 0 \\ 5 \end{bmatrix}$$

and

$$\mathbf{v} = \begin{bmatrix} 1 \\ 9 \\ 2 \\ 1 \end{bmatrix}.$$

Compute $2\mathbf{u} - 3\mathbf{v}$.

$$2\mathbf{u} - 3\mathbf{v} = 2 \begin{bmatrix} 2 \\ -1 \\ 0 \\ 5 \end{bmatrix} - 3 \begin{bmatrix} 1 \\ 9 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 0 \\ 10 \end{bmatrix} - \begin{bmatrix} 3 \\ 27 \\ 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ -29 \\ -6 \\ 7 \end{bmatrix}.$$

Let us now analyze the structure we have in a very abstract and mathematical way. What we do is *generalize* properties of one structure to another. Think about what we do, when we add vectors. In words, it is just *adding componentwise*. But adding componentswise is adding two numbers, just the way we are used to. Each property we are used to with ordinary addition, just carry over to vectors! Let's get more concrete:

2.1.4 Theorem (Algebraic Properties of Vectors)
Let a, b be scalars (i.e. real numbers) and $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$. Then the following holds:

(a) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

(b) $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.

(c) $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$.

(d) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.

(e) $a(b\mathbf{u}) = (ab)\mathbf{u}$.

(f) $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.

(g) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u}$.

(h) $1\mathbf{u} = \mathbf{u}$.

Proof: Let

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

Then

$$\mathbf{u}+\mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix} = \begin{bmatrix} v_1 + u_1 \\ v_2 + u_2 \\ \vdots \\ v_n + u_n \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \mathbf{v}+\mathbf{u}.$$

This proves (a).
For (b) consider

$$a(\mathbf{u}+\mathbf{v}) = a\begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_2 \end{bmatrix} = \begin{bmatrix} a(u_1 + v_1) \\ a(u_2 + v_2) \\ \vdots \\ a(u_n + v_n) \end{bmatrix} = \begin{bmatrix} au_1 + av_1 \\ au_2 + av_2 \\ \vdots \\ au_n + av_2 \end{bmatrix} = a\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + a\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = a\mathbf{u}+a\mathbf{v}.$$

Try the remaining ones as extra practice! □

Adding multiples of different vectors will be an important concept, so that it gets is own name.

2.1.5 Definition

Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ be vectors, and $c_1, c_2, \ldots, c_m$ be scalars. We then call

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_m\mathbf{u}_m$$

a **linear combination** of (the given) vectors. Note, that we do *not* exclude the case, that some (or all) scalars are zero.

2.1.6 Example

(a) What can we obtain when we linear combine 'one' vector? Let's plug in the definition. Let $\mathbf{u}$ be a vector in $\mathbb{R}^n$. Then

$$a\mathbf{u}$$

with a in $\mathbb{R}$ is by Definition 2.1.5 a linear combination of $\mathbf{u}$. What if $a = 1$? This tells us that $\mathbf{u}$ is a linear combination of $\mathbf{u}$! For $a = 0$, we see, that $\mathbf{0}$ is a linear combination of $\mathbf{u}$!

(b) Let

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

be three vectors in $\mathbb{R}^3$. Then

$$3\mathbf{e}_1 - 1\mathbf{e}_2 + 2\mathbf{e}_3 = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$$

is a linear combination of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$.

2.1.7 Example

Remember our milk example from the beginning of this chapter. We had 2%-fat milk whose ingredients could be displayed as

$$\begin{bmatrix} 5 \\ 13 \\ 8 \end{bmatrix}$$

and the whole milk ingredients,

$$\begin{bmatrix} 8 \\ 12 \\ 8 \end{bmatrix}.$$

How much of each kind of milk do I have to drink, to ingest $3g$ of fat, $19/3g$ of carbs and $4g$ of protein?
What are we looking for? We need to find the amount x_1 of 2%-fat milk and the amount x_2 of whole milk so that the given amounts of fat, carbs and protein are ingested. When finding the combined ingredients, we introduced adding to vectors, so we do the same here: We want the following to be true:

$$x_1 \begin{bmatrix} 5 \\ 13 \\ 8 \end{bmatrix} + x_2 \begin{bmatrix} 8 \\ 12 \\ 8 \end{bmatrix} = \begin{bmatrix} 3 \\ 19/3 \\ 4 \end{bmatrix}.$$

This must be true in each row, so let us write it down row wise or component wise.

$$\begin{array}{rcrcr} 5x_1 & + & 8x_2 & = & 3 \\ 13x_1 & + & 12x_2 & = & 19/3 \\ 8x_1 & + & 8x_2 & = & 4 \end{array}$$

This is a linear system! Let us find the corresponding augmented matrix:

$$\left[\begin{array}{rr|r} 5 & 8 & 3 \\ 13 & 12 & 19/3 \\ 8 & 8 & 4 \end{array}\right] \sim \left[\begin{array}{rr|r} 5 & 8 & 3 \\ 0 & -44 & -22/3 \\ 0 & -24 & -4 \end{array}\right]$$

$$\sim \left[\begin{array}{rr|r} 5 & 8 & 3 \\ 0 & 1 & 1/6 \\ 0 & 1 & 1/6 \end{array}\right] \sim \left[\begin{array}{rr|r} 5 & 0 & 10/6 \\ 0 & 1 & 1/6 \\ 0 & -24 & -4 \end{array}\right] \sim \left[\begin{array}{rr|r} 1 & 0 & 1/3 \\ 0 & 1 & 1/6 \\ 0 & 0 & 0 \end{array}\right]$$

So the solution is $x_1 = 1/3$ and $x_2 = 1/6$. In context, this means, that we need to drink 1/3 cup of 2% fat-milk and 1/6 cup of whole milk to ingest the desired amounts.
As solution we therefore have

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{x} = \begin{bmatrix} 1/3 \\ 1/6 \end{bmatrix}.$$

To finish this section, let us define a new operation between a matrix and a vector. This is a *new product*.

2.1.8 Definition (Matrix-Vector-Product)
Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ be vectors in $\mathbb{R}^n$. If

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_m \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}.$$

Then the matrix-vector product between A and $\mathbf{x}$ is defined to be

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_m\mathbf{a}_m.$$

2.1.9 Remark
(a) The product is *only* defined, if the number of columns of A equals the number of components of $\mathbf{x}$. In the definition above, this means that $\mathbf{x}$ must be an element from $\mathbb{R}^m$.
(b) Have a second look at the Definition 2.3.2 of the span. Compare this with the definition of the matrix-vector product of Definition 2.1.8. We notice the IMPORTANT fact, that the result of a matrix-vector product is in fact a linear combination of the columns of the matrix!!! In other words, any product $A \cdot \mathbf{b}$ for an (m, n)-matrix A and a vector $\mathbf{b} \in \mathbb{R}^n$ lies in the span of the columns of A.

2.1.10 Example
Find $A, \mathbf{x}$ and $\mathbf{b}$, so that the equation $A\mathbf{x} = \mathbf{b}$ corresponds to the system of equations

$$\begin{array}{rcrcrcrcr} 4x_1 & - & 3x_2 & + & 7x_3 & - & x_4 & = & 13 \\ -x_1 & + & 2x_2 & & & + & 6x_4 & = & -2 \\ & & x_2 & - & 3x_3 & - & 5x_4 & = & 29 \end{array}$$

First we collect coefficients:

$$\begin{bmatrix} 4 & -3 & 7 & -1 \\ -1 & 2 & 0 & 6 \\ 0 & 1 & -3 & -5 \end{bmatrix}$$

Then we collect variables:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

Finally collect the values on the right hand side of the equality sign.

$$\mathbf{b} = \begin{bmatrix} 13 \\ -2 \\ 29 \end{bmatrix}$$

-Put on table three equivalent ways of displaying a linear system.
We will soon see a proof for a theorem that is very elegant just because we are now able to use this new representation of linear systems with the matrix-vector product. Let usexplore an example:
Even though we treated homogeneous and inhomogeneous systems as being very different, we are about to discover, that they are very much related.

2.1.11 Example
Consider the following example for a homogeneous (blue column) and inhomogenous (red column) system, written in matrix form:

$$\left[\begin{array}{cccc|c|c} 2 & -6 & -1 & 8 & 0 & 7 \\ 1 & -3 & -1 & 6 & 0 & 6 \\ -1 & 3 & -1 & 2 & 0 & 4 \end{array}\right] \sim \left[\begin{array}{cccc|c|c} 2 & -6 & -1 & 8 & 0 & 7 \\ 0 & 0 & 1 & -4 & 0 & -5 \\ 0 & 0 & -3 & 12 & 0 & 15 \end{array}\right] \sim$$

$$\left[\begin{array}{cccc|c|c} 2 & -6 & -1 & 8 & 0 & 7 \\ 0 & 0 & 1 & -4 & 0 & -5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \sim \left[\begin{array}{cccc|c|c} 2 & -6 & 0 & 4 & 0 & 2 \\ 0 & 0 & 1 & -4 & 0 & -5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \sim$$

$$\left[\begin{array}{cccc|c|c} 1 & -3 & 0 & 2 & 0 & 1 \\ 0 & 0 & 1 & -4 & 0 & -5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

We set the free variables $x_4 := t_1$ and $x_2 := t_2$ and obtain as solution set for the respective systems

$$S_h = \{s_1 \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s_2 \begin{bmatrix} -2 \\ 0 \\ 4 \\ 1 \end{bmatrix} \mid t_1, t_2 \in \mathbb{R}\}$$

$$S_i = \{ \begin{bmatrix} 1 \\ 0 \\ -5 \\ 0 \end{bmatrix} + s_1 \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s_2 \begin{bmatrix} -2 \\ 0 \\ 4 \\ 1 \end{bmatrix} \mid t_1, t_2 \in \mathbb{R} \}.$$

Compare those sets, what do we note? The set 'differ' only by the constant vector

$$\begin{bmatrix} 1 \\ 0 \\ -5 \\ 0 \end{bmatrix}.$$

This phenomenon can be proven and is very helpful for computations.

Now we are ready to state and proof the relationship between homogeneous and inhomogeneous systems:

2.1.12 Theorem
Let $\mathbf{x}_p$ be one arbitrary solution to a linear system of the form

$$A\mathbf{x} = \mathbf{b}.$$

Then all solutions $\mathbf{x}_g$ to this system are of the form

$$\mathbf{x}_g = \mathbf{x}_p + \mathbf{x}_h,$$

where $\mathbf{x}_h$ is a solution to the associated homogeneous *system $A\mathbf{x} = \mathbf{0}$.*

Proof: Let $\mathbf{x}_g$ be any solution to the system. We must prove, that $\mathbf{x}_g$ can be written as proposed. To do so, we first consider the vector $\mathbf{x}_g - \mathbf{x}_p$. We have by Lemma the distributeive property (here without proof)

$$A(\mathbf{x}_g - \mathbf{x}_p) = A\mathbf{x}_g - A\mathbf{x}_p = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

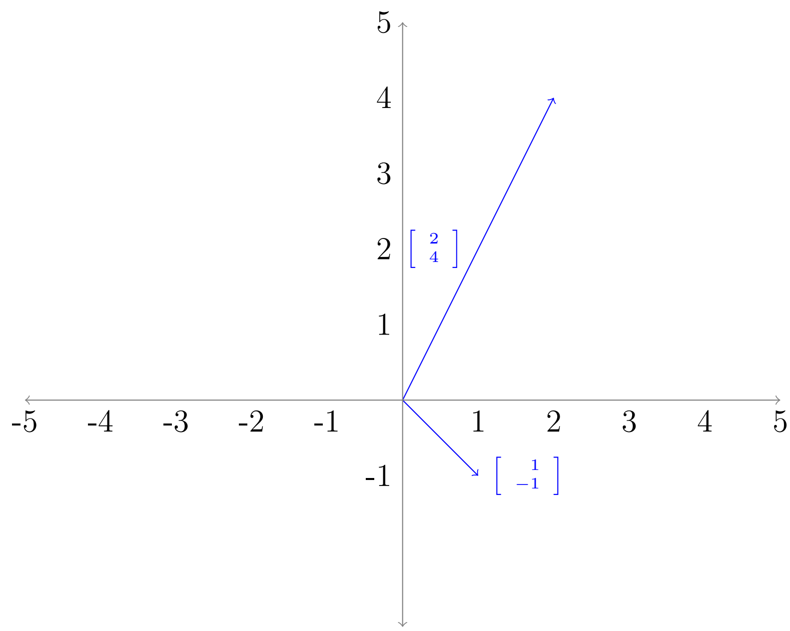
From this, we conclude, that $\mathbf{x}_g - \mathbf{x}_p$ is a solution $\mathbf{x}_h$ to the associated homogeneous system, so that we can write

$$\mathbf{x}_g - \mathbf{x}_p = \mathbf{x}_h.$$

Solving the last equation for $\mathbf{x}_g$, gives us $\mathbf{x}_g = \mathbf{x}_p + \mathbf{x}_h$. □

2.2 Geometry of $\mathbb{R}^n$

We are dealing with vectors over $\mathbb{R}^n$. This set is equipped with a huge package of additional properties. One of these properties - at least for $n = 1, 2, 3$ - is that it allows a geometric interpretation of vectors, addition of vectors, and scalar multiplication. Let us investigate $\mathbb{R}^2$.
A vector $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ of $\mathbb{R}^2$ can be understood as an arrow from the origin to the point (x_1, x_2).



It is now clear, that a length can be attached to a vector in $\mathbb{R}^n$. Good old Pythagoras shows that if $\mathbf{u} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$, its length $|\mathbf{u}|$ in the plane is given by $|\mathbf{u}| = \sqrt{2^2 + 4^2} = \sqrt{20}$.

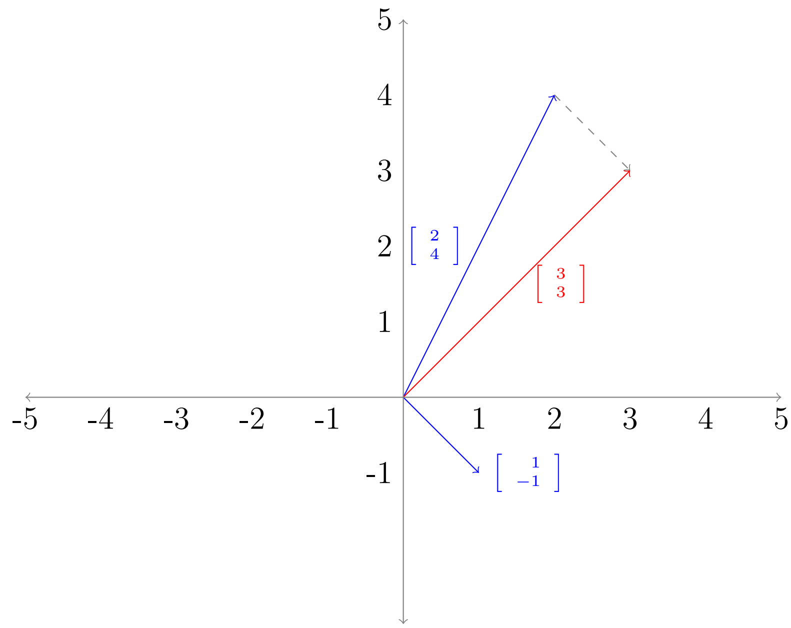
2.2.1 Definition (Length of a vector)
Let

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

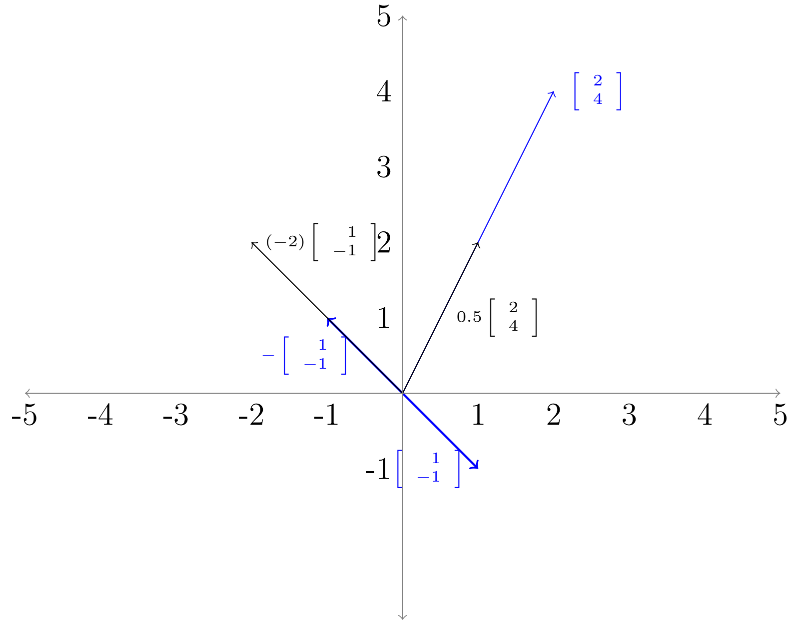
be a vecotr in $\mathbb{R}^n$. Then the **Length $|\mathbf{u}|$ of $\mathbf{u}$** is defined to be

$$|\mathbf{u}| = \sqrt{u_1^2 + u_2^2 + \ldots + u_n^2}.$$

How is adding vectors being reflected in the picture? We move one of the two vectors without changing direction or length, such that the tip of the other vector touches the tail of the moved vector. The result, i.e. the sum of the two vectors starts in the origin and ends at the tip of the moved vector.



Let us finish this section with visualizing, what scalar mulitplication means. Let $\mathbf{u}$ be a vector in $\mathbb{R}^n$ and let $c \in \mathbb{R}$. Then we have
(a) If $c = 0$, then $c\mathbf{u} = 0$, just a point in the origin.
(b) If $c > 0$, then $c\mathbf{u}$ points into the *same* direction as $\mathbf{u}$, and its length is the length of $\mathbf{u}$ multiplied by c.
(c) If $c < 0$, then $c\mathbf{u}$ points into the *opposite* direction as $\mathbf{u}$, and its length is the length of $\mathbf{u}$ multiplied by $|c|$.



2.3 Span

2.3.1 Example
Consider the following two vectors

$$\mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \text{and} \quad \mathbf{w} = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$$

We aim at answering the following question. Is $\mathbf{w}$ a linear combination of $\mathbf{u}, \mathbf{v}$, i.e. is there a solution to

$$x_1\mathbf{u} + x_2\mathbf{v} = \mathbf{w}?$$

Let's do the computation.
We like to find a solution to

$$\begin{bmatrix} 2x_1 \\ -x_1 \\ x_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ -x_2 \\ 0x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$$

The corresponding linear system is

$$\begin{array}{rcrcr} 2x_1 & + & x_2 & = & 3 \\ -x_1 & - & x_2 & = & 4 \\ x_1 & & & = & 0 \end{array}$$

The associated augmented matrix is

$$\left[\begin{array}{rr|r} 2 & 1 & 3 \\ -1 & -1 & 4 \\ 1 & 0 & 0 \end{array}\right] \sim \left[\begin{array}{rr|r} 1 & 0 & 0 \\ 2 & 1 & 3 \\ -1 & -1 & 4 \end{array}\right]$$

$$\left[\begin{array}{rr|r} 1 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & -1 & 4 \end{array}\right] \sim \left[\begin{array}{rr|r} 1 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 7 \end{array}\right]$$

The last row of the matrix reveals a contradiction. So the answer must be: No matter which combination of $\mathbf{u}$ and $\mathbf{w}$ you consider, you will never reach the desired one.

The previous example immediately raises the following question: How can we describe the set of all linear combinations of a set of given vectors? How can one determine, if the set of all linear combinations of a given set of vectors in $\mathbb{R}^n$ is the whole of $\mathbb{R}^n$? We will investigate this kind of questions in this section.

2.3.2 Definition (Span)
Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a set of vectors in $\mathbb{R}^n$. The **span** of this set is denoted by $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$, and is defined to be the set of all linear combinations

$$s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \ldots + s_m\mathbf{u}_m,$$

where $s_1, s_2, \ldots, s_m$ can be any real numbers. If $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_m\} = \mathbb{R}^n$, we say that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_m\}$ *spans* $\mathbb{R}^n$.

2.3.3 Example
(a) Let $\mathbf{u} = \mathbf{0} \in \mathbb{R}^n$. Then

$$\text{span}\{\mathbf{u}\} = \{a\mathbf{0} \mid a \in \mathbb{R}\} = \{\mathbf{0}\}$$

is a set with only one element.

(b) Let $\mathbf{u} \neq \mathbf{0} \in \mathbb{R}^n$. Then

$$\text{span}\{\mathbf{u}\} = \{a\mathbf{u} \mid a \in \mathbb{R}\}$$

is a set with infinitely many elements. Have a look back at the geometric interpretation on $a \cdot \mathbf{u}$. This shows that span$\{\mathbf{u}\}$ is a line through the origin.

Just by following Example 2.3.1, we can generally pose the following theorem:

2.3.4 Theorem
Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ and $\mathbf{w}$ be vectors in $\mathbb{R}^n$.
Then $\mathbf{w}$ is an element of span$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$, *if and only if the linear system represented by the augmented matrix*

$$\left[\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_m & \mathbf{w} \end{array}\right]$$

has a solution.

Proof: Let us plug in the definition of $\mathbf{v}$ being an element of the set span$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$:
$\mathbf{w} \in \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ if and only if there are scalars $s_1, s_2, \ldots, s_m$, such that

$$s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \ldots + s_m\mathbf{u}_m = \mathbf{w}.$$

But this is true, if and only if $(s_1, s_2, \ldots, s_m)$ is a solution to the system represented by

$$\left[\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_m & \mathbf{w} \end{array}\right]$$

This is what we wanted to show. □

2.3.5 Example
Is

$$\begin{bmatrix} 0 \\ -2 \\ 8 \end{bmatrix}$$

an element of

$$\text{span}\{\begin{bmatrix} -1 \\ -1 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ -5 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}\}?$$

Consider the following augmented matrix representing a linear system:

$$\left[\begin{array}{rrr|r} 1 & 1 & -1 & 0 \\ 2 & 3 & -1 & -2 \\ -1 & -5 & -3 & 8 \end{array}\right] \to \left[\begin{array}{rrr|r} 1 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 \\ 0 & -4 & -4 & 8 \end{array}\right]$$

$$\left[\begin{array}{rrr|r} 1 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right] \to \left[\begin{array}{rrr|r} 1 & 0 & -2 & 2 \\ 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Let us translate that back to a linear system.

$$\begin{array}{rcrcr} x_1 & = & 2 & + & 2s \\ x_2 & = & -2 & - & s \\ x_3 & = & & & s \end{array}$$

In vector form, this solution set can be written as

$$S = \{\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 0 \end{bmatrix} + s \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} | s \in \mathbb{R}\}.$$

It seems that sometimes, one could take out an element from a set $S := \{\mathbf{u}_1, \mathbf{u}_2, \dots \mathbf{u}_m\}$ of vectors in $\mathbb{R}^n$ and still obtain the same span. Which properties do such elements have? Here is the answer:

2.3.6 Theorem

Let $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$ *and* $\mathbf{u}$ *be vectors in* $\mathbb{R}^n$. *If* $\mathbf{u}$ *is in* $\operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$, *then*

$$\operatorname{span}\{\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\} = \operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}.$$

One could also say, that $\mathbf{u}$ *is redundant in* $\operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$.

Proof: Whenever we want to show equality of two sets, we do that by showing that one set is contained in the other and vice versa. To that end, let $S_0 := \operatorname{span}\{\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ and $S_1 := \operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$. So let us first show $S_1 \subseteq S_0$. Let $\mathbf{v}$ be any element in S_1. Then there are scalars $a_1, a_2, \dots, a_m$, such that $\mathbf{v} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \dots + a_m\mathbf{u}_m$. This sum can also be written as $\mathbf{v} = 0\mathbf{u} + a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \dots + a_m\mathbf{u}_m$, so that $\mathbf{v}$ lies in S_0.
Let us now show that $S_0 \subseteq S_1$. Fix $\mathbf{w} \in S_0$, so there are scalars, $b_0, b_1, b_2 \dots, b_m$ such that

$$\mathbf{w} = b_0\mathbf{u} + b_1\mathbf{u}_1 + b_2\mathbf{u}_2 + \dots + b_m\mathbf{u}_m.$$

Remember, that we also assumed that $\mathbf{u}$ is an element of S_1. That means, that there are scalars $c_1, c_2, \ldots, c_m$ such that

$$\mathbf{u} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_m\mathbf{u}_m.$$

Let us plug in this representation of $\mathbf{u}$ in the representation of $\mathbf{w}$. We then get

$$\mathbf{w} = b_0(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_m\mathbf{u}_m) + b_1\mathbf{u}_1 + b_2\mathbf{u}_2 + \ldots + b_m\mathbf{u}_m.$$

Apply the distributive property and rearrange that expression, so that you get

$$\mathbf{w} = (b_0c_1 + b_1)\mathbf{u}_1 + (b_0c_2 + b_2)\mathbf{u}_2 + \ldots + (b_0c_m + b_m)\mathbf{u}_m.$$

Now have a look at the previous equation. This is a linear combination of the vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$, hence it lies in S_1!
Now we have shown $S_0 \subseteq S_1$ and $S_1 \subseteq S_0$, hence those two sets are equal. □

We are still exploring, what one can say about $\operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$, in particular exploring what a certain choice of vectors reveals about their span.

2.3.7 Theorem
Let $S := \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a set of vectors in $\mathbb{R}^n$. If $m < n$, then this set does not span $\mathbb{R}^n$. If $m \geq n$, then the set might *span $\mathbb{R}^n$ or it might not. If not, we cannot say more without additional information about the vectors.*

Proof: Let $\mathbf{a}$ be any vector in $\mathbb{R}^n$. Let us check, what we need to generate this vector with elements of S. By Theorem 2.3.4 his means, looking at the system, whose augmented matrix looks like that:

$$A = \left[\begin{array}{cccc|c} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_m & \mathbf{a} \end{array}\right].$$

What do we usually do, when handling an augmented matrix? We aim for echelon form! Imagine, we did that for A. We then end up with an equivalent matrix of the form

$$\left[\begin{array}{cccc|c} 1 & \star & \ldots & \star & a_1 \\ 0 & 1 & \star & \star & a_2 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & a_n \end{array}\right].$$

The important thing to understand is, that because $m < n$, we *will* get at least one zero row. If you choose $\mathbf{a}$ smartly, such that *after* getting to the echelon form, you have $a_m \neq 0$, you will have created an inconsistent system. In our setting, an inconsistent system means finding no solution to generating **a** with elements of S. But this means, that there are vectors in $\mathbb{R}^n$ which do not lie in the span S. If $m \geq n$, then the matrix in echelon form may represent either a matrix with zero rows or not. If there are zero rows, we always find vectors **a**, such that the associated system becomes inconsistent. If not, we always find solutions, hence the span is $\mathbb{R}^n$. $\square$

2.3.8 Example
Is

$$\text{span}\{\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 3 \\ -3 \\ 2 \end{bmatrix}\} = \mathbb{R}^3?$$

If that were true, any vector $[a_1, a_2, a_3]^t$ of $\mathbb{R}^3$ would have to be a linear combination of the given vectors. We make use of Theorem 2.3.4 and solve the following system:

$$\left[\begin{array}{rrrr|l} 1 & -2 & 0 & 3 & a_1 \\ -1 & 2 & 0 & -3 & a_2 \\ 1 & 0 & -3 & 2 & a_3 \end{array}\right] \to \left[\begin{array}{rrrr|r} 1 & -2 & 0 & -3 & a_1 \\ 0 & 0 & 0 & 0 & a_1 + a_2 \\ 0 & 2 & -3 & -1 & -a_1 + a_3 \end{array}\right]$$

$$\left[\begin{array}{rrrr|r} 1 & -2 & 0 & -3 & a_1 \\ 0 & 2 & -3 & -1 & -a_1 + a_3 \\ 0 & 0 & 0 & 0 & a_1 + a_2 \end{array}\right]$$

The last row of the matrix reveals a contradiction. Whenever you choose a vector in $\mathbb{R}^3$ such that $a_1 + a_3 \neq 0$, you will not be able to linear combine the given vectors to that one. In particular, they cannot span the whole $\mathbb{R}^3$.

2.3.9 Remark (!!)
If you follow the previous example, you see that the considering the equivalent matrix in echelon form actually presented an understanding of whether there is a solution or not. So, whenever you are actually computing the span, go over to echelon form and look for rows which have only zero entries. If there are any, then you'll instantly know, that not the whole of $\mathbb{R}^n$ can be spanned, because there will be choices for vectors which have nonzero entries at precisely those rows.

To summarize the previous results ang get a wider perspective, let us collect the following equivalences.

2.3.10 Theorem

Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ *and* $\mathbf{b}$ *be vectors in* $\mathbb{R}^n$*. Then the following statements are equivalent:*

(a) The vector $\mathbf{b}$ *is in* $\operatorname{span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$.

(b) The vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_m\mathbf{a}_m = \mathbf{b}$ *has at least one solution.*

(c) The linear system corresponding to $\left[\begin{array}{cccc|c} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_m & \mathbf{b} \end{array}\right]$ *has at least one solution.*

(d) The equation $A\mathbf{x} = \mathbf{b}$ *with* A *and* $\mathbf{x}$ *given as in Definition 2.1.8, has at least one solution.*

Proof: (a) $\iff$ (c): This is Theorem 2.3.4

(b) $\iff$ (a): This is rewriting system in augmented matrix form to linear equation form.

(c) $\iff$ (d): This is Definition 2.1.8. □

2.3.11 Problem

Here is some more for you to practice:

(a) Do $\mathbf{u}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ *and* $\mathbf{u}_2 = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$ *span* $\mathbb{R}^2$*?*

(b) Find a vector in $\mathbb{R}^3$ *that is not in*

$$\operatorname{span}\{\begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -2 \end{bmatrix}\}.$$

How do you know a priori, that such a vector must exist?

(c) Find all values of h*, such that the set of vectors* $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ *spans* $\mathbb{R}^3$*, where*

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ h \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}$$

(d) Determine, if the equation $A\mathbf{x} = \mathbf{b}$ *has a solution for any choice of* $\mathbf{b}$*, where*

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 2 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix}.$$

Do the same for

$$A = \begin{bmatrix} 1 & 3 \\ -2 & -6 \end{bmatrix}.$$

2.4 Linear Independence

The previous section and this section are central key parts in this course. Before those sections, we learned how to solve linear equation and how to interpret equivalent systems/matrices in terms of the solution set. I therefore understand those sections as providing strong tools within linear algebra. But these current sections actually investigate the essence of linear algebra and deal with fundamental concepts.
Let us assume that we are given three vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$, such that

$$\mathbf{u}_3 = 2\mathbf{u}_1 - \mathbf{u}_2.$$

We can solve these equations for the other vectors as well:

$$\mathbf{u}_1 = 1/2\mathbf{u}_2 + 1/2\mathbf{u}_3 \quad \text{or} \quad \mathbf{u}_2 = \mathbf{u}_1 - \mathbf{u}_3\ .$$

We see, that the representation of one of the vectors *depends* on the other two. Another way of seeing this, is - because the equations are equivalent to

$$2\mathbf{u}_1 - \mathbf{u}_2 - \mathbf{u}_3 = \mathbf{0},$$

- is that the system

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 = \mathbf{0}$$

has the *non trivial* solution $(2, -1, -1)$. This is exactly the idea behind linear dependence/independence:

2.4.1 Definition (!!Linear Independence)
Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a set of vectors in $\mathbb{R}^n$. If the only solution to the vector equation

$$x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \ldots + x_m\mathbf{u}_m = \mathbf{0}$$

is the trivial solution given by $x_1 = x_2 = \ldots = x_m = 0$, then the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ is **linearly independent**. If there are non trivial solutions, then the set is **linearly dependent**.

2.4.2 Remark
(a) Note that the definition is very abstract: It is based on the *absence* of *non trivial* solutions.
(b) When are two vectors $\mathbf{u}_1, \mathbf{u}_2$ linearly (in-)dependent? There are two cases to discuss:

(i) At least one of the vectors is the zero vector. Let $\mathbf{u}_1 = \mathbf{0}$. Then $1 \cdot \mathbf{0} + 0 \cdot \mathbf{u}_2 = \mathbf{0}$. Compare that with the definition of linear independence. It fits then negation of that definition! This is because $x_1 = 1, x_2 = 0$ is *not* the trivial solution. So these two vectors are linearly dependent and we can write the equation above as $\mathbf{u}_1 = 0\mathbf{u}_2$.
(ii) Neither vector is the zero vector. Let us assume that they are linearly *dependent*. Then there are scalars c_1, c_2 not equal to 0 (assume one is zero and you will see that then at least one of the vectors must be the zero vector), such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 = 0.$$

So we can solve for $\mathbf{u}_1$ or $\mathbf{u}_2$. Let us solve for $\mathbf{u}_1$, so that we get

$$\mathbf{u}_1 = c_2/c_1\mathbf{u}_2.$$

How can we interpret this last equation: Two vectors are linearly dependent, whenever one vector is a multiple (here the factor is c_2/c_1) of the other.
(c) One single vector $\mathbf{u} \neq \mathbf{0}$ is linearly independent. We see that by plugging in the definition for linear independence for one vector:

$$c\mathbf{u} = \mathbf{0}$$

has only $c = 0$ as solution, as we assume that $\mathbf{u} \neq \mathbf{0}$. In other words, this equation has only the trivial solution $c = 0$, which makes it linearly independent.
(d) A different view on linear independence is to ask whether the zero vector is a non-trivial linear combination of the given vectors. So actually, we are dealing with questions about the span. Make use of Theorem 2.3.4 to get the following theorem!

2.4.3 Lemma

Let $S := \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a set of vectors in $\mathbb{R}^n$. Consider the system for determining if this set is linearly independent. Set

$$A = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_m \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}.$$

Then S is linearly independent, if and only if the homogeneous linear system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

2.4.4 Example
When you need to determine if a set of vectors is linearly independent or not, you *usually* just do, what the definition of linear independence says: You solve the associated system: Let us do this with a concrete example: Let

$$\mathbf{u}_1 = \begin{bmatrix} -1 \\ 4 \\ -2 \\ -3 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 3 \\ -13 \\ 7 \\ 7 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2 \\ 1 \\ 9 \\ -5 \end{bmatrix}.$$

So let us solve the system

$$x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + x_3\mathbf{u}_3 = \mathbf{0}.$$

The augmented matrix corresponding to that system is

$$\left[\begin{array}{rrr|r} -1 & 3 & -2 & 0 \\ 4 & -13 & 1 & 0 \\ -2 & 7 & 9 & 0 \\ -3 & 7 & -5 & 0 \end{array}\right] \sim \left[\begin{array}{rrr|r} -1 & 3 & -2 & 0 \\ 0 & -1 & -7 & 0 \\ 0 & 1 & 13 & 0 \\ 0 & -2 & 1 & 0 \end{array}\right]$$

$$\left[\begin{array}{rrr|r} -1 & 3 & -2 & 0 \\ 0 & -1 & -7 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 15 & 0 \end{array}\right] \sim \left[\begin{array}{rrr|r} -1 & 3 & -2 & 0 \\ 0 & -1 & -7 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & -13 & 0 \end{array}\right]$$

$$\left[\begin{array}{rrr|r} -1 & 3 & -2 & 0 \\ 0 & -1 & -7 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Back substitution now shows, that the unique solution is the trivial solution $x_1 = x_2 = x_3 = 0$. Therefore, the $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ are linearly independent.

2.4.5 Theorem
Let $S := \{\mathbf{0}, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a set of vectors in $\mathbb{R}^n$. Then the set is linearly dependent.

Proof: Let us just plug in the definition for linear independence, that is, determine the solutions for

$$x_0\mathbf{0} + x_1\mathbf{u}_1 + \ldots + \mathbf{u}_m = \mathbf{0}.$$

We are looking for non trivial solutions. Observe, that $x_0\mathbf{0} = \mathbf{0}$ for *any* x_0. So why not suggesting $(1, 0, 0, \dots, 0)$ as solution. It works and it is non trivial. So, the system has a non trivial solution, hence it is linearly dependent. □

2.4.6 Example
Consider

$$\mathbf{u}_1 = \begin{bmatrix} 16 \\ 2 \\ 8 \end{bmatrix} \quad \mathbf{u}_2 = \begin{bmatrix} 22 \\ 4 \\ 4 \end{bmatrix} \quad \mathbf{u}_3 = \begin{bmatrix} 18 \\ 0 \\ 4 \end{bmatrix} \quad \mathbf{u}_4 = \begin{bmatrix} 18 \\ 2 \\ 6 \end{bmatrix}$$

Are these vectors linearly independent?
We set up the equation for testing linear independence.

$$\left[\begin{array}{cccc|c} 16 & 22 & 18 & 18 & 0 \\ 2 & 4 & 0 & 2 & 0 \\ 8 & 4 & 4 & 6 & 0 \end{array}\right] \to \left[\begin{array}{cccc|c} 8 & 11 & 9 & 9 & 0 \\ 0 & -5 & 9 & 1 & 0 \\ 0 & 0 & 4 & 1 & 0 \end{array}\right]$$

There is at least one free variable, which means for homogeneous systems that there is not only the trivial solution. We therefore conclude that the vectors are linearly dependent.

We find in this example one situation, when we actually know that a set is linearly independent *without* solving a system:

2.4.7 Theorem
Suppose that $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ is a set of vectors in $\mathbb{R}^n$. If $n < m$, then the set is linearly dependent.

Proof: Consider the system $x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \dots + x_m\mathbf{u}_m = \mathbf{0}$. As this is an homogeneous system, we will not get any inconsistency. Because of the assumption, this system has less rows than columns. Hence, there must be free variables, thus infinitely many solutions. In particular, the system does not only have the trivial solution, hence the vectors are linearly dependent. □

Let us now make a connection between 'span' and 'linear independence'.

2.4.8 Theorem
Let $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ be a set of vectors in $\mathbb{R}^n$. Then the set is linearly dependent, if and only if one of the vectors in the set is in the span of the other vectors.

Proof: Suppose, the set is linearly dependent. Then there exist scalars $c_1, c_2, \ldots, c_m$, not all zero, such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_m\mathbf{u}_m = \mathbf{0}.$$

Note that this is the equation for linear independence. Without loss of generality, let us assume that $c_1 \neq 0$. Then solve the equation for $\mathbf{u}_1$, to get

$$\mathbf{u}_1 = -\frac{c_2}{c_1}\mathbf{u}_2 - \ldots - \frac{c_m}{c_1}\mathbf{u}_m.$$

This is an expression which shows that u_1 is a linear combination of the others. Hence, $\mathbf{u}_1$ is in the span of $\{\mathbf{u}_2, \ldots, \mathbf{u}_m\}$. Therefore, the 'forward' direction of the assertion is true.
Let us now suppose that one of the vectors is in the span of the remaining vectors. Again, without loss of generality we assume that $\mathbf{u}_1$ is in the span of $\{\mathbf{u}_2, \ldots, \mathbf{u}_m\}$. Hence, there are scalars $b_2, b_3, \ldots, b_m$, such that

$$\mathbf{u}_1 = b_2\mathbf{u}_2 + \ldots + b_m\mathbf{u}_m.$$

This equation is equivalent to

$$\mathbf{u}_1 - b_2\mathbf{u}_2 - \ldots - b_m\mathbf{u}_m = \mathbf{0}.$$

Note that $\mathbf{u}_1$ has 1 as coefficient, so we see that the zero vector is a non trivial linear combination of the vectors, which means by definition, that they are linearly dependent. This shows the 'backward direction'. □

2.4.9 Remark
Note that the previous theorem does *not* mean, that *every* vector in a linearly dependent set must be a linear combination of the others. There is just *at least one* that is a linear combination of the others.

We can collect the previous theorems, reinterpret them and summarize as follows:

2.4.10 Theorem
Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ and $\mathbf{b}$ be vectors in $\mathbb{R}^n$. Then the following statements are equivalent.
(a) The set $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ is linearly independent.
(b) The vector equation $x_1\mathbf{a}_1 + x_2\mathbf{x}_2 + \ldots + x_m\mathbf{a}_m = \mathbf{b}$ has at most one solution.

(c) The linear system corresponding to $\left[\begin{array}{cccc|c} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_m & \mathbf{b} \end{array}\right]$ *has at most one solution.*
(d) The equation $A\mathbf{x} = \mathbf{b}$ *with* $\left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_m \end{array}\right]$ *has at most one solution.*

Proof: (b) $\iff$ (c) $\iff$ (d) is obvious, because it is just rewriting linear systems in equation-/ matrix-form.
So we need to proof that from (a), (b) follows and vice versa. Let us assume that (a) is true and then show that (b) is true. To that end, we assume the contrary and lead that to a contradiction. So assume, that $x_1\mathbf{a}_1 + x_2\mathbf{x}_2 + \ldots + x_m\mathbf{a}_m = \mathbf{b}$ has more than one solution. We pick two different solutions, $(r_1, r_2, \ldots, r_m)$ and $(s_1, s_2, \ldots, s_m)$, so that

$$\begin{aligned} r_1\mathbf{a}_1 + r_2\mathbf{a}_2 + \ldots + r_m\mathbf{a}_m &= \mathbf{b} \\ s_1\mathbf{a}_1 + s_2\mathbf{a}_2 + \ldots + s_m\mathbf{a}_m &= \mathbf{b} \end{aligned}$$

Thus, we have

$$r_1\mathbf{a}_1 + r_2\mathbf{a}_2 + \ldots + r_m\mathbf{a}_m = s_1\mathbf{a}_1 + s_2\mathbf{a}_2 + \ldots + s_m\mathbf{a}_m,$$

and hence

$$(r_1 - s_1)\mathbf{a}_1 + (r_2 - s_2)\mathbf{a}_2 \ldots + (r_m - s_m)\mathbf{a}_m = \mathbf{0}.$$

Without loss of generality, we may assume that $r_1 \neq s_1$, so that we see, that there is a non trivial linear combination of the zero vector. But by assumption, $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ are linearly independent, which gives us the contradiction. Therefore, the assumption in the number of solutions must have been wrong.
Let us now assume (b) and show that (a) holds. Choose $\mathbf{b} = \mathbf{0}$. As (b) holds, we know that $x_1\mathbf{a}_1 + x_2\mathbf{x}_2 + \ldots + x_m\mathbf{a}_m = \mathbf{0}$ has at most one solution. As the trivial solution is already one, there cannot be another solution, which gives us per definition the linear independence of the vectors. □

2.4.11 Theorem (!THE BIG THEOREM!)
Let $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ *be a set of* n *vectors in* $\mathbb{R}^n$*. (Note that this time we have the same number* n *of vectors and entries in a column.) Furthermore, let* $A = \left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{array}\right]$*. Then the following are equivalent:*
(a) $\mathcal{A}$ *spans* $\mathbb{R}^n$*.*
(b) $\mathcal{A}$ *is linearly independent.*
(c) $A\mathbf{x} = \mathbf{b}$ *has a unique solution for all* $\mathbf{b}$ *in* $\mathbb{R}^n$

Proof: First we show, that (a) $\iff$ (b). If $\mathcal{A}$ is linearly dependent, then without loss of generality, $\mathbf{a}_1$ is a linear combination of the remaining vectors in $\mathcal{A}$ by Theorem 2.4.8. Hence, by Theorem 2.4.8, we have

$$\text{span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\} = \text{span}\{\mathbf{a}_2, \ldots, \mathbf{a}_n\}.$$

By assumption (a), we have

$$\text{span}\{\mathbf{a}_2, \ldots, \mathbf{a}_n\} = \mathbb{R}^n,$$

which is a contradiction to Theorem 2.3.7. Therefore, the assumption about $\mathcal{A}$ being linearly dependent must be wrong, and hence (a) $\Rightarrow$ (b).
For (b) $\Leftarrow$ (a), we may assume that $\mathcal{A}$ is linearly independent. If $\mathcal{A}$ does not span $\mathbb{R}^n$, we find a vector $\mathbf{a}$, which is not a linear combinations of vectors in $\mathcal{A}$. Therefore, Theorem 2.4.8 applies, and $\{\mathbf{a}, \mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is linearly independent. But these are $n+1$ linearly independent vectors in $\mathbb{R}^n$, which cannot be by Theorem 2.4.7, this is a contradiction. Hence (a) $\rightarrow$ (b).
We now show, that (a) $\Rightarrow$ (c). Note, as we have already shown (a) $\Rightarrow$ (b), that we can use (a) and (b). By Theorem 2.3.10(a), we now, that $A\mathbf{x} = \mathbf{b}$ has *at least* one solution for every $\mathbf{b}$ in $\mathbb{R}^n$. But Theorem 2.4.10(b), implies, that $A\mathbf{x} = \mathbf{b}$ has *at most* one solution. Hence, there is exactly one solution to the equation.
Finally, it remains to show (c) $\Rightarrow$ (a). So let (c) be true, so that there is a unique solution to *every* vector in $\mathbb{R}^n$. But this is just saying that any vector of $\mathbb{R}^n$ lies in the span of $\mathcal{A}$, hence (a) is true. □

Chapter 3

Matrices

3.1 Linear Transformations

So far we have always fixed a natural number n and then moved around within $\mathbb{R}^n$. We never passed a bridge from $\mathbb{R}^n$ to $\mathbb{R}^m$, say. But with what we are going to explore in this chapter, this will be possible. We will learn that certain functions will allow us to go from one vector space to another without losing too much structure.

3.1.1 Definition ((Co)-Domain, Image, Range, Linear Transformation/Homomorphism)
Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a function, that takes a vector of $\mathbb{R}^m$ as input and produces a vector in $\mathbb{R}^n$ as output. We then say, that $\mathbb{R}^m$ is the **domain of** T and $\mathbb{R}^n$ is the **codomain of** T. For $\mathbf{u} \in \mathbb{R}^m$, we call $T(\mathbf{u})$ the **image of u under** T. The set of the images of all vectors in the domain is called the **range** of T. Note that the range is a subset of the codomain.
A function $T : \mathbb{R}^m \to \mathbb{R}^n$ is a **linear transformation**, or a **linear homomorphism**, if for all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^m$, and all scalars c, the following holds:
(a) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$.
(b) $T(c\mathbf{u}) = cT(\mathbf{u})$.

3.1.2 Example
Check if the following functions are linear homomorphism.
(a) $T_1 : \mathbb{R}^1 \to \mathbb{R}^2, [u] \mapsto \begin{bmatrix} u \\ -3u \end{bmatrix}$.

Let us check the first condition. We need to take to vectors $[u_1], [u_2] \in \mathbb{R}$ and see if the function respects addition.

$$T_1([u_1+u_2]) = \begin{bmatrix} u_1 + u_2 \\ -3(u_1+u_2) \end{bmatrix} = \begin{bmatrix} u_1 \\ -3u_1 \end{bmatrix} + \begin{bmatrix} u_2 \\ -3u_2 \end{bmatrix} = T_1([u_1]+T_1([u_2])).$$

Let us check the second condition. Let $c \in \mathbb{R}$ be a scalar and let $[u] \in \mathbb{R}^1$.

$$T_1(c[u]) = T_1[cu] = \begin{bmatrix} cu \\ -3(cu) \end{bmatrix} = c \begin{bmatrix} u \\ -3u \end{bmatrix} = cT_1([u]).$$

Both condition are satisfied, T_1 is a linear homomorphism.
(b) $T_2 : \mathbb{R}^n \to \mathbb{R}^m, \mathbf{u} \mapsto \mathbf{0}$.
Let us check the conditions.

$$T_2(\mathbf{u}_1 + \mathbf{u}_2) = \mathbf{0} = \mathbf{0} + \mathbf{0} = T_2(\mathbf{u}_1) + T_2(\mathbf{u}_2).$$

Finally

$$cT_2(\mathbf{u}) = c\mathbf{0} = \mathbf{0} = T_2(c\mathbf{u}),$$

hence this is also a linear homomorphism.
(c) $T_3 : \mathbb{R}^2 \to \mathbb{R}^2, [u_1, u_2]^t \mapsto [u_1^2, u_2^2]^t$.
We know by the binomial equations that squaring does not preserve addition. So we are already expecting that the first condition will be violated.
Check the following vectors out:

$$T_3(\begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 4 \\ 4 \end{bmatrix}) = T_3(\begin{bmatrix} 6 \\ 7 \end{bmatrix}) = \begin{bmatrix} 36 \\ 49 \end{bmatrix},$$

but

$$T_3(\begin{bmatrix} 2 \\ 3 \end{bmatrix}) + T_3(\begin{bmatrix} 4 \\ 4 \end{bmatrix}) = \begin{bmatrix} 4 \\ 9 \end{bmatrix} + \begin{bmatrix} 16 \\ 16 \end{bmatrix} = \begin{bmatrix} 25 \\ 32 \end{bmatrix}.$$

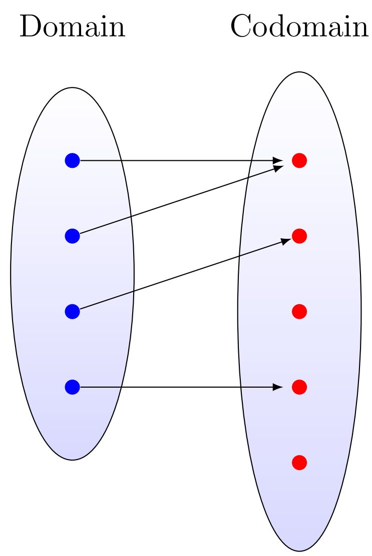
We can stop here because the first condition is not satisfied and we can already conclude that T_3 is not a linear homomorphism.
(d) $T_4 : \mathbb{R}^2 \to \mathbb{R}^2, [u_1, u_2]^t \mapsto \begin{bmatrix} u_1 + u_2 \\ 1 \end{bmatrix}$.
Let us check the first condition.

$$T_4(\begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \end{bmatrix}) = T_4(\begin{bmatrix} 4 \\ 4 \end{bmatrix}) = \begin{bmatrix} 4 \\ 1 \end{bmatrix},$$

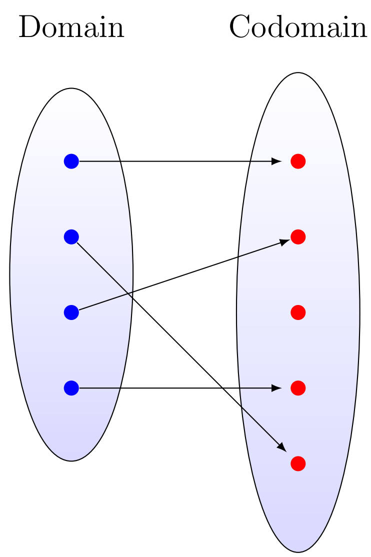
but

$$T_4(\begin{bmatrix} 2 \\ 2 \end{bmatrix}) + T_4(\begin{bmatrix} 2 \\ 2 \end{bmatrix}) = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

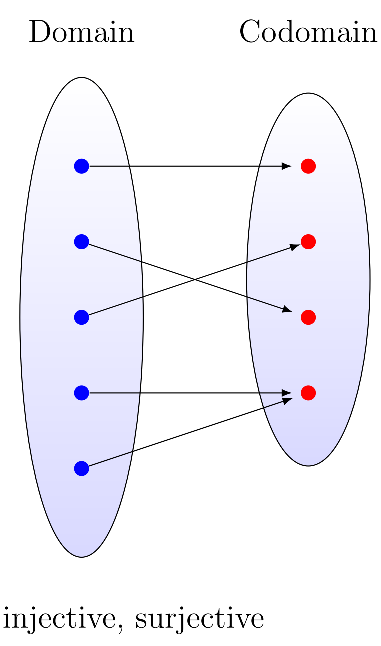
and hence it is not a linear homomorphism.
But check that it is in fact a homomorphism if we map the second component to 0 instead of 1.



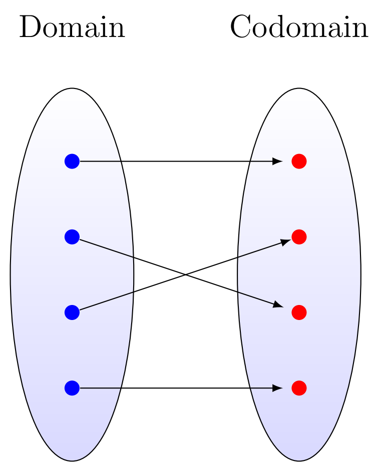
(a) not injective, not surjective



(b) injective, not surjective



(c) not injective, surjective



(d) injective, surjective

3.1.3 Definition (Injective/Surjective Homomorphism)
Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear homomorphism.
(a) T is called **injective** or **one-to-one**, if for every vector $\mathbf{w}$ in $\mathbb{R}^n$ there exists *at most* one vector $\mathbf{u}$ in $\mathbb{R}^m$ such that $T(\mathbf{u}) = \mathbf{w}$.
(b) T is called a **surjective homomorphism** or **onto**, if for *every* vector $\mathbf{w}$ in $\mathbb{R}^n$, there exists *at least* one vector $\mathbf{u}$ in $\mathbb{R}^m$, such that $T(\mathbf{u}) = \mathbf{w}$.
(c) T is called an **isomorphism** or **bijective**, if it is injective and surjective.

3.1.4 Lemma
Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear homomorphism. Then T is injective, if and only if $T(\mathbf{u}) = T(\mathbf{v})$ implies that $\mathbf{u} = \mathbf{v}$.

Proof: Put $\mathbf{w} := T(\mathbf{u})$. If T is injective, then $\mathbf{u}$ and $\mathbf{v}$ both are mapped to the same image, hence they must be equal by the definition of injectivity.
Assume now that there are two vectors $\mathbf{u}$ and $\mathbf{v}$ which have the same image $\mathbf{w}$ under T. That is, $T(\mathbf{u}) = T(\mathbf{v})$. By assumption, this means, that $\mathbf{u} = \mathbf{v}$, so that every possible image has at most one preimage, which is the definition for T being injective. □

3.1.5 Definition (Kernel)
Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a homomorphism. Then the **kernel of** T is defined to be the set of vectors $\mathbf{x}$ in $\mathbb{R}^m$, which satisfy

$$T(\mathbf{x}) = \mathbf{0}.$$

The kernel is denoted by $\ker(T)$.

3.1.6 Theorem
Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear homomorphism. Then T is injective if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution, in other words, if and only if $\ker(T) = \{\mathbf{0}\}$.

Proof:
If T is injective, then there is at most one solution to $T(\mathbf{x}) = \mathbf{0}$. But T is linear, so $\mathbf{0} = 0T(\mathbf{x}) = T(0\mathbf{x}) = T(\mathbf{0})$, so that $\mathbf{0}$ is already a solution. That means, that $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.
Let us now suppose that $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution. Let $\mathbf{u}, \mathbf{v}$ be vectors such that $T(\mathbf{u}) = T(\mathbf{v})$. Then $\mathbf{0} = T(\mathbf{u}) - T(\mathbf{v}) = T(\mathbf{u} - \mathbf{v})$. By our assumption, we conclude that $\mathbf{u} - \mathbf{v} = \mathbf{0}$, hence $\mathbf{u} = \mathbf{v}$. □

3.1.7 Remark (Tool!)
Determining whether a linear homomorphism is injective or not just means finding out, if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

3.1.8 Lemma
Suppose that $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_m \end{bmatrix}$ is an (n, m)-matrix, and let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

Then $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$.

3.1.9 Theorem (!Matrix Product as Homomorphism)
Let A be an (n, m)-matrix and define $T(\mathbf{x}) = A\mathbf{x}$. Then $T : \mathbb{R}^m \to \mathbb{R}^n$ is a linear homomorphism.

Proof: We look at the definition and see that we need to prove two properties: So let $\mathbf{u}, \mathbf{v}$ be two arbitrary vectors in $\mathbb{R}^m$.
(a) $T(\mathbf{u} + \mathbf{v}) = A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = T(\mathbf{u}) + T(\mathbf{v})$.
(b) $c(T(\mathbf{u})) = cA\mathbf{u} = (cA)\mathbf{u} = A(c\mathbf{u}) = T(c\mathbf{u})$.

□

The following theorem gives a tool for deciding whether a vector lies in the range of a linear transformation or not.

3.1.10 Theorem
Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_m \end{bmatrix}$ *be an* (n, m)*-matrix, and let* $T : \mathbb{R}^m \to \mathbb{R}^n, \mathbf{x} \mapsto A\mathbf{x}$ *be a linear transformation. Then*
(a) The vector $\mathbf{w}$ *is in the range of* T *if and only if* $A\mathbf{x} = \mathbf{w}$ *is a consistent linear system.*
(b) $\mathrm{range}(T) = \mathrm{span}\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$.

Proof: (a) A vector $\mathbf{w}$ is in the range of T, if and only if there exists a vector $\mathbf{v} \in \mathbb{R}^m$, such that

$$T(\mathbf{v}) = A\mathbf{v} = \mathbf{w}.$$

But this equation is true, if and only if $\mathbf{v}$ is a solution to the system

$$A\mathbf{x} = \mathbf{w}.$$

(b) By Theorem 2.3.10(a),(d), we know that $A\mathbf{x} = \mathbf{w}$ is consistent, if and only if $\mathbf{w}$ is in $\mathrm{span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$, which just means what the assertion states. □

Let us find conditions, for when a homomorphism is injective or surjective.

3.1.11 Theorem
Let A *be a* (n, m)*-matrix and define the homomorphism* $T : \mathbb{R}^m \to \mathbb{R}^n$, $\mathbf{x} \mapsto A\mathbf{x}$*. Then*
(a) T *is injective, if and only if the columns of* A *are linearly independent.*
(b) If $n < m$*, then* T *is not injective.*
(c) T *is surjective, if and only if the columns of* A *span the codomain* $\mathbb{R}^n$*.*
(d) If $n > m$*, then* T *is not surjective.*

Proof: (a) By Theorem 3.1.6, T is injective, if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution. By Theorem 2.4.3, $T(\mathbf{x})$ has only the trivial solution, if and only if the columns of A are linearly independent.
(b) If A has more columns than rows, we can apply Theorem 2.4.7. The columns of A are therefore linearly dependent. Hence, by (a) of this Theorem, T is not injective.
(c) T is surjective, if and only if its range is $\mathbb{R}^n$. By Theorem 3.1.10, the range of T equals the span of the columns of A. Hence, the range of T is the whole of $\mathbb{R}^n$, if the columns of A span $\mathbb{R}^n$.
(d) By Theorem 2.3.7 and the assumption $n > m$, the columns do not span $\mathbb{R}^n$, hence T is not onto. □

Let us try an example to see, how we tackle questions about injectivity and surjectivity.

3.1.12 Example

(a) Consider the homomorphism

$$T : \mathbb{R}^2 \to \mathbb{R}^3, \mathbf{x} \mapsto \begin{bmatrix} 1 & 3 \\ -1 & 0 \\ 3 & 3 \end{bmatrix} \mathbf{x}.$$

We determine, if T is injective or surjective. Let us first consider injectivity. We apply Theorem 3.1.11 and find the echelon form of

$$\left[\begin{array}{rr|r} 1 & 3 & 0 \\ -1 & 0 & 0 \\ 3 & 3 & 0 \end{array}\right] \sim \left[\begin{array}{rr|r} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

This means, that the associated linear system has only the trivial solution, so that the columns of A are linearly independent. Hence, we conclude that T is injective.

The reasoning for T not being surjective is easier: because $3 > 2$, T cannot be surjective.

(b) Is the homomorphism $T : \mathbb{R}^3 \to \mathbb{R}^3, \mathbf{x} \mapsto \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \end{bmatrix} \mathbf{x}$ surjective? By Theorem 3.1.11 we need to determine, if the columns of the matrix span $\mathbb{R}^3$. Let us find the echelon form of

$$\left[\begin{array}{rrr|r} 2 & 1 & 1 & \star \\ 1 & 2 & 0 & \star \\ 1 & 3 & 0 & \star \end{array}\right] \sim \left[\begin{array}{rrr|r} 2 & 1 & 1 & \star \\ 0 & 1 & -1/3 & \star \\ 0 & 0 & -2 & \star \end{array}\right].$$

Whatever vector $[\star, \star, \star]$ we start with, the system will always be consistent. We thus conclude, that T is surjective.

The following theorem is important. We introduced homomorphisms and we saw that $\mathbf{x} \mapsto A\mathbf{x}$ for a matrix A is a homomorphism. But the next theorem proves that *any* homomorphism can be written as a matrix-vector-product-homomorphism.

3.1.13 Theorem

Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a function (not yet a homomorphism!). There is an (n, m)-matrix A, such that $T(\mathbf{x}) = A\mathbf{x}$ for any $\mathbf{x} \in \mathbb{R}^n$, if and only if T is a homomorphism.

Proof: We have already proven one direction in Theorem 3.1.9. So let us assume that T is a homomorphism. We need to find a matrix A that satisfies $A\mathbf{x} = T(\mathbf{x})$ for any $\mathbf{x}$.
Consider the *standard vectors*

$$\mathbf{e}_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_3 := \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \mathbf{e}_n := \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

! Note that any vector in $\mathbb{R}^m$ can be written as a linear combination of the standard vectors, as can be seen as follows:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \ldots + x_m \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_m\mathbf{e}_m.$$

Now consider the matrix

$$A := \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \ldots & T(\mathbf{e}_n) \end{bmatrix}.$$

We then have from the linear property of T that

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_m\mathbf{e}_m) \\ &= x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \ldots + x_mT(\mathbf{e}_m) \\ &= A\mathbf{x}. \end{aligned}$$

3.1.14 Example
Show that

$$T : \mathbb{R}^2 \to \mathbb{R}^3, \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \begin{bmatrix} x_2 \\ 2x_1 - 3x_2 \\ x_1 + x_2 \end{bmatrix}$$

is a homomorphism.

We could verify the defining properties of homomorphisms (see Definition 3.1.1). Instead we find a matrix such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x}$ and apply Theorem 3.1.13.

$$T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} x_2 \\ 2x_1 - 3x_2 \\ x_1 + x_2 \end{bmatrix} = \begin{bmatrix} 0x_1 + x_2 \\ 2x_1 - 3x_2 \\ x_1 + x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 2 & -3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

So

$$A = \begin{bmatrix} 0 & 1 \\ 2 & -3 \\ 1 & 1 \end{bmatrix}$$

is the desired matrix and T is indeed a linear transformation.

3.1.15 Theorem (The Big Theorem, Version 2)
Let $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_n\} \subseteq \mathbb{R}^n$, $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}$ and $T : \mathbb{R}^n \to \mathbb{R}^n, \mathbf{x} \mapsto A\mathbf{x}$. Then the following is equivalent:
(a) $\mathcal{A}$ spans $\mathbb{R}^n$.
(b) $\mathcal{A}$ is linearly independent.
(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for any $\mathbf{b}$ in $\mathbb{R}^n$.
(d) T is surjective.
(e) T is injective.

Proof: In Theorem 2.4.11, we have already proven the equivalences (a) $\iff$ (b) $\iff$ (c). From Theorem 3.1.11(c) we see, that (a) and (d) are equivalent. From Theorem 3.1.11(a) we see that (b) and (e) are equivalent. □

3.2 Matrix Algebra

In Definition 1.4.1, we introduced matrices. We used matrices as a mean to deal with systems of linear equations. They provided a tool to find solutions to a linear system by performing Gauss elimination. We also introduced the matrix-vector product in Definition 2.1.8. This is already a first hint, that the set of matrices over $\mathbb{R}$ provides a very rich algebraic structure. We are going to explore these structures systematically in this section.

3.2.1 Definition
Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \text{ and } B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{bmatrix}$$

be two (n, m)-matrices over $\mathbb{R}$, and let $c \in \mathbb{R}$. Then addition and scalar multiplication of matrices are defines as follows:

(a)**Addition:** $$A + B = \begin{bmatrix} (a_{11} + b_{11}) & (a_{12} + b_{12}) & \dots & (a_{1m} + b_{1m}) \\ (a_{21} + b_{21}) & (a_{22} + b_{22}) & \dots & (a_{2m} + b_{2m}) \\ \vdots & \vdots & & \vdots \\ (a_{n1} + b_{n1}) & (a_{n2} + b_{n2}) & \dots & (a_{nm} + b_{nm}) \end{bmatrix}.$$

(b)**Scalar multiplication:** $$cA = \begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1m} \\ ca_{21} & ca_{22} & \dots & ca_{2m} \\ \vdots & \vdots & & \vdots \\ ca_{n1} & ca_{n2} & \dots & ca_{nm} \end{bmatrix}$$

Just as we did with vectors, we find some algebraic properties which matrices inherited from the corresponding properties in $\mathbb{R}$. Compare with Theorem 2.1.4.

3.2.2 Theorem
Let s, t be scalars, A, B and C be (n, m)-matrices. Let $\mathbf{0}_{nm}$ be the (n, m)- **zero-matrix**, *which has all 0's at entries. Then*
(a) $A + B = B + A$
(b) $s(A + B) = sA + sB$
(c) $(s + t)A = sA + sB$
(d) $(st)A = s(tA)$
(e) $A + \mathbf{0}_{nm} = A$

Proof: Mimic proof of Theorem 2.1.4. □
The operations 'addition' and 'scalar mulitplication', did not hit us by surprise. What we already knew about vectors is easily generalized to matrices.

What is really new is the following way to define a *new* and unexpected matrix product. Instead of a componentwise multiplication, we will introduce a convolution product. This new product reveals a lot of algebraic structure that inhibits the set of (n, m)-matrices.

3.2.3 Definition (Matrix-Matrix-Product)
Let A be an (n, k)-matrix and $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_m \end{bmatrix}$ be a (k, m)-matrix. Then the **matrix product** AB is the (n, m)-matrix given by

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \dots & A\mathbf{b}_m \end{bmatrix}.$$

3.2.4 Remark
Note that the matrix product for two matrices is *only* defined if the number of columns of the first matrix equals the number of rows of the second matrix.

3.2.5 Example
Let

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 4 & 2 \end{bmatrix}$$

and compute the matrix product AB.
We defined AB to be $\begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 \end{bmatrix}$. So let us compute the three matrix vector products:

$$\begin{aligned} A\mathbf{b}_1 &= 2\begin{bmatrix} 2 \\ -1 \end{bmatrix} + 0\begin{bmatrix} 0 \\ 3 \end{bmatrix} &= \begin{bmatrix} 4 \\ -2 \end{bmatrix} \\ A\mathbf{b}_2 &= 1\begin{bmatrix} 2 \\ -1 \end{bmatrix} + 4\begin{bmatrix} 0 \\ 3 \end{bmatrix} &= \begin{bmatrix} 2 \\ 11 \end{bmatrix}, \\ A\mathbf{b}_3 &= 0\begin{bmatrix} 2 \\ -1 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 3 \end{bmatrix} &= \begin{bmatrix} 0 \\ 6 \end{bmatrix} \end{aligned}$$

so that

$$AB = \begin{bmatrix} 4 & 2 & 0 \\ -2 & 11 & 6 \end{bmatrix}.$$

Even though the definition of the matrix product works fine, there is a formula, which makes it much more practical - even if you do not think so when seeing the formula. It is all about using your both hands in order to get used to the matrix product.

3.2.6 Lemma
Let A be an (n,k)-matrix and $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_m \end{bmatrix}$ be a (k,m)-matrix. Then the i,j-entry of $C = AB$ for $1 \le i \le n$ and $1 \le j \le m$ is given by

$$C_{ij} = \sum_{t=1}^{k} A_{it} B_{tj}.$$

Now we are ready to inevestigate the set of matrices with et another product and find out which structures are revealed.

3.2.7 Definition (Zero Matrix, Identity Matrix)
(a) The (n,m)-matrix having only zero entries is called the **zero** (n,m)**-matrix** and denoted by $\mathbf{0}_{nm}$.
(b) The **identity** (n,n)**-matrix** is of the form

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

and denoted by $\mathbf{I}_n$ or $\mathbf{I}$, if the dimensions are obvious.
(c) We introduce the short notation $[a_{ij}]$ for

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}.$$

3.2.8 Theorem
Let s be a scalar, and let A, B and C be matrices. Then each of the following holds, whenever the dimensions of the matrix are such that the operation is defined.
(a) $A(BC) = (AB)C$
(b) $A(B + C) = AB + AC$
(c) $(A + B)C = AC + BC$
(d) $s(AB) = (sA)B = A(sB)$
(e) $A\mathbf{I} = A$
(f) $\mathbf{I}A = A$

Proof: All proofs apply Lemma 3.2.6, so we illustrate only the parts (c) and (e).
(c) Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be (n,m)-matrices and let $C = [c_{ij}]$ be an (m,k)-matrix. Then the (ij)-entry of $(A+B)C$ is

$$(A+B)C_{ij} = \sum_{t=1}^{m}(A_{it}+B_{it})C_{tj} = \sum_{t=1}^{m}(A_{it}C_{tj}+B_{it}C_{tj}) = \sum_{t=1}^{m} A_{it}C_{tj}+\sum_{t=1}^{m} B_{it}C_{tj}.$$

But the last term is just the (ij)-entry of $AC+BC$, just what we wanted to proof.
(e) The (i,j)-entry of $A\mathbf{I}_m$ is by Lemma 3.2.6

$$(A\mathbf{I}_m)_{ij} = \sum_{t=1}^{m} A_{it}\mathbf{I}_{tj}.$$

But $\mathbf{I}_{tj}$ is only not zero if $t=j$, so that the sum can be displayed as

$$(A\mathbf{I}_m)_{ij} = A_{ij},$$

so the assertion follows. □
We will now see a phenonemon, that we have not encountered before. We need to get used to the fact that different products lead to different behaviour.

3.2.9 Theorem
Let A, B and C be matrices with dimensions so that all products are well defined.
(a) It is possible that $AB \neq BA$, i.e. the set of matrices is not **commutative** *under the matrix product.*
(b) $AB = \mathbf{0}$ does not *imply that $A = \mathbf{0}$ or $B = \mathbf{0}$.*
(c) $AC = BC$ does not *imply that $A = B$ or $C = \mathbf{0}$.*

Proof: (a) We give examples that prove the statement right. Consider

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 0 \end{bmatrix} \text{and } B = \begin{bmatrix} 0 & 1 \\ 4 & -1 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} -4 & 3 \\ 0 & -1 \end{bmatrix}$$

and
$$BA = \begin{bmatrix} -1 & 0 \\ 9 & -4 \end{bmatrix}.$$
(b) Consider
$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \text{ and } B = \begin{bmatrix} -4 & 6 \\ 2 & -3 \end{bmatrix}.$$
Then $AB = \mathbf{0}_{22}$, but neither A nor B is the zero-vector.
(c) Consider
$$A = \begin{bmatrix} -3 & 3 \\ 11 & -3 \end{bmatrix} \text{ and } B = \begin{bmatrix} -1 & 2 \\ 3 & 1 \end{bmatrix} \text{ and } C = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}.$$
Even though $A \neq B$ and C is not the zero matrix, we have $AC = BC$.

□

Let us continue defining operations on the set of matrices.

3.2.10 Definition
The **transpose** of an (n, m)-matrix A, denoted by A^t, A^T and also A^{tr}, is the (m, n)-matrix defined by $A^t_{ij} = A_{ji}$. Practically, it is interchanging rows and columns.

3.2.11 Example
(a) Consider
$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}.$$
Then
$$A^t = \begin{bmatrix} 1 & 5 & 9 \\ 2 & 6 & 10 \\ 3 & 7 & 11 \\ 4 & 8 & 12 \end{bmatrix}.$$
(b)
$$\begin{bmatrix} 2 \\ 3 \\ -1 \\ 2 \\ 4 \end{bmatrix}^t = [2 \ 3 \ -1 \ 2 \ 4].$$
We have the following properties for transposing.

3.2.12 Theorem
Let A, B be (n,m)-matrices, C be an (m,k)-matrix and s be a scalar. Then
(a) $(A+B)^t = A^t + B^t$
(b) $(sA)^t = sA^t$
(c) $(AC)^t = C^t A^t$

Proof: (a) and (b) are straightforward. (c), even not difficult regarding the idea, is very technical (apply Lemma 3.2.6) and therefore left out. □

3.2.13 Definition (Symmetric Matrices)
A matrix A that satisfies $A^t = A$, is called **symmetric**. Note that it must be an (n,n)-matrix.

3.2.14 Definition (Power of a matrix)
We define the **k-th power** of a matrix A to be the product

$$A^k = A \cdot A \cdots A,$$

where we have k factors.

3.2.15 Definition (Diagonal Matrix, Upper/Lower Triangular Matrix)
(a) A **diagonal (n,n)-matrix** is of the form

$$\begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ 0 & 0 & a_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}.$$

(b) An **upper triangular (n,n)-matrix** is of the form

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}.$$

(c) A **lower triangular** (n, n)**-matrix** is of the form

$$\begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ a_{21} & a_{22} & 0 & \dots & 0 \\ a_{31} & a_{32} & a_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}.$$

We introduced these matrices because they behave in a certain way under multiplication with matrices of the same shape.

3.2.16 Theorem
(a) If

$$A = \begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ 0 & 0 & a_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}.$$

is a diagonal matrix and $k \geq 1$ an integer, then

$$A^k = \begin{bmatrix} a_{11}^k & 0 & 0 & \dots & 0 \\ 0 & a_{22}^k & 0 & \dots & 0 \\ 0 & 0 & a_{33}^k & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn}^k \end{bmatrix}.$$

(b) If A, B are upper (lower) triangular (n, n)-matrices, then AB is also an upper (lower) triangular (n, n)-matrix. In particular, if $k \geq 1$ is an integer, then A^k is also an upper (lower) triangular (n, n)-matrix.

Proof: (a) We proof this assertion by induction. If $k = 1$, then $A = A^1 = A^k$ is a diagonal matrix, so the assertion is true for $k = 1$. Let us assume that the assertion is true for all integers up to $k - 1$. We need to proof that it is also true for k. Let us write $A^k = (A^{k-1}) \cdot A$. For A^{k-1}, the induction hypothesis holds and we know that this is a diagonal matrix.

$$\begin{aligned} A^k = A^{k-1}\cdot A &= \begin{bmatrix} a_{11}^{k-1} & 0 & 0 & \dots & 0 \\ 0 & a_{22}^{k-1} & 0 & \dots & 0 \\ 0 & 0 & a_{33}^{k-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn}^{k-1} \end{bmatrix} \begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ 0 & 0 & a_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}^{k} & 0 & 0 & \dots & 0 \\ 0 & a_{22}^{k} & 0 & \dots & 0 \\ 0 & 0 & a_{33}^{k} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn}^{k} \end{bmatrix}. \end{aligned}$$

The last equality results from just performing matrix multiplication.
(b) We proof the assertion for upper triangular matrices. The proof for the lower triangular matrices works just the same. We need to show that below the diagonal, there are only zero entries. The indices of any entry $(AB)_{ij}$ below the diagonal satisfy $i > j$. So let $i > j$ and determine what AB_{ij} is. By Lemma 3.2.6, we have

$$(AB)_{ij} = \sum_{t=1}^{n} a_{it}b_{tj}.$$

Now consider the following four cases:
(i) $i > t$: Then $a_{it} = 0$, so the product $a_{it}b_{tj} = 0$.
(ii) $t > j$: Then $b_{tj} = 0$, so is the product $a_{it}b_{tj}$.
(iii) $t \geq i$. But then we have $t \geq i > j$ by assumption, so that $b_{tj} = 0$, hence the product $a_{it}b_{tj} = 0$.
(iv) $j \geq t$. But then $i > j \geq t$, so that $a_{it} = 0$ in this case, hence the product $a_{it}b_{tj} = 0$.
These are all possible cases for t, which means, that

$$(AB)_{ij} = \sum_{t=1}^{n} a_{it}b_{tj} = 0$$

for $i > j$. □

3.2.17 Remark

As we have seen, the set of (n, n)-matrices satisfies the following properties for any matrices A, B, C and scalars $a, b \in \mathbb{R}$.

(a) $A + B = B + A$
(b) $A + \mathbf{0} = A$
(c) $(AB)C = A(BC)$
(d) $A\mathbf{I} = A$
(e) $A(B + C) = AB + AC$
(f) $(A + B)C = AC + BC$
(g) $a(A + B) = aA + aB$
(h) $(a + b)A = aA + bA$
(i) $(ab)A = a(bA)$
(j) $1A = A$
(k) $a(AB) = (aA)B = A(aB)$
Because of these properties, we call the set of (n, n)-matrices over $\mathbb{R}$ an **$\mathbb{R}$-algebra**.

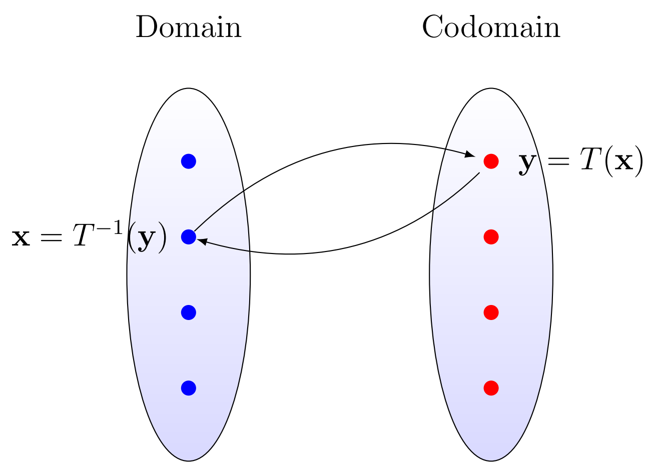
3.3 Inverses

In practical applications such as encoding data, encrypting messages, logisistics, linear homomorphisms are used to handle data. But the nature of these applications request, that we can get back from an image to the preimage. This is the scope of this section. We will determine the properties a homomorphism needs to have in order to "go back", i.e. in order to have an *inverse*. Moreover, we will learn how to determine the inverse of a homomorphism.

3.3.1 Definition (Invertible homomorphism, Inverse)
A homomorphism $T : \mathbb{R}^m \to \mathbb{R}^n$ is **invertible**, if T is an isomorphism, i.e. injective and surjective. When T is invertible, the **inverse** $T^{-1} : \mathbb{R}^n \to \mathbb{R}^m$ is defined by

$$T^{-1}(\mathbf{y}) = \mathbf{x} \text{ if and only if } T(\mathbf{x}) = \mathbf{y}.$$

What we have is the following picture.



3.3.2 Definition (Identity function)
A function f that satisfies $f(x) = x$ for all x in the domain of f is called the **identity function**. It is denoted by **id**.

3.3.3 Remark
If T is an invertible homomorphism, then the inverse function is uniquely determined. Its domain is the range of T, which is equal to the codomain of T, and its codomain is the domain of T. Moreover,

$$T(T^{-1})(\mathbf{y}) = \mathbf{y} \text{ and } T^{-1}(\mathbf{x}) = \mathbf{x},$$

hence, $T(T^{-1}) = \mathbf{id}$ is the identity on the range of T and $T^{-1}(T) = \mathbf{id}$ is the identity on the domain of T.

3.3.4 Theorem
Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a homomorphism. Then
(a) If T has an inverse, then $m = n$.
(b) If T is invertible, then T^{-1} is also an invertible homomorphism.

Proof: (a) If T has an inverse, then it is surjective and injective. By Theorem 3.1.11(b) and (d) we have $n \geq m$ and $n \leq m$, hence equality.
(b) We need to show that $T^{-1}(\mathbf{u}+\mathbf{v}) = T^{-1}(\mathbf{u})+T^{-1}(\mathbf{v})$ for all $\mathbf{v}, \mathbf{u}$. But T is an isomorphism, so there are unique $\tilde{\mathbf{u}}, \tilde{\mathbf{v}}$, such that $T(\tilde{\mathbf{u}}) = \mathbf{u}$ and $T(\tilde{\mathbf{v}}) = \mathbf{v}$.

Therefore, we have

$$\begin{aligned} T^{-1}(\mathbf{u}+\mathbf{v}) &= T^{-1}(T(\tilde{\mathbf{u}})) + T(\tilde{\mathbf{v}})) \\ &= T^{-1}(T(\tilde{\mathbf{u}}+\tilde{\mathbf{u}}) \\ &= \tilde{\mathbf{u}}+\tilde{\mathbf{v}} \\ &= T^{-1}(\mathbf{u}) + T^{-1}(\mathbf{v}). \end{aligned}$$

Moreover, we need to show that T^{-1} respects scalar multiplication. So similarly to the above, we have $T^{-1}(r\mathbf{u}) = T^{-1}(rT(\tilde{\mathbf{u}})) = T^{-1}(T(r\tilde{\mathbf{u}})) = r\tilde{\mathbf{u}} = r(T^{-1}(\mathbf{u}))$. □

Remember Theorem 3.1.13? Every homomorphism T can be written in matrix-form, meaning there is a matrix A such that $T(\mathbf{x}) = A\mathbf{x}$ for any vector $\mathbf{x}$ in the domain.

What does this mean in the world of inverses?

3.3.5 Lemma

Let $\mathbf{id}_{\mathbb{R}^n}$ be the identity on $\mathbb{R}^n$. Then the associated matrix for this homomorphism is the identity matrix.

Proof: For any vector in $\mathbb{R}^n$ we need to find a matrix A such that $\mathbf{id}(\mathbf{x}) = A\mathbf{x}$. By plugging in the standard vectors, the assertion follows. □

3.3.6 Theorem

Suppose that T is an invertible homomorphism with associated matrix A_T. Then the inverse T^{-1} has an associated matrix $A_{T^{-1}}$ which satisfies $A_T A_{T^{-1}} = \mathbf{I}$.

Proof: By Lemma 3.3.5 and Remark 3.3.3, the assertion follows. □

3.3.7 Definition (Invertible Matrix, Singular Matrix)

Let A be an (n, n)-matrix. Then A is called **invertible or nonsingular**, if there exists an (n, n)-matrix B such that $AB = \mathbf{I}_n$. If A has no inverse, it is called **singular**.

3.3.8 Corollary

Let T be an invertible homomorphism. Then the associated matrix A is also invertible. □

3.3.9 Theorem

Let A be an invertible matrix and let B be a matrix with $AB = \mathbf{I}$. Then $BA = \mathbf{I}_n$. Moreover, B is the unique matrix such that $AB = BA = \mathbf{I}$.

Proof: Let $\mathbf{x} \in \mathbb{R}^n$. Then

$$AB(A\mathbf{x}) = \mathbf{I}A(\mathbf{x}) = A\mathbf{x} \Rightarrow A(BA\mathbf{x}) = A\mathbf{x} \Rightarrow A(BA\mathbf{x} - \mathbf{x}) = \mathbf{0}.$$

But A is invertible, in particular it is injective. So $A\mathbf{y} = \mathbf{0}$ has only the trivial solution. Therefore, $A(BA\mathbf{x} - \mathbf{x}) = \mathbf{0}$ means that $BA\mathbf{x} - \mathbf{x}$ must be the trivial vector, hence $BA\mathbf{x} = \mathbf{x}$. Since $\mathbf{x} \in \mathbb{R}^n$ was arbitrary, we have that BA is the identity matrix.
Now suppose, that there is a matrix C with $AB = AC = \mathbf{I}$. Then

$$B(AB) = B(AC) \Rightarrow (BA)B = (BA)C \Rightarrow B = C,$$

as we have just proven that $BA = \mathbf{I}$. Hence $B = C$, so it is unique. □

3.3.10 Definition (Inverse matrix)
Let A be an invertible matrix. Then the uniquely determined matrix B such that $AB = BA = \mathbf{I}$ is called the **inverse of** A and is denoted by $B = A^{-1}$.

3.3.11 Theorem
Let A and B be invertible (n, n)-matrices, and C, D be (n, m)-matrices.
(a) A^{-1} is invertible with $(A^{-1})^{-1} = A$.
(b) AB is invertible with $(AB)^{-1} = B^{-1}A^{-1}$.
(c) If $AC = AD$, then $C = D$. (Compare that to Theorem 3.2.9).
(d) If $AC = \mathbf{0}_{nm}$, then $C = \mathbf{0}_{nm}$.

Proof: (a) We know that $A^{-1}A = \mathbf{I}$. So by Theorem 3.3.9, the inverse of the inverse, which is $(A^{-1})^{-1} = A$.
(b) Note that

$$(AB)(B^{-1}A^{-1}) = A\mathbf{I}A^{-1} = AA^{-1} = \mathbf{I}.$$

Therefore, the inverse of (AB) is $(AB)^{-1} = B^{-1}A^{-1}$.
(c) Consider $C = \mathbf{I}C = A^{-1}AC = A^{-1}AD = \mathbf{I}D = D$.
(d) As in (c). □
How do we find the inverse of a matrix. As we will see, there is an algorithm that provides a way to find an inverse of a matrix, if it exists.
For this, it is convenient to introduce the following notion.

3.3.12 Definition
The vector

$$\mathbf{e}_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^n,$$

where the nonzero entry is in the i-th position, is called the **i-th standard vector of** $\mathbb{R}^n$.

Let A be an invertible matrix and denote for the time being the inverse of A by $B = [\mathbf{b}_1 \ \dots \ \mathbf{b}_n]$. Then, by definition, we have $AB = \mathbf{I} = [\mathbf{e}_1 \ \dots \ \mathbf{e}_n]$. Remember the definition of a matrix product, Definition 3.2.3. From that, we can read off, that

$$A\mathbf{b}_1 = \mathbf{e}_1, \ \dots \ A\mathbf{b}_n = \mathbf{e}_n.$$

But this reads as "$\mathbf{b}_1$ is the solution to $A\mathbf{x} = \mathbf{e}_1$, ..., $\mathbf{b}_n$ is the solution to $A\mathbf{x} = \mathbf{e}_n$. Each of these n systems, could have been solved one at a time by performing Gauss-Jordan elimination to each of the systems

$$\left[\begin{array}{ccc|c} \mathbf{a}_1 & \dots & \mathbf{a}_n & \mathbf{e}_i \end{array} \right] \text{ for all } 1 \leq i \leq n.$$

But we can save some time and work by doing that simultaneously with one *large augmented matrix*:

$$\left[\begin{array}{ccc|ccc} \mathbf{a}_1 & \dots & \mathbf{a}_n & \mathbf{e}_1 & \dots & \mathbf{e}_n \end{array} \right].$$

If we are done with Gauss-Jordan elimination, then we will have

$$[A|\,\mathbf{I}_n] \text{ transformed to } [\mathbf{I}_n|A^{-1}\,].$$

3.3.13 Example
Find the inverse of

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 2 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix}.$$

So we need to Gauss-Jordan-eliminate the following large augmented matrix

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -2 & 1 & 0 & 0 \\ 2 & 1 & 2 & 0 & 1 & 0 \\ 2 & 1 & 3 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & -2 & 1 & 0 & 0 \\ 0 & 1 & 6 & -2 & 1 & 0 \\ 0 & 1 & 7 & -2 & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -2 & 1 & 0 & 0 \\ 0 & 1 & 6 & -2 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -2 & 2 \\ 0 & 1 & 0 & -2 & 7 & -6 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{array}\right]$$

We therefore conclude that the right hand side of the augmented matrix is the inverse

$$A^{-1} = \begin{bmatrix} 1 & -2 & 2 \\ -2 & 7 & -6 \\ 0 & -1 & 1 \end{bmatrix}$$

3.3.14 Example

How do we see, if a matrix does not have an inverse? Let us have a look at an example.

Does

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ -2 & 3 & 3 \end{bmatrix}$$

have an inverse? Start with the algorithm:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ -2 & 3 & 3 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 3 & 3 & 2 & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 & -3 & 1 \end{array}\right]$$

We see with the third row that this is an inconsistent system. It therefore has no solution, hence no inverse.

For $(2, 2)$-matrices there is a quick formula for finding the inverse. We will later learn how with the help of *determinants*, the formula is correct.

3.3.15 Remark
Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

be a $(2, 2)$-matrix. If $ad - bc \neq 0$, then the inverse of A can be computed by:

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Proof: Just compute AA^{-1} and see that it is the identity matrix. □
We had a lot of statements about linear systems relating (augmented) matrices. Let us see how this relation is specified with *inverse* matrices.
Let us finish the section with yet a further version of the Big Theorem.

3.3.16 Theorem (The Big Theorem, Version 3)
Let $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_n\} \subseteq \mathbb{R}^n$, $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}$ *and* $T : \mathbb{R}^n \to \mathbb{R}^n, \mathbf{x} \mapsto A\mathbf{x}$.
Then the following is equivalent:
(a) $\mathcal{A}$ *spans* $\mathbb{R}^n$.
(b) $\mathcal{A}$ *is linearly independent (remember: i.e.* $A\mathbf{x} = \mathbf{0}$ *has only one solution).*
(c) $A\mathbf{x} = \mathbf{b}$ *has a unique solution for any* $\mathbf{b}$ *in* $\mathbb{R}^n$ *given by* $\mathbf{x} = A^{-1}\mathbf{b}$.
(d) T *is surjective.*
(e) T *is injective.*
(f) A *is invertible.*

Proof: We have already the equivalences (a),(b),(c),(d),(e). Therefore we need to proof that T is injective if and only if A is invertible. But T is injective if and only if (we are in the setting of the Big Theorem "$n = m$") T is surjective, if and only if A is invertible. □

Chapter 4

Subspaces

4.1 Introduction to Subspaces

We have seen homomorphisms, that are surjective, and we have seen sets of vectors that span the whole of $\mathbb{R}^n$. But what if a homomorphism is *not* surjective? What if a set of vectors does not span the whole Euclidean space? Which structural properties do we then find in the range of that homomorphism and in the span of the vectors, respectively? This is the scope of this section, we will study so called *subspaces*.

4.1.1 Definition (subspace)
A subset S of $\mathbb{R}^n$ is a **subspace**, if S satisfies the following three conditions:
(i) S contains the zero vector $\mathbf{0}$.
(ii) If $\mathbf{u}$ and $\mathbf{v}$ are in S, then so is $\mathbf{u} + \mathbf{v}$.
(iii) If $\mathbf{u}$ is in S and r any real number, then $r\mathbf{u}$ is also in S.

4.1.2 Example
(a) ϕ, the empty set is **not** a subspace because it does not satisfy the first property fromthe definition of subspace.
(b) $\{\mathbf{0}\} \subset \mathbb{R}^n$ is a subspace.
(c) $\mathbb{R}^n$ is a subspace.
(d) The set of all vectors that consists of integer components is **not** a subsapce, because it violates (iii) of the definition. We sometimes say that (b) and (c) are the **trivial subspaces** of $\mathbb{R}^n$.

4.1.3 Example
Let us consider the set vectors S in $\mathbb{R}^3$, that have a zero in the third component, i.e.

$$S = \{\begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \mid a, b \in \mathbb{R}\}.$$

Is S a subspace? Let us check the defining properties:
(i) $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ is an element of S, because it has a zero in the last component.
(ii) Consider $\begin{bmatrix} a_1 \\ b_1 \\ 0 \end{bmatrix}, \begin{bmatrix} a_2 \\ b_2 \\ 0 \end{bmatrix}$. Then

$$\begin{bmatrix} a_1 \\ b_1 \\ 0 \end{bmatrix} + \begin{bmatrix} a_2 \\ b_2 \\ 0 \end{bmatrix} = \begin{bmatrix} a_1 + a_2 \\ b_1 + b_2 \\ 0 \end{bmatrix},$$

which is again an element of S.
(iii) Let r be a scalar and $\begin{bmatrix} a \\ b \\ 0 \end{bmatrix}$. Then

$$r \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} = \begin{bmatrix} ra \\ rb \\ 0 \end{bmatrix},$$

which again is an element in S.
Hence, we showed that S is indeed a subspace of $\mathbb{R}^3$.
If we had chosen the last component to be 1, then we easily would have seen that no defining property of a subspace would had been satisfied.

Which sets do we easlily recognize as subspaces?

4.1.4 Theorem
Let $S = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ *be a subset of* $\mathbb{R}^n$. *Then* S *is a subspace of* $\mathbb{R}^n$.

Proof: (a) Since $0\mathbf{u}_1 + \ldots + 0\mathbf{u}_m$ lies in the span, the zero vector is an element of the span.

(b) Let $\mathbf{v}, \mathbf{w}$ be elements of S. Then there are scalars $r_1, \ldots, r_n$ and $s_1, \ldots, s_n$ such that

$$\mathbf{v} = r_1\mathbf{u}_1 + \ldots + r_n\mathbf{u}_n \text{ and } \mathbf{w} = s_1\mathbf{u}_1 + \ldots + s_n\mathbf{u}_n.$$

The aum of the two vectors is then

$$\mathbf{u}+\mathbf{w} = r_1\mathbf{u}_1+\ldots+r_n\mathbf{u}_n+s_1\mathbf{u}_1+\ldots+s_n\mathbf{u}_n = (r_1+s_1)\mathbf{u}_1+\ldots+(r_n+s_n)\mathbf{u}_n.$$

But this last expression is easily identified as an element lying in the span of $\mathbf{u}_1, \ldots, \mathbf{u}_n$.
(c) Let c be an scalar and let $\mathbf{v} = r_1\mathbf{u}_1+\ldots+r_n\mathbf{u}_n$ be an element in S. Then

$$c\mathbf{v} = c(r_1\mathbf{u}_1 + \ldots + r_n\mathbf{u}_n) = (cr_1)\mathbf{u}_1 + \ldots + (cr_n)\mathbf{u}_n$$

is again represented as an element of the span. □

4.1.5 Example
Let

$$A = \begin{bmatrix} 1 & 3 & 4 & 0 \\ 0 & 2 & 4 & 4 \\ 1 & 1 & 0 & -4 \end{bmatrix}.$$

Let us determine the solution to the corresponding homogeneous system:

$$\left[\begin{array}{cccc|c} 1 & 3 & 4 & 0 & 0 \\ 0 & 2 & 4 & 4 & 0 \\ 1 & 1 & 0 & -4 & 0 \end{array}\right] \sim \left[\begin{array}{cccc|c} 1 & 3 & 4 & 0 & 0 \\ 0 & 2 & 4 & 4 & 0 \\ 0 & -2 & -4 & -4 & 0 \end{array}\right] \sim \left[\begin{array}{cccc|c} 1 & 3 & 4 & 0 & 0 \\ 0 & 2 & 4 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

We deduce, that x_3, x_4 are free variables, which means that we set them to be $x_3 = s_1$ and $x_4 = s_2$. Backward substitution gives us then $x_1 = 2s_1 + 6s_2$ and $x_2 = -2s_1 - 2s_2$. The general solution is therefore

$$\mathbf{x} = s_1 \begin{bmatrix} 2 \\ -2 \\ 1 \\ 0 \end{bmatrix} + s_2 \begin{bmatrix} 6 \\ -2 \\ 0 \\ 1 \end{bmatrix}.$$

This can also be written in the form

$$S = \text{span}\{\begin{bmatrix} 2 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 \\ -2 \\ 0 \\ 1 \end{bmatrix}\}.$$

Note also, that the zero vector corresponds to the trivial solution, which lies in S. By the algorithm in Remark 4.1.4, we thus conclude, that S is a subspace of $\mathbb{R}^3$.

The result of the previous example holds in general, as we see in the following theorem.

4.1.6 Theorem
If A is an (n, m)-matrix, then the set of solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$ is a subspace of $\mathbb{R}^m$.

Proof: (a) The zero vector represents the trivial solution and is hence an element of the set of solutions.
(b) Suppose $\mathbf{u}, \mathbf{v}$ are both solutions to the system. Then

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

which shows, that the sum of two solutions is again a solution to the homogeneous system.
(c) Let $c \in \mathbb{R}$ and $\mathbf{u}$ be a solution. Then

$$A(c\mathbf{u}) = c(A\mathbf{u}) = c\mathbf{0} = \mathbf{0},$$

which shows, that the set of solutions is also closed under scalar multiplication.
Overall, we have that the set of solutions is a subspace. □

4.1.7 Definition (Null Space)
If A is an (n, m)-matrix, then the set of solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$ is called the **null space** of A and denoted by $\text{null}(A)$.

Now that we have that the solutions to a system form a subspace, we can ask and investigate which other sets we have encountered so far also form subspaces. That leads us directly into homomorphisms, their range and their *kernel*.

4.1.8 Theorem
Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a homomorphism. Then the kernel of T is a subspace of the domain $\mathbb{R}^m$ and the range of T is a subspace of the codomain $\mathbb{R}^n$.

Proof: Let $A = [\mathbf{a}_1, \ldots \mathbf{a}_m]$ be the associated (n, m)-matrix wrt T. Then any element of the kernel of T satisfies $A\mathbf{x} = \mathbf{0}$. Thus, we see that elements of the kernel are solutions to the homogeneous system associated with A, i.e. we have

$$\ker(T) = \text{null}(A).$$

By Theorem 4.1.6, we thus have that the kernel is in fact a subspace of $\mathbb{R}^m$. In Theorem 3.1.10, we saw that the range of T and the span of the columns of A were the same, i.e.

$$\text{range}(T) = \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}.$$

□

4.1.9 Example

Determine the kernel and the range of the homomorphism $T : \mathbb{R}^2 \to \mathbb{R}^3$, $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \begin{bmatrix} 2x_1 - 10x_2 \\ -3x_1 + 15x_2 \\ x_1 - 5x_2 \end{bmatrix}$. The associated matrix wrt T is

$$A = \begin{bmatrix} 2 & -10 \\ -3 & 15 \\ 1 & -5 \end{bmatrix}.$$

Because $\ker(T) = \text{null}(A)$, we just need to find the solution to

$$\left[\begin{array}{cc|c} 2 & 10 & 0 \\ -3 & 15 & 0 \\ 1 & -5 & 0 \end{array}\right].$$

But this matrix is equivalent to

$$\left[\begin{array}{cc|c} 1 & -5 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right].$$

Any solution to this system is of the form

$$\mathbf{x} = s \begin{bmatrix} 5 \\ 1 \end{bmatrix},$$

which means that the kernel of T is

$$\ker(T) = \text{span}\{\begin{bmatrix} 5 \\ 1 \end{bmatrix}\}.$$

What is th range of T? By Theorem 3.1.10, the range of T is just the span of the columns of A. Hence we have

$$\text{range}(T) = \text{span}\{\begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} -10 \\ 15 \\ -5 \end{bmatrix}\}.$$

Let us close this section with the Big Theorem extended by the results of this section.

4.1.10 Theorem (Big Theorem, Version 4)
Let $\mathcal{A} := \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ be a set of n vectors in $\mathbb{R}^n$, let $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix}$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be given by $T(\mathbf{x}) = Ax$. Then the following are equivalent:
(a) $\mathcal{A}$ spans $\mathbb{R}^n$.
(b) $\mathcal{A}$ is linearly independent.
(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for any $\mathbf{b} \in \mathbb{R}^n$.
(d) T is surjective.
(e) T is injective.
(f) A is invertible.
(g) $\ker(T) = \{\mathbf{0}\}$.
(h) $\text{null}(A) = \{\mathbf{0}\}$.

Proof: (a) through (f) has already been proven in Theorem 3.3.16. Theorem 3.1.6 gives the last equivalence. □

4.2 Basis and Dimension

When we like to find a set of vectors in $\mathbb{R}^n$ such that their span is the whole of $\mathbb{R}^n$, we need to choose 'enough' vectors, just saying 'more is better' in that sense. On the other hand, when dealing with linearly independent sets, we rather choose only few such vectors, 'less is better'. As we will see in this section, there is a point that exactly meets both needs and both directions of adding vectors and taking vectors out.

4.2.1 Definition (!!!!Basis!!!!)
A set $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ is called a **basis** for a subspace S if
(a) $\mathcal{B}$ spans S.
(b) $\mathcal{B}$ is linearly independent.

Let us first look at some properties of bases.

4.2.2 Theorem
Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a basis for a subspace S. Then every vector $\mathbf{v}$ in S can be written as a linear combination

$$\mathbf{v} = r_1\mathbf{u}_1 + \ldots + r_m\mathbf{u}_m$$

in **exactly one** *way.*

Proof: $\mathcal{B}$ is per assumption a basis for S, hence, its vectors span S. Assume therefore there are two ways to present $\mathbf{v} \in S$:

$$\mathbf{u} = r_1\mathbf{u}_1 + \ldots + r_m\mathbf{u}_m \text{ and } \mathbf{u} = s_1\mathbf{u}_1 + \ldots + s_m\mathbf{u}_m.$$

Reorganizing this equation gives us

$$(r_1 - s_1)\mathbf{u}_1 + \ldots + (r_m - s_m)\mathbf{u}_m = \mathbf{0}.$$

But as $\mathbf{B}$ is assumed to be a basis it has linearly independent vectors, thus only the trivial solution to the previous equation. That means, that $r_1 = s_1, \ldots, r_m = s_m$. □

4.2.3 Example
Consider the set $\mathcal{A} := \{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\}$. Let $S = \text{span}(\mathcal{A})$. We easily see, that $\begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}$ is an element of S. But

$$\begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

The set $\mathcal{A}$ provides two different ways of representing the vector $\begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}$, hence $\mathcal{A}$ cannot be a basis for S.

4.2.4 Definition (Standard Basis)
Let

$$\mathbf{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

be the i-th standard vector for $1 \leq i \leq n$. Then $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is called the **standard basis** of $\mathbb{R}^n$.

Before we proceed we need a technical lemma.

4.2.5 Lemma
Let A and B be equivalent matrices. Then the subspace spanned by the rows of A is the same as the subspace spanned by the rows of B.

Is there always a basis for a subspace? Yes there is, and here is how you find it:

4.2.6 Theorem
Let $S = \text{span}\{\text{u}_1, \ldots, \text{u}_m\}$. A basis of S can be obtained in the following way:
STEP 1: *Use the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ to form the* **rows** *of a matrix A.*
STEP 2: *Transform A to echelon form B.*
STEP 3: *The non-zero rows give a basis for S.*

Proof: The new rows still span S, by Lemma 4.2.5. The linearly independence is given by the fact that we choose the non-zero rows which are in echelon form, giving us only the trivial solution when considering the associated homogenous system. Be careful: We are now stating linear independence for rows instead of columns! □

4.2.7 Example
Find a basis for the following subspace:

$$\text{span}\{\begin{bmatrix} 1 \\ 0 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ -6 \\ 5 \end{bmatrix}\}.$$

Let us put them into a matrix

$$\begin{bmatrix} 1 & 0 & -2 & 3 \\ -2 & 2 & 4 & 2 \\ 3 & -1 & -6 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 & 3 \\ 0 & 2 & 0 & 8 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We therfore have as a basis for the subspace:

$$\mathcal{B} = \{ \begin{bmatrix} 1 \\ 0 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \\ 8 \end{bmatrix} \}.$$

4.2.8 Lemma
Suppose $U = [\mathbf{u}_1 \ \dots \ \mathbf{u}_m]$ and $V = [\mathbf{v}_1 \ \dots \ \mathbf{v}_m]$ are two equivalent matrices. Then any linear dependence that exists among the vectors $\mathbf{u}_1, \dots, \mathbf{u}_m$ also exists among the vectors $\mathbf{v}_1, \dots, \mathbf{v}_m$. This means, that if WLOG $\mathbf{u}_1 = r_2\mathbf{u}_2 + \dots + r_m\mathbf{u}_m$, then $\mathbf{v}_1 = r_2\mathbf{v}_2 + \dots + r_m\mathbf{v}_m$.

There is another way of determining the basis of a subspace, involving the *columns* instead of the *rows*.

4.2.9 Theorem
Let $S = \text{span}\{\mathbf{u}_1, \dots \mathbf{u}_m\}$. Then the following algorithm provides a basis for S.
STEP 1: *Use the vectors $\mathbf{u}_1, \dots, \mathbf{u}_m$ to form the* **columns** *of a matrix A.*
STEP 2: *Transform A to echelon form B. The set of the pivot columns will be linearly independent (and the remaining will be linearly dependent form the pivot columns).*
STEP 3: *The set of columns of A corresponding to the pivot columns of B form a basis of S.*

4.2.10 Example
Find a basis for the following subspace:

$$\text{span}\{ \begin{bmatrix} 1 \\ 0 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ -6 \\ 5 \end{bmatrix} \}.$$

Following the previous algorithm, we transform the following matrix in echelon form:

$$\begin{bmatrix} 1 & -2 & 3 \\ 0 & 2 & -1 \\ -2 & 4 & -6 \\ 3 & 2 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \\ 0 & 8 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The pivot columns are the first and second column, so $\mathcal{B} := \{ \begin{bmatrix} 1 \\ 0 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 4 \\ 2 \end{bmatrix} \}$ is a basis for the subspace.

4.2.11 Remark

We obviously have two options of determining a basis for a given subspace. The first algorithm has the tendency to give back vectors with more zero entries, which is sometimes favorable. The second algorithm gives us a basis which is a subset of the original vectors. Deciding which algorithm to use depends on what you like to do with the basis afterwards.

4.2.12 Theorem

Let S be a subspace of $\mathbb{R}^n$. Then any basis of S has the same number of vectors.

Proof: We just give the idea to the proof. We assume the contrary, so that there are at least two bases, one of which has less elements than the other. Denote this basis by $\mathcal{U}$, the one with more elements by $\mathcal{V}$. By the definition of the basis, each element of $\mathcal{V}$ can be expressed as a linear combination of elements of $\mathcal{U}$. Consider then the homogeneous equation

$$a_1\mathbf{v}_1 + \ldots + a_m\mathbf{v}_m = \mathbf{0}$$

of elements of $\mathcal{V}$ and subsitute their expression in terms of elements of $\mathcal{U}$. Because of the linear independence of $\mathcal{U}$, this gives a system of homogeneous equation with mire variables than equations, hence infinitely many solutions, a contradiction to the linear independence of $\mathcal{V}$. □

4.2.13 Definition (Dimension)

Let S be a subspace of $\mathbb{R}^n$. Then the **dimension** of S is the uniquely determined number of vectors in any basis of S.

4.2.14 Remark
(a) The zero subspace $\{\mathbf{0}\}$ has no basis and its dimension is defined to be 0.
(b) One finds the dimension of a space, by determining a basis and counting its elements.

4.2.15 Theorem
Let $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a set of vectors in a subspace S of $\mathbb{R}^n$.
(a) If $\mathcal{U}$ is linearly independent, then either $\mathcal{U}$ is a basis for S or additional vectors can be added to $\mathcal{U}$ to form a basis of S.
(b) If $\mathcal{U}$ spans S, then either $\mathcal{U}$ is a basis for S or vectors can be removed from $\mathcal{U}$ to form a basis for S.

Proof: (a) If $\mathcal{U}$ spans S, it is a basis. If not, then select a vector $\mathbf{s}_1 \in S$ that does not lie in the span of $\mathcal{U}$ and add it to $\mathcal{U}$. Denote the new set by $\mathcal{U}_2$. As $\mathbf{s}_1$ is not in the span of $\mathcal{U}$, we know that $\mathcal{U}_2$ is linearly independent. If $\mathcal{U}_2$ spans S, then it is a basis. If not, repeat the process until the set spans S, which gives then a basis for S.
(b) Apply the algorithm of Theorem 4.2.9. □
Let us go through an example of how to find a basis.

4.2.16 Example
Let

$$\mathbf{x}_1 = \begin{bmatrix} 7 \\ 3 \\ -6 \\ 4 \end{bmatrix}, \quad A = \begin{bmatrix} -3 & 3 & -6 & -6 \\ 2 & 6 & 0 & -8 \\ 0 & -8 & 4 & 12 \\ -3 & -7 & -1 & 9 \\ 2 & 10 & -2 & -14 \end{bmatrix}.$$

Note that $\mathbf{x}_1$ lies in the null space of A. Find a basis of null(A) that includes $\mathbf{x}_1$. Find a basis for null(A).
After using Gauss algorithm we see, that the null space of A has the following form:

$$\text{null}(A) = \text{span}\{\begin{bmatrix} -1 \\ 3 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 2 \\ 0 \end{bmatrix}\}.$$

Finding a basis that includes $\mathbf{x}_1$, is now possible with the 'Column-Algorithm' described in Theorem 4.2.9.
We therefore consider the following matrices:

$$\begin{bmatrix} 7 & -1 & -3 \\ 3 & 3 & 1 \\ -6 & 0 & 2 \\ 4 & 2 & 0 \end{bmatrix} \sim \begin{bmatrix} 3 & 0 & -1 \\ 0 & 3 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The pivot columns of the previous matrix are the first and the second, hence we take the first and the second matrix of the original matrix to find that

$$\{ \begin{bmatrix} 7 \\ 3 \\ -6 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ 0 \\ 2 \end{bmatrix} \}$$

is a basis for null(A).

4.2.17 Theorem
Let $\mathcal{U}$ be a set of m vectors of a subspace S of dimension m. If $\mathcal{U}$ is either linearly independent or spans S, then $\mathcal{U}$ is a basis for S.

Proof: If $\mathcal{U}$ is linearly independent, but does not span S, then by Theorem 4.2.15, we can add an additional vector, so that $\mathcal{U}$ together with the new vector stays linearly independent. But this gives a linearly independent subset of S with more than m elements, which is a contradiction to the assumption that $\dim(S) = m$.
If $\mathcal{U}$ spans S, but is not linearly independent, there is a vector in $\mathcal{U}$, that lies in the span of the others. Remove this vector from $\mathcal{U}$ to obtain the set $\mathcal{U}_1$. Repeat this process until ou get to a linearly independent set $\mathcal{U}_*$, which still spans S. Hence $\mathcal{U}_*$ is a basis for S contradicting the assumption that the dimension is m. □

4.2.18 Theorem
Suppose that S_1, S_2 are both subspaces of $\mathbb{R}^n$ and that S_1 is a subset of S_2. Then $\dim(S_1) \leq \dim(S_2)$. Moreover, $\dim(S_1) = \dim(S_2)$ if and only if $S_1 = S_2$.

Proof: Assume that $\dim(S_1) > \dim(S_2)$. Let $\mathcal{B}_{S_1}$ be a basis for S_1. Then $S_1 = \text{span}(\mathcal{B}_1) \subseteq S_2$. Then by Theorem 4.2.15 (a), additional vectors can be added to $\mathcal{B}_1$ to obtain a basis for S_2. But this is a contradiction to the assumption $\dim(S_1) > \dim(S_2)$.
If $\dim(S_1) = \dim(S_2)$, then any basis for S_1 must be already a basis for S_2. □

4.2.19 Theorem
Let $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a set of vectors in a subspace S of dimension k.
(a) If $m < k$, then $\mathcal{U}$ does not span S.
(b) If $m > k$, then $\mathcal{U}$ is not linearly independent.

Proof: (a) Assume that $\mathcal{U}$ spans S. By Theorem 4.2.15(b) we conclude, that m must be greater than the dimension of S, a contradiction to $m < k$.
(b) Assume that $\mathcal{U}$ is linearly independent. By Theorem 4.2.15(a), we can add vectors to $\mathcal{U}$ to obtain a basis, a contradiction to $m > k$. □

4.2.20 Theorem (Big Theorem, Version 5)
Let $\mathcal{A} := \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ be a set of n vectors in $\mathbb{R}^n$, let $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix}$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be given by $T(\mathbf{x}) = Ax$. Then the following are equivalent:
(a) $\mathcal{A}$ spans $\mathbb{R}^n$.
(b) $\mathcal{A}$ is linearly independent.
(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for any $\mathbf{b} \in \mathbb{R}^n$.
(d) T is surjective.
(e) T is injective.
(f) A is invertible.
(g) $\ker(T) = \{\mathbf{0}\}$.
(h) $\mathrm{null}(A) = \{\mathbf{0}\}$.
(i) $\mathcal{A}$ is a basis for $\mathbb{R}^n$.

Proof:
We only need to prove the equivalence of (a)-(g) with (h). By Definition 4.2.1, (a) plus (b) are equivalent with (h). □

4.3 Row and Column Spaces

In the previous section we introduced two algorithms fo obtaining a basis of a subspace. One involved the *rows* of a matrix, the other the *columns*. In this section, we will learn how the columns and the rows of a matrix are structurally connected.

4.3.1 Definition (Row Space, Column Space)
Let A be an (n, m)-matrix.
(a) The **row space** of A is the subspace of $\mathbb{R}^m$ spanned by the row vectors of A and is denoted by $\mathrm{row}(A)$.

(b) The **column space** of A is the subspace of $\mathbb{R}^n$ spanned by the column vectors of A and is denoted by $\mathrm{col}(A)$.

4.3.2 Example
Consider the matrix

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

Then the row space is

$$\mathrm{span}\{[1,\ -1,\ 2,\ 1], [0,\ 1,\ 0,\ 1]\}$$

and the column space is

$$\mathrm{span}\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\}.$$

Note that in general the column space and the row space are **not** equal.

The algorithms for providing a basis can now be reformulated as follows:

4.3.3 Theorem
Let A be a matrix and B an echelon form of A.
(a) The nonzero rows of B form a basis for $\mathrm{row}(A)$.
(b) The columns of A corresponding to the pivot columns of B form a basis for $\mathrm{col}(A)$. □

Shedding a slightly different view on Theorem 4.2.6 and 4.2.9, and Theorem 4.3.3.

4.3.4 Theorem
For any matrix A, the dimension of the row space equals the dimension of the column space.

Proof: WLOG we may assume that A is already in echelon form. By Theorem 4.3.3(a), we know that the dimension of the row space of A is equal to the nonzero rows. But remembering the stair-like pattern form of matrices in echelon form, we conclude that each pivot nonzero row has exactly one pivot position, hence a pivot column. Thus, the number of pivot columns is equal to the number of nonzero rows. By Theorem 4.3.3(b), the statement follows. □

4.3.5 Definition (Rank)
Let A be a matrix. Then the **rank** of A is the dimension of the row (or column) space. It is denoted by rank(A).

Let us have a closer look at the numbers we can now attach to any matrix.

4.3.6 Example
Consider

$$\begin{bmatrix} -1 & 4 & 3 & 0 \\ 2 & -4 & -4 & 2 \\ 2 & 8 & 2 & 8 \end{bmatrix}.$$

What is the nullity and what is the rank of A?
For determining both numbers, we first need to find an equivalent matrix in echelon form. It is easy to see that A is equivalent to

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

The previous matrix has two nonzero rows, so that $\text{rank}(A) = 2$. On the other hand, if we associate A with a homogeneous system, we see that there are two free variables, hence $\text{nullity}(A) = 2$.

It is not by coincidence, that the nullity and the rank turned out to be 2 and 2. Here is why:

4.3.7 Theorem (!Rank-Nullity Theorem!)
Let A be an (n, m)-matrix. Then

$$\text{rank}(A) + \text{nullity}(A) = m.$$

Proof: Assume WLOG, that A is in echelon form. We have already seen that the number of pivot columns is equal to the rank of A. On the other hand, each *non*pivot column corresponds to a free variable. Hence the number of nonpivot columns equals the nullity of A. But the number of pivot columns plus the number of nonpivot columns is equal to the number of columns which is m. This translates to 'the rank plus the nullity equals m'. □

4.3.8 Example
Let A be a $(3, 15)$-matrix.
- The maximum rank of A is 3.
- The minimum nullity of A is 12.
In particular, the homomorphism that sends a vector $\mathbf{u} \in \mathbb{R}^{15}$ to $A\mathbf{u}$ $(\in \mathbb{R}^3)$ cannot be injective, because the nullity of A is at the same time the dimension of the kernel of that homomorphism.

Let us apply the previous theorem to homomorphisms.

4.3.9 Theorem
Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a homomorphism. Then

$$\dim(\ker(T)) + \dim(\text{range}(T)) = m.$$

Proof: Let A be the associated matrix to T. The nullity of A equals the dimension of the kernel of T. Moreover, the range of T equals the span of the columns of A, which is by definition $\text{col}(A)$. Hence, the statement follows. □

4.3.10 Theorem
Let A be an (n, m)-matrix and $\mathbf{b}$ be a vector in $\mathbb{R}^n$.
(a) The system $A\mathbf{x} = \mathbf{b}$ is consistent, if and only if $\mathbf{b}$ is $\text{col}(A)$.
(b) The system $A\mathbf{x} = \mathbf{b}$ has a unique solution, if and only if $\mathbf{b}$ is in $\text{col}(A)$ and the columns of A are linearly independent.

Proof: (a) This is true by Theorem 2.3.10 and the fact that any vector of the form $A\mathbf{u}$ is a linear combination of the columns of A, hence in $\text{col}(A)$.
(b) By (a) we have at least one solution, by Theorem 2.4.10, we have at most one solution, hence a unique solution. □

4.3.11 Theorem (Big Theorem, Version 6)
Let $\mathcal{A} := \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ be a set of n vectors in $\mathbb{R}^n$, let $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix}$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be given by $T(\mathbf{x}) = Ax$. Then the following are equivalent:
(a) $\mathcal{A}$ spans $\mathbb{R}^n$.
(b) $\mathcal{A}$ is linearly independent.
(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for any $\mathbf{b} \in \mathbb{R}^n$.
(d) T is surjective.
(e) T is injective.
(f) A is invertible.

(g) $\ker(T) = \{\mathbf{0}\}$.
(h) $\text{null}(A) = \{\mathbf{0}\}$.
(i)$\mathcal{A}$ *is a basis for* $\mathbb{R}^n$.
(j) $\text{col}(A) = \mathbb{R}^n$.
(k) $\text{row}(A) = \mathbb{R}^n$.
(l) $\text{rank}(A) = n$.

Proof: With version 5 we had already the equivalences (a)-(h). By Theorem 4.3.4 and Definition 4.3.5, we see that (i)-(k) are equivalent. Moreover, (a) and (i) are obviously equivalent as it states the same. □

Chapter 5

Determinants

5.1 The Determinant Function

We have already attached to numbers to a matrix, such as the rank and the nullity. In this chapter, we will introduce yet another powerful number to a square matrix - the *determinant*.
Before defining this new term we need to introduce some technicalities.

5.1.1 Definition
Let A be an (n,n)-matrix. Then M_{ij} denotes the $(n-1,n-1)$-matrix that we get from A by deleting the row and the column containing the entry a_{ij}.

5.1.2 Example
Determine $M_{2,3}$ and $M_{2,4}$ for the following matrix.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

The entry $a_{2,3}$ is equal to 7. We therefore have

$$M_{2,3} = \begin{bmatrix} 1 & 2 & 4 \\ 9 & 10 & 12 \\ 13 & 14 & 16 \end{bmatrix}.$$

The entry $a_{2,4}$ is equal to 8, hence

$$M_{2,4} = \begin{bmatrix} 1 & 2 & 3 \\ 9 & 10 & 11 \\ 13 & 14 & 15 \end{bmatrix}$$

5.1.3 Definition (Determinant, Cofactor)
(a) Let $A = [a_{11}]$ be a $(1,1)$-matrix. Then the **determinant** of A is given by

$$\det(A) = | a_{11} | := a_{11}.$$

(b) Let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

be a $(2,2)$-matrix. Then the **determinant** of A is given by

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

(c) Let

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

be a $(3,3)$-matrix. Then the **determinant** of A is given by

$$\begin{aligned} \det(A) &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{31}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{23}a_{32}a_{11} - a_{33}a_{12}a_{21}. \end{aligned}$$

(d) Let A be an (n,n)-matrix and let a_{ij} be its (i,j)th entry. Then the **cofactor** C_{ij} of a_{ij} is given by

$$C_{ij} = (-1)^{i+j} \det(M_{ij}).$$

The determinant $\det(M_{ij})$ is called the **minor of** a_{ij}.

(e) Let n be an integer and fix **either** some $1 \leq i \leq n$ **or** $1 \leq j \leq n$. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

be an (n, n) matrix. Then the **determinant** of A is given by
either
(i) $\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in}$, where C_{ij} are the cofactors of $a_{i1}, a_{i2}, \ldots, a_{in}$, respectively, ('Expand across row i'),
or
(ii) $\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \ldots + a_{nj}C_{nj}$,
where C_{ij} are the cofactors of $a_{1j}, a_{2j}, \ldots, a_{nj}$, respectively, ('Expand down column j').
These formulas are referred to collectively as the **cofactor expansions**.

5.1.4 Remark
Note that the definition of a general determinant is used *recursively*, because in its definition it is reduced to the determinant of an $(n-1, n-1)$-matrix.

5.1.5 Example
Find $\det(A)$ for

$$A = \begin{bmatrix} 3 & 0 & -2 & -1 \\ -1 & 2 & 5 & 4 \\ 6 & 0 & -5 & -3 \\ -6 & 0 & 4 & 2 \end{bmatrix}$$

Let us first find the M_{1j}'s.

$$M_{1,1} = \begin{bmatrix} 2 & 5 & 4 \\ 0 & -5 & -3 \\ 0 & 4 & 2 \end{bmatrix} \qquad M_{1,2} = \begin{bmatrix} -1 & 5 & 4 \\ 6 & -5 & -3 \\ -6 & 4 & 2 \end{bmatrix}$$

$$M_{1,3} = \begin{bmatrix} -1 & 2 & 4 \\ 6 & 0 & -3 \\ -6 & 0 & 2 \end{bmatrix} \qquad M_{1,4} = \begin{bmatrix} -1 & 2 & 5 \\ 6 & 0 & -5 \\ -6 & 0 & 4 \end{bmatrix}$$

We can now use the formula of Definition 5.1.3(c), to get:

$$\begin{aligned}
\det(M_{1,1}) &= -5\cdot 2\cdot 2-3\cdot 5\cdot 0+4\cdot 0\cdot 4-4\cdot(-5)\cdot 0-4\cdot(-3)\cdot 2-2\cdot 0\cdot 5=4\\
\det(M_{1,2}) &= 10+90+96-120-12-60=4\\
\det(M_{1,3}) &= 0+36+0-0-0-24=12\\
\det(M_{1,4}) &= 0+60+0-0-0-48=12
\end{aligned}$$

Using Definition 5.1.3(d), we get

$$\det(A)=3\cdot(-1)^{1+1}\cdot 4+0\cdot(-1)^{1+2}\cdot 4+(-2)\cdot(-1)^{1+3}(12)+(-1)\cdot(-1)^{1+4}\cdot 12=12-24+12=0.$$

Now that we have understood the consept of calculating the determinant of a matrix, we give a generalized version of such a calculation. This comes in very handy, if there are rows or columns with many zeros.

5.1.6 Theorem
For any integer n we have $\det(I_n)=1$.

Proof: We proof that by induction.
$n=1$: We have $I_1=[1]$, so that $\det(I_1)=1$.
Let the statement be true for $n-1$. Then

$$\det(I_n)=1\cdot C_{1,1}+0\cdot C_{1,2}+\ldots+0\cdot C_{1,n}=C_{1,1}=(-1)^2\det(M_{1,1})=\det(M_{1,1}).$$

But $M_{1,1}=I_{n-1}$, and using the assumption of the induction, we get

$$\det(I_n)=\det(M_{1,1})=\det(I_{n-1})=1.$$

□

Why bothering with introducing a complicated formula all around the entries of a square matrix? Because the determinant features some really powerful tools and properties attached to matrices. Here are a few:

5.1.7 Theorem
Let A,B be (n,n)-matrices.
(a) A is invertible, if and only if $\det(A)\neq 0$.
(b) $\det(A\cdot B)=\det(A)\cdot\det(B)$.
(c) If A is invertible, then

$$\det(A^{-1})=\frac{1}{\det(A)}.$$

(d) $\det(A^t) = \det(A)$.

□

5.1.8 Example
Find the determinant of

$$A = \begin{bmatrix} -2 & 1 & 0 & 13 \\ 1 & 1 & 0 & 7 \\ -7 & 3 & 0 & 2 \\ 2 & 1 & 0 & 9 \end{bmatrix}$$

We expand down the third column, which gives us:
$\det(A) = 0 \cdot C_{1,3} + 0 \cdot C_{2,3} + 0 \cdot C_{3,3} + 0 \cdot C_{4,3} = 0$, hence we see that the matrix is not invertible.

What did we learn from the previous example?

5.1.9 Theorem
Let A be a square matrix.
(a) If A has a zero column or zero row, then $\det(A) = 0$.
(b) If A has two identical rows or columns then $\det(A) = 0$.

Proof: (a) Is clear by Theorem **??**.
(b) We use induction.
The statement makes only sense for $n \geq 2$.
Let A is a $(2,2)$-matrix with two identical rows or columns. WLOG let A have two identical rows, so that

$$\begin{bmatrix} a & b \\ a & b \end{bmatrix}.$$

Hence, $\det(A) = ab - ab = 0$.
Let the assertion be true for $n-1$ and assume that A is an (n,n)-matrix with two identical rows. Take a row that is different from those two identical ones and expand across this row. We will get a sum of determinants of $(n-1, n-1)$-matrices with two identical rows. We thus have by induction a sum of zeros. □
There is one special case when calculating the determinant is particularly easy.

5.1.10 Theorem
Let A be a triangular (n,n)-matrix. Then $\det(A)$ *is the product of the terms along the diagonal.*

Proof: We use induction. If $n = 1$, then the statement is trivial. Let the statement be true for $n-1$ and assume that WLOG A is an upper triangular (n,n)-matrix. We expand down the first column. This will get to

$$\det(A) = a_{1,1} \cdot C_{1,1}.$$

$C_{1,1}$ is the determinant of an upper triangular $(n-1, n-1)$-matrix. By induction, the statement follows. □

5.1.11 Theorem (The Big Theorem, Version 7)
Let $\mathcal{A} := \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ be a set of n vectors in $\mathbb{R}^n$, let $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix}$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be given by $T(\mathbf{x}) = Ax$. Then the following are equivalent:
(a) $\mathcal{A}$ spans $\mathbb{R}^n$.
(b) $\mathcal{A}$ is linearly independent.
(c) $A\mathbf{x} = \mathbf{b}$ has a unique solution for any $\mathbf{b} \in \mathbb{R}^n$.
(d) T is surjective.
(e) T is injective.
(f) A is invertible.
(g) $\ker(T) = \{\mathbf{0}\}$.
(h) $\text{null}(A) = \{0\}$.
(i) $\mathcal{A}$ is a basis for $\mathbb{R}^n$.
(j) $\text{col}(A) = \mathbb{R}^n$.
(k) $\text{row}(A) = \mathbb{R}^n$.
(l) $\text{rank}(A) = n$.
(m) $\det(A) \neq 0$.

Proof: We have already the equivalences (a)-(k). The equivalence (l) is included by Theorem 5.1.7