Notes On Linear Algebra

Nikos Apostolakis

The City University of New York
Email address: nikolaos.apostolakis@bcc.cuny.edu

2010 *Mathematics Subject Classification.* Primary

Thanks to Nick's Pizza for their timely deliveries.

ABSTRACT. Linear Algebra the Way I Like it.

Contents

CHAPTER 1

Linear systems

We say that two equations are *equivalent* if they have the same *solution set*, and we use the symbol $\iff$ to denote equivalency of equations. For example

$$3x = 9 \iff x = 3,$$

since both equations have the same solution set, namely $\{3\}$. We use the symbol $\implies$ to indicate that every solution of the equation on the left side is also a solution of the equation at the right side. For example

$$x = 3 \implies x^2 = 9.$$

Notice that it is not true that

$$x^2 = 9 \implies x = 3,$$

because -3 is a solution of the left equation but not of the right.

For an equation with more than one variables a solution is an *assignment* of a value to each of the variables that make the equation true. For example assigning $x = 3$ and $y = 4$ is a solution of the equation

$$10x + 3y = 42.$$

Usually there is an implicit order among the variables and we use *ordered tuples* to denote solutions. If our variables are x, y, and z then we write $(1, -2, 4)$ to denote the assignment $x = 1$, $y = -2$ and $z = 4$.

REMARK 1. Notice that whether two equations are equivalent depends on the *domain of definition*, in other words where the variables are supposed to vary. For example if x is a real variable (i.e. $x \in \mathbb{R}$) then

$$x^3 = 1 \iff x = 1.$$

But if x is a complex variable (i.e. $x \in \mathbb{C}$) then these equations are not equivalent because there are three cubic roots of unity.

1.1. Linear systems

DEFINITION 1. A *linear equation* with n variables $x_1, x_2, \ldots, x_n$ is an equation that is an equation of the form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = c,$$

where $a_1, \ldots, a_n$ and c are real numbers[1].

The numbers $a_1, \cdots, a_n$ are called the *coefficients* and c is called the *constant*.

If the constant is 0 we say that the equation is *homogeneous*.

In this part of the class we'll study *systems of linear equations*, namely we'll address the questions:

- How can we solve a linear system?
- What sets appear as solution sets of linear systems?

Let's wet our appetite by looking at a single linear equation.

[1] In this equation the symbols $a_1, \ldots, a_n$ and c are *parameters* while $x_1, \ldots, x_n$ are *variables*. Unlike variables, parameters are considered to have constant (but unspecified) values.

1.1.1. One variable. A linear equation of one variable has the form

$$a\,x = c. \tag{1.1}$$

We have two cases:

(a) **Non-zero coefficient.** If $a \neq 0$ then we can divide both sides by a (or equivalently multiply by a^{-1}):

$$a\,x = c \iff x = \frac{c}{a}.$$

So in this case we have a *unique solution*.

(b) **Zero coefficient.** If $a = 0$ we have two subcases:

  (a) **Non-zero constant.** If $c \neq 0$ then there are no solutions, in other words the solution set is the empty set $\varnothing$.

  (b) **Zero constant.** If $c = 0$ then all numbers are solutions, in other words the solution set is the set of real numbers $\mathbb{R}$.

In summary we have:

The solution set of $a\,x = c$

The solution set of a linear equation with one variables is

- a point, or
- the empty set $\varnothing$, or
- the whole line $\mathbb{R}$.

The case of non-zero coefficient is the *generic* case, most linear equations have a unique solution. “Wait a minute”, I here you say, “what do you mean *most*?”. Here is what I mean: we can represent the linear equation $a\,x = c$ by the point $(a, c) \in \mathbb{R}^2$, and conversely we can think of any point of $\mathbb{R}$ as representing a linear equation. So the point $(3, 4)$ represents the equation $3\,x = 4$ and the point $(0, 3)$ represents the equation $0\,x = 3$.

So we identified the set of linear equations with the Cartesian plane$\mathbb{R}^2$, the coefficient a in horizontal axis and the constant c in the vertical. The equations with zero coefficient are then represented by the vertical axis, a one-dimensional[2] subspace of the two-dimensional space. The equation $0x = 0$ is represented by a single point $(0, 0)$ a zero-dimensional subspace. “Most” points are outside the vertical axis, so most equations have a unique solution. Furthermore, the generic equation that doesn’t have a unique solution has no solutions at all.

1.1.2. Two variables. A linear equation with two variables, say x, y has the form

$$a\,x + b\,y = c, \tag{1.2}$$

with $a, b, c \in \mathbb{R}$.

Let’s first look at a particular equation, for example

$$2\,x - 3\,y = 0. \tag{1.3}$$

If we divide by the coefficient of y and move the x-term to the right side we get the equivalent equation

$$y = \frac{2}{3}x.$$

[2] Later in the class we will define what this means.

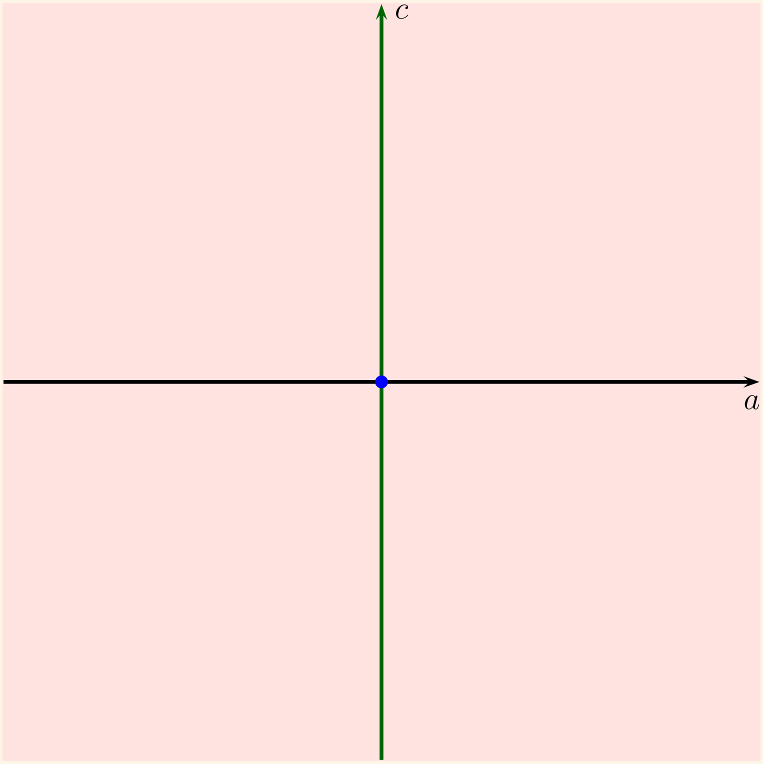


FIGURE 1. The space of linear equations with one variable.

The solution set of this equation, obviously, consists of all pairs where the second coordinate is two-thirds of the first coordinate. So the solution set is

$$S = \left\{ \left(x, \frac{2}{3}x \right) : x \in \mathbb{R} \right\}.$$

We can write the solution set in parametric form as follows

$$\left\{ \begin{aligned} x &= t \\ y &= \frac{2}{3}t \end{aligned} \right. \qquad t \in \mathbb{R}. \tag{1.4}$$

This form makes it clear that the solution set is one-dimensional, in the sense that a solution is completely determined once we choose a value for t.

Using *vector* notation we can express the solution set as

$$\begin{pmatrix} x \\ y \end{pmatrix} = t \begin{pmatrix} 1 \\ \frac{2}{3} \end{pmatrix}. \tag{1.5}$$

We will explain this in more detail later, but for the moment here is a quick explanation. We write coordinates vertically as columns: instead of (x, y) we write $\begin{pmatrix} x \\ y \end{pmatrix}$ and instead of $(1, 2/3)$ we write $\begin{pmatrix} 1 \\ 2/3 \end{pmatrix}$. Later in the class we will say that these are *column vectors*. In the right hand side of (1.5) we have *scalar multiplication*: we multiply a number and a vector by multiplying each coordinate of the vector with the number. So,

$$t\begin{pmatrix}1\\ \frac{2}{3}\end{pmatrix}=\begin{pmatrix}t\\ \frac{2}{3}t\end{pmatrix}.$$

Finally two vectors are equal if their corresponding coordinates are equal. Equation (1.5) is therefore just a rewriting of the system of equations (1.4).

The operation of *vector addition* for column vectors is also defined coordinate-wise:

$$\begin{pmatrix}a\\ b\end{pmatrix}+\begin{pmatrix}c\\ d\end{pmatrix}=\begin{pmatrix}a+c\\ b+d\end{pmatrix}.$$

The solution set S is a special subset of $\mathbb{R}^2$. It has two special properties, namely, it is closed under scalar multiplication and vector addition. This means that if we multiply a solution by a number the result is a solution, and if we add two solutions we get another solution. Indeed, if s is a real number we have

$$s\begin{pmatrix}t\\ \frac{2}{3}t\end{pmatrix}=\begin{pmatrix}st\\ \frac{2}{3}st\end{pmatrix}=st\begin{pmatrix}1\\ \frac{2}{3}\end{pmatrix},$$

so a scalar times a solution is a solution. And,

$$\begin{pmatrix}t_1\\ \frac{2}{3}t_1\end{pmatrix}+\begin{pmatrix}t_2\\ \frac{2}{3}t_2\end{pmatrix}=\begin{pmatrix}t_1+t_2\\ \frac{2}{3}t_1+\frac{2}{3}t_2\end{pmatrix}=\begin{pmatrix}t_1+t_2\\ \frac{2}{3}\left(t_1+t_2\right)\end{pmatrix},$$

so adding two solutions gives a solution. These two properties can be summarized by saying:

$$\boxed{S \text{ is a Vector Space.}}$$

Actually the solution set of any homogeneous linear equation is closed under scalar multiplication and vector addition.

THEOREM 1.1.1. *The solution set of any* homogeneous *linear equation is closed under scalar multiplication and vector addition.*

PROOF. Let

$$a_1x_1+\ldots+a_nx_n=0 \tag{1.6}$$

be a homogeneous equation and $(v_1,\ldots,v_n)$, $(w_1,\ldots,w_n)$ be two solutions. This means that

$$a_1v_1+\ldots+a_nv_n=0,\ \text{and } a_1w_1+\ldots+a_nw_n=0.$$

Adding these two equations we get

$$a_1v_1+\ldots+a_nv_n+a_1w_1+\ldots+a_nw_n=0.$$

Now taking common factors gives

$$a_1\left(v_1+w_1\right)+\ldots+a_n\left(v_n+w_n\right)=0.$$

Therefore $(v_1+w_1,\ldots,v_n+w_n)$ is a solution of (1.6).

Now let t be any number, then

$$a_1tv_1+\ldots+a_ntv_n=t\left(a_1v_1+\ldots+a_nv_n\right)=0,$$

therefore $s\left(v_1,\ldots,v_n\right)$ is a solution of (1.6). □

Consider now the equation

$$2\,x - 3\,y = 6. \tag{1.7}$$

Notice that this equation has the same coefficients as Equation (1.3). Entirely similarly as before we have that the solution set is

$$S' = \left\{\left(x, \frac{2}{3}x\right) + 6 : x \in \mathbb{R}\right\}.$$

In parametric form the solution is

$$\begin{cases} x &= t \\ y &= \dfrac{2}{3}t + 2 \end{cases} \qquad t \in \mathbb{R}, \tag{1.8}$$

and in vector notation:

$$\begin{pmatrix} x \\ y \end{pmatrix} = t\begin{pmatrix} 1 \\ \frac{2}{3} \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix}. \tag{1.9}$$

Equations (1.5) and (1.9) are very similar, they differ by the vector $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$. Where did that come from?

The answer will be revealed if we graph the equations, see Figure 2.

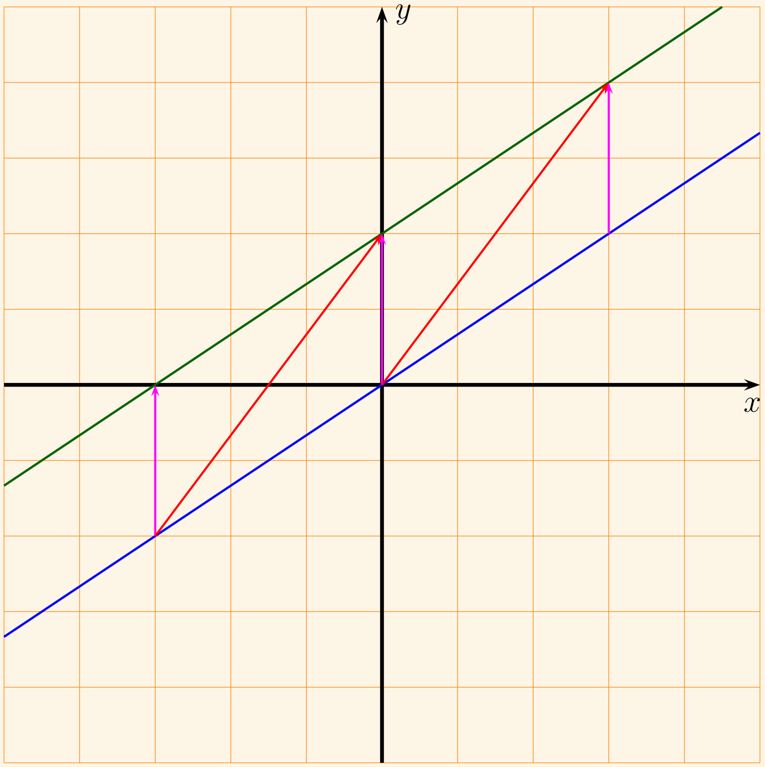


FIGURE 2. The solution sets of Equations (1.3) and (1.7).

We see then that $(0, 2)$ is the y–intercept of the line with equation (1.7). That's where $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$ came from. Geometrically, Equation (1.9) says that the graph the green line is obtained from the blue line by a vertical *translation* of two units.

There is nothing special about the y-intercept: take any other point of the blue line, for example the point with coordinates $(3,4)$. If we translate the blue line using the vector with components $(3,4)$ we will again get the green line. This connection is explored further in Section 1.2.3.

In general if at least one of the coefficients in non-zero the equation (1.2) has a one-dimensional solution set. Indeed, if $a \neq 0$, we can divide by a and move the y-term to the right side to get

$$x = -\frac{b}{a}y + \frac{c}{a}.$$

Similarly as above, we get that the general solution is

$$\begin{pmatrix} x \\ y \end{pmatrix} = t\begin{pmatrix} \frac{b}{a} \\ 1 \end{pmatrix} + \begin{pmatrix} \frac{c}{a} \\ 0 \end{pmatrix}. \tag{1.10}$$

1.1.3. The case of zero coefficients. In the trivial case $a = b = 0$, we have the equation

$$0\,x + 0\,y = c.$$

Clearly if $c \neq 0$ there are no solutions, and if $c = 0$ all points $(x,y) \in \mathbb{R}^2$ are solutions.

The solution set of $a\,x + b\,y = c$

The solution set of a linear equation with two variables is

- a line, or
- the empty set $\varnothing$, or
- the whole plane $\mathbb{R}^2$.

Three or more variables. Let's again look at a generic example first. Consider the equation

$$2\,x + 3\,y - z = 1.$$

Solving for z we get

$$z = 2\,x + 3\,y - 1.$$

The solution set

$$S = \left\{(x, y, 2\,x + 3\,y - 1) : (x,y) \in \mathbb{R}^2\right\}$$

has now two free parameters

Using vector notation we have

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = s\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + t\begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}. \tag{1.11}$$

As in the one variable and two variable cases, if all the coefficients are zero we either have no solutions (when the constant is non-zero), or the solution set is $\mathbb{R}^3$ (when the constant is zero).

Clearly this pattern continues in all dimensions. The solution set of a generic[3] linear equation with n unknowns has $n-1$ independent parameters. If all coefficients are 0 then if the constant is non-zero the solution set is empty, and if the constant is 0 the solution set is $\mathbb{R}^n$.

[3] i.e. with at least one non-zero coefficient.

The solution set of $a_1 x_1 + \cdots + a_n x_n = c$

The solution set of a linear equation with n variables is

- an $(n-1)$-dimensional subspace, or
- the empty set $\varnothing$, or
- the whole space $\mathbb{R}^n$.

1.1.4. Systems of linear equations. An $m \times n$ *linear system* is a collection of m linear equations with n variables. So,

$$\begin{cases} 2x - 3y + 4z = 0 \\ x + y - z = -6 \end{cases}$$

is a 2×3 system, while

$$\begin{cases} x - y = 5 \\ -3x + 2y = 2 \\ 9x + \pi y = -2 \end{cases}$$

is a 3×2 system. A *solution* of a system is a *common* solution of all the equations, and we say that two systems are equivalent if they have the same solution sets.

Elementary row operations

- Interchange two equations.
- Multiply one equation by a non-zero scalar.
- Replace E_k by $E_k + E_\ell$.

THEOREM 1.1.2. *Application of an elementary row operation gives an equivalent system.*

PROOF. The first two operations don't change the solution set of any equation, so the resulting system is equivalent to the original.

Let S be the original system and S' the system that we get by replacing E_k with $E_k + E_\ell$. It's easy to see that $S \implies S'$. Indeed, a common solution of E_k and E_ℓ is also a solution of $E_k + E_\ell$.

Now notice that we can go from S' to S by multiplying E_ℓ with -1 and adding it to $E_k + E_\ell$. Therefore $S' \implies S$ as well. □

REMARK 2 (**An often used combination**). In practice we often perform the following combination of the second and third operation:

(a) multiply E_k by a non-zero scalar λ_k,
(b) multiply E_ℓ by a non-zero scalar λ_ℓ,
(c) replace E_k with $E'_k + E'_\ell$,
(d) change E'_ℓ back to E_ℓ by multiplying it with λ_ℓ^{-1}.

The combined effect of these row operations is to replace E_k by $\lambda_k E_k + \lambda_\ell E_\ell$.

EXAMPLE 1. Consider the following 2×2 system

$$\begin{cases} 2x + 3y = 5 \\ 7x - 3y = 4 \end{cases}.$$

Multiply the first equation by -7:

$$\begin{cases} 14x + 21y = 35 \\ 7x - 3y = 4 \end{cases}.$$

Multiply the second equation by -2:

$$\begin{cases} 14x + 21y = 35 \\ -14x + 6y = -8 \end{cases}.$$

Replace the second equation by the sum of the first and the second:

$$\begin{cases} 14x + 21y = 35 \\ 27y = 27 \end{cases}.$$

Divide the first equation by 7:

$$\begin{cases} 2x + 3y = 5 \\ 27y = 27 \end{cases}.$$

Now divide the second equation by 27:

$$\begin{cases} 2x + 3y = 5 \\ y = 1 \end{cases}.$$

Now let's multiply the second equation by -3 and add it to the first in one step:

$$\begin{cases} 2x = 2 \\ y = 1 \end{cases}.$$

Finally we divide the first equation by 2:

$$\begin{cases} x = 1 \\ y = 1 \end{cases}.$$

We arrived at a system whose solution set is obvious: $S = \{(1,1)\}$.

It turns out that every linear system can be solved by applying a finite number of elementary row operations. Example 1 contains all the ingredients for an algorithm that solves all linear systems.

EXAMPLE 2. Let's solve the system

$$\begin{cases} 2x - y + 3z = 1 \\ -4x + 7y + 5z = 13 \end{cases}.$$

We first add 7 times the first equation to the second:

$$\begin{cases} 2x - y + 3z = 1 \\ 10x + 26z = 30 \end{cases}.$$

Now add the second equation to -5 times the first, and then divide the second equation by 10:

$$\begin{cases} - y + 11z = 25 \\ x + \frac{13}{10}z = 3 \end{cases}.$$

Now we multiply the first equation with -1 and (for aesthetic reasons) we interchange the equations:

$$\begin{cases} x \qquad + \frac{13}{10}z = 3 \\ \qquad y - 11z = -25 \end{cases}.$$

The final step is to move the z-terms to the right side:

$$\begin{cases} x \qquad = -\frac{13}{10}z + 3 \\ \qquad y = 11z - 25 \end{cases}.$$

So we have a one-dimensional solution set. In vector form:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = t \begin{pmatrix} -\frac{13}{10} \\ 11 \\ 1 \end{pmatrix} + \begin{pmatrix} 3 \\ -25 \\ 0 \end{pmatrix}.$$

Let's also see what can happen when we have more equations than unknowns.

EXAMPLE 3. Consider the 3×2 system:

$$\begin{cases} x - \ y = 5 \\ -3x + 2y = 2 \\ 9x + 7y = -2 \end{cases}.$$

We use the first equation to eliminate x from the second and third. To do this we add 3 times the first equation to the second, and -9 times the first equation to the third.

$$\begin{cases} x - \quad y = 5 \\ \ \ - \quad y = 17 \\ \qquad 16y = -47 \end{cases}.$$

Solving the second and third equations for y we get

$$\begin{cases} x - y = 5 \\ \qquad y = -17 \\ \qquad y = \frac{47}{16} \end{cases}.$$

The second and third equations are in contradiction, they cannot both be true. Therefore the system has no solutions.

Problems in many areas of mathematics (and other sciences) reduce to solving linear systems.

EXAMPLE 4 (**Finding the equation of a line)**. Find the line that passes through $(3,7)$ and $(-4,2)$.

SOLUTION. Let

$$a\,x + b\,y + c = 0$$

be the equation of the line, where a, b, and c are real numbers and at least one of the a, b is non-zero. Substituting the coordinates of the given points we get the system

$$\begin{cases} 3a + 7b + c = 0 \\ -4a + 2b + c = 0 \end{cases}.$$

Multiplying the first equation by 4 and the second by 3 gives

$$\begin{cases} 12a + 28b + 4c = 0 \\ -12a + 6b + 3c = 0 \end{cases}.$$

We then replace the second equation with the sum of the two equations, and multiply the first by $1/4$ and we get

$$\begin{cases} 3a + 7b + c = 0 \\ 34b + 7c = 0 \end{cases}.$$

We've eliminated a from the second equation, and now we'll eliminate b from the first. Now replace the first equation by 34 times the first equation plus -7 times the first:

$$\begin{cases} 102a - 15c = 0 \\ 34b + 7c = 0 \end{cases}.$$

Now divide the first equation by 102 and the second by 34 coefficients to get

$$\begin{cases} a - 5c/34 = 0 \\ b + 7c/34 = 0 \end{cases}.$$

This means that a one-dimensional solution set:

$$S = \left\{ \left(\frac{5}{34}c, -\frac{7}{34}c, c \right) : c \in \mathbb{R} \right\}.$$

When $c = 0$ we get the solution $(0,0,0)$ that doesn't satisfy the requirement that at least one of a, b is non-zero. So any equation of the form

$$\frac{5c}{34}x - \frac{7c}{34}y + c = 0, \qquad c \neq 0$$

is an equation of the line that passes through this two points. The simplest of all these equations is, arguably, obtained for $c = 34$:

$$5x - 7y + 34 = 0.$$

□

EXAMPLE 5 (**Determining a quadratic polynomial by three values**). For the polynomial $p(x) = a\,x^2 + b\,x + c$ we have that $p(1) = 3$, $p(-1) = 1$, and $p(2) = 10$. Find the coefficients of p.

SOLUTION. We have the system,

$$\begin{cases} a + b + c = 3 \\ a - b + c = 1 \\ 4a + 2b + c = 10 \end{cases}.$$

Rather than working with the system itself we will work with its *augmented matrix*:

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & -1 & 1 & 1 \\ 4 & 2 & 1 & 10 \end{array}\right)$$

Think of it like this: we pretend that the variables a, b, and c as well as the additions symbols are invisible and that the equal signs "=" have been replaced by vertical bars.

We use the following strategy: First we get an *upper triangular* matrix: we use a_{11} to kill all the other entries in the first column. Then we use a_{22} to kill everything bellow it, and so on until we get all entries below the diagonal to be 0.

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & -1 & 1 & 1 \\ 4 & 2 & 1 & 10 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & -2 & 0 & = 2 \\ 0 & -2 & -3 & -2 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & -2 & 0 & = 2 \\ 0 & 0 & -3 & 0 \end{array}\right)$$

The next step is then to go back and kill all the entries above the diagonal until we are are left with a *diagonal matrix*. We will start with the lowest diagonal entry a_{33} and we use it to kill a_{23} and a_{13}.

In our case, a_{23} is already 0, so we go to a_{13}: we multiply the third row by $1/3$ and add it to the first. Next we go to a_{22} and use it to kill a_{12}: we multiply the second row by $1/2$ and add it to the first.

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & 3 \\ 0 & -2 & 0 & -2 \\ 0 & 0 & -3 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & -2 & 0 & -2 \\ 0 & 0 & -3 & 0 \end{array}\right)$$

Now that we have a diagonal matrix we can easily solve, just divide each row by its first non-zero entry:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

So the solution of the system is $a = 2$, $b = 1$, and $c = 0$. So our polynomial is

$$p(x) = 2x^2 + x.$$

We can verify that indeed, $p(1) = 3$, $p(-1) = 1$, and $p(2) = 10$. □

EXAMPLE 6. Let's again consider a quadratic binomial $p(x) = a\,x^2 + b\,x + c$, and suppose that we now are given that $p(1) = 2$, $p(-1) = -2$, and $p(2) = 4$. What is the polynomial now?

SOLUTION. Entirely similarly as before we get the system:

$$\begin{cases} a + b + c = 2 \\ a - b + c = -2 \\ 4a + 2b + c = 4 \end{cases}.$$

with /augmented/ matrix

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 1 & -1 & 1 & -2 \\ 4 & 2 & 1 & 4 \end{array}\right)$$

As before we want to first obtain a triangular matrix.

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

So we get the solution $a = 0$, $b = 2$, and $c = 0$. Even though the system has a solution the polynomial we obtain $p(x) = 2x$ is not really quadratic. □

REMARK 3. Notice that the two systems in the previous two examples have the same coefficients and that the procedure we used to solve them was identical: we performed *the exact same* row operations. So even though the solutions are different the solution sets have the same *nature*: they both consist of a single solution.

EXAMPLE 7. Consider the 3×3 system:

$$\begin{cases} x - 3y + 2z = 4 \\ 2x + 5y - z = -3 \\ 3x + 2y + z = 1 \end{cases}$$

Let's again do our thing.

$$\left(\begin{array}{ccc|c} 1 & -3 & 2 & 4 \\ 2 & 5 & -1 & -3 \\ 3 & 2 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -3 & 2 & 4 \\ 0 & 11 & -5 & -11 \\ 0 & 11 & -5 & -11 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -3 & 2 & 4 \\ 0 & 11 & -5 & -11 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Now let's divide the second row by 11.

$$\left(\begin{array}{ccc|c} 1 & -3 & 2 & 4 \\ 0 & 1 & -5/11 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 7/11 & 1 \\ 0 & 1 & -5/11 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Notice that the last row is all zeros. What does this mean? If we make the variables visible again the last equation is now the trivial equation

$$0\,x + 0\,y + 0\,z = 0.$$

This is a tautology[4], and its presence does not really affect the solution set. So we might as well delete the third row to get the system

$$\begin{cases} x \qquad + \frac{7}{11}z = 1 \\ \qquad y - \frac{5}{11}z = -1 \end{cases}.$$

So we have a one-parameter family of solutions. That is, the solution set is 1-*dimensional*:

$$S = \left\{ \left(1 - \frac{7}{11}z, -1 + \frac{5}{11}z, z\right) : z \in \mathbb{R} \right\}.$$

We can write this in "vector form" as follows:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = t \begin{pmatrix} -7/11 \\ 5/11 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}.$$

EXAMPLE 8. Let's solve to solve the system

$$\begin{cases} 2x_1 + 3x_2 - 3x_3 + 5x_4 = 2 \\ -4x_1 + 7x_2 + x_3 = -7 \\ 3x_2 + 2x_4 = 1 \\ -2x_1 + 13x_2 - 2x_3 + 7x_4 = 10 \end{cases}.$$

We have the augmented matrix

$$\left(\begin{array}{cccc|c} 2 & 3 & -3 & 5 & 2 \\ -4 & 7 & 1 & 0 & -7 \\ 0 & 3 & 0 & 2 & 1 \\ -2 & 13 & -2 & 7 & 10 \end{array}\right).$$

We use $a_{11} = 2$ to kill all other entries in the column and get

[4] This means that the equation is true for all values of the variables

$$\left(\begin{array}{cccc|c} 2 & 3 & -3 & 5 & 2 \\ 9 & 13 & -5 & -10 & -11 \\ 0 & 3 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 12 \end{array}\right).$$

Look at the last row

$$\left(\begin{array}{cccc|c} 0 & 0 & 0 & 0 & 12 \end{array}\right)$$

all the coefficients are 0 but the constant is non-zero. If we make the variables visible again we see that the last equation is:

$$0\,x_1 + 0\,x_2 + 0\,x_3 + 0\,x_4 = 12.$$

This equation has no solutions, and so the system has no solutions either. The solution set is thus the empty set $\varnothing$.

The last two examples show that rows with all but, possibly, the last entries 0 are important.

The importance of zeros

If in the process of solving a linear system we arrive at an augmented matrix with a row of the form

$$\left(\begin{array}{cccc|c} 0 & 0 & \dots & 0 & c \end{array}\right)$$

then

- If $c \neq 0$ the system is *inconsistent*.
- If $c = 0$ we can delete that row from the matrix.

Before continuing with the theory (and practice) of linear systems we take a detour to properly introduce matrices. In our first encounter, matrices appeared to be just a convenient book-keeping device, but appearances are deceptive sometimes. Matrices play a fundamental role in linear algebra.

1.2. Matrices of linear systems

The *matrix form* of an $m \times n$ linear system

$$\left\{\begin{array}{lcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & c_2 \\ \quad\vdots \qquad\qquad \vdots & \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & c_m \end{array}\right.$$

is

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix},$$

or in more compact form

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

A is called the *matrix* of the system, $\mathbf{x}$ the *vector of unknowns*, and $\mathbf{c}$ the vector of constants. The *augmented matrix* of the system is the matrix A with an extra column that contains the constants.

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & c_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & c_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & c_m \end{array}\right).$$

The algorithm for solving a linear system consists of using elementary row operations to transform the augmented matrix of the system into a special form, the so-called *row-echelon* form. Roughly speaking, a matrix in row-echelon form exhibits a staircase pattern[5].

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 11 & 3 & -6 \\ 0 & -9 & 3 \\ 0 & 0 & 4 \end{pmatrix}, \quad \begin{pmatrix} -8 & -11 & 32 & 5 \\ 0 & 0 & 1 & -9 \\ 0 & 0 & 0 & 33 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 3 & -7 & 0 & 1 & 0 \\ 0 & 0 & 2 & -42 & 6 & 11 \\ 0 & 0 & 0 & 5 & -3 & -69 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

DEFINITION 2. A *zero row* is a row with all entries 0. The *leading entry* of a non-zero row is the first non-zero entry in that row.

(Reduced) Echelon form

We say that the matrix $A = (a_{ij})$ is in echelon form if it satisfies the following two conditions:

(a) The zero rows are at the bottom of the matrix.
(b) All the entries below the leading entry of a non-zero row are 0.
(c) The leading entry of a non-zero row is in a column to the right of any leading entry above it.

We say that a matrix is in *reduced echelon form* if it is in echelon form, and it satisfies the following two additional properties:

(c) All leading entries are equal to 1.
(d) If a column contains a leading 1, all other entries in that column are 0.

If the augmented matrix of a system is in echelon form then the system is easy to solve, by using *back-substitution*.

EXAMPLE 9. Consider the system with augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 2 & -2 \end{array}\right).$$

The corresponding system is

[5] The term *echelon* comes from the French word “échelle” that means “ladder”.

$$\begin{cases} x + 2y + 3z = 0 \\ y + z = 2 \\ 2z = -2 \end{cases}.$$

The last equation is practically solved: dividing by 2 gives $z = -1$. We now substitute the value of z back to the first and second equation:

$$\begin{cases} x + 2y \quad - 3 = 0 \\ y \quad - 1 = 2 \\ z \quad = -1 \end{cases}.$$

We then solve the second equation and we find $y = 3$. Substituting back into the first equation gives

$$\begin{cases} x \qquad + 3 = 0 \\ y \qquad = 3 \\ z \quad = -1 \end{cases}.$$

We finally solve the first equation to get

$$\begin{cases} x \qquad = -3 \\ y \quad = 3 \\ z = -1 \end{cases}.$$

On the other hand, a system whose augmented matrix is in reduced echelon form is super-easy to solve, in fact it's solved already!

EXAMPLE 10. Consider the system with augmented matrix

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 3 & 6 \\ 0 & 1 & 0 & -7 & 0 \\ 0 & 0 & 1 & 0 & -3 \end{array}\right).$$

The system is

$$\begin{cases} x_1 \qquad + 3x_4 = 6 \\ x_2 \quad - 7x_4 = 0 \\ x_3 \qquad = -3 \end{cases},$$

and all we need to do to solve it is to move the terms containing the free variable x_4 to the right hand side:

$$\begin{cases} x_1 \qquad = -3x_4 + 6 \\ x_2 \quad = 7x_4 \\ x_3 = -3 \end{cases}.$$

From these two examples it is clear that if we are able to put the augmented matrix of a system into echelon form (reduced or not) then we can solve it. We will shortly see that we can put any matrix in (reduced) echelon form, and that the procedure for doing so is *algorithmic*, we have actually being applying this procedure already. So we have two slightly different methods for solving linear systems: either we stop once we get any echelon form, and use back substitution, or we go all the way to reduced echelon form. The first method is called *Gauss Elimination* and the second *Gauss-Jordan Elimination*.

DEFINITION 3. We say that two matrices A and B are *row equivalent*, and write $A \sim B$, if B is obtained from A after the application of finitely many elementary row operations.

THEOREM 1.2.1. *Row equivalence is an equivalence relation. In other words, it enjoys the following properties:*

(a) It is reflexive. *This means that every matrix is row equivalent to itself:*

(b) It is symmetric. *This means that if A is equivalent to b then B is also equivalent to A:*

$$\forall A, B \quad A \sim B \implies B \sim A.$$

(c) It is transitive. *This means that if A is equivalent to B and B is equivalent to C then A is also equivalent to C:*

$$\forall A, B, C \quad A \sim B \text{ and } B \sim C \implies A \sim C.$$

PROOF. Reflexivity holds because we can get A by applying zero elementary row operations to A.

Symmetry holds because all elementary row operations are reversible.

Transitivity holds because if we can go from A from B and from B to C then we can clearly go from A to C: start from A and perform the row operators needed to go to B but don't stop, perform the operations needed to go from B to C. □

We already have seen the procedure for getting the reduced echelon form of a matrix in our example. Let's prove that it always work.

THEOREM 1.2.2 (**Kill below first, then kill above**).

(a) Every matrix is row equivalent to a matrix in echelon form.

(b) Every matrix in row echelon form is row equivalent to a matrix in reduced row echelon form. Therefore, every matrix is equivalent to a matrix in reduced echelon form.

PROOF. We will prove that every matrix has an echelon form and then we will show that any echelon matrix is row equivalent to a reduced echelon matrix.

(a) Starting with a_{11} we scan the first row for non-zero entries. If there isn't any then we proceed to the second row, and scan it starting with its leftmost entry. We continue until we either find a row that has a non-zero entry or we have scanned the whole matrix without succeeding. In the later case, all the rows of our matrix are zero rows and so the matrix is already in reduced echelon form.

If we are successful then the entry we find, say a_{ij}, is the leading entry of its row. We then scan all the entries below and to the left, that is all the entries $a_{k\ell}$ with $k < i$ and $\ell > j$, searching for non-zero entries. If we find such a non-zero $a_{k\ell}$ we restrict our search to the entries below and to the left of $a_{k\ell}$. Since every time we find such an $a_{k\ell}$ we move below end to the left, we keep decreasing the number of entries we are searching. Since there are finitely many entries in our matrix, this cannot go on forever, eventually we'll find a non-zero entry only zero columns to the left of it. Call that entry the *pivot* and denote it by p. Since $p \neq 0$ we can use row operations to kill all the entries bellow it in its column. Since p is the topmost and leftmost non-zero entry all the other entries in its column and all the entries of the column left of p are now zero. Make the row of p the first row using row operations.

We repeat the process restricting attention to the entries below and to the right of the first row. This process eventually will terminate because every time we find a new pivot we decrease the size of the matrix we concentrate on.

The matrix we get at the end of this procedure is in echelon form. Indeed, there cannot be a zero row above a non-zero row because our procedure picks all non-zero rows and moves them to the top just below the first non-zero row. All the entries below a leading entry p are zero because we have either killed them when we first found p, or they were 0 already. Finally, p is to the left of the leading entries of the rows above it, because otherwise it would have been killed.

(b) Let A be an echelon matrix. We start by dividing each non-zero row by its leading entry, to obtain an echelon matrix with all leading entries 1. Because all the entries to the left of the leading 1s are zero, we can kill all entries above the rightmost leading 1 (that is the leading 1 of the last non-zero row) without changing anything in the columns to the left of it, and in particular without changing the leading entries of the rows above the last non-zero row.

We then restrict attention to the entries above and to the left, and keep going. Again at every step we reduce the size of the matrix we are concentrating on, and therefore the procedure will terminate. The final matrix is obviously in reduced echelon form.

□

Gauss and Gauss-Jordan Elimination

When we solve a system, using either Gauss, or Gauss-Jordan, Elimination we modify the algorithm described above in two ways.

(a) If we scan a row and find no non-zero entries, we just discard that row.

(b) If the leading entry is in the last column we stop the procedure and declare that the solution set is $\varnothing$.

REMARK 4. I've made some choices in the description of the procedure above because I wanted to present it as an *algorithm*, a procedure that can be performed without any thought. Other choices are possible.

For example, dividing each of the rows of an echelon matrix by the leading entry could be done at any point of the procedure. If the algorithm is to be performed by an infallible entity it seems efficient to divide at the beginning of the procedure.

However doing so may introduce unwieldy fractions, that could cause more errors when the algorithm is executed by not-so-infallible beings. In such cases it may actually be more efficient to not divide until the end so as to minimize the probability of error.

In general, just because a procedure can be executed without any thought, it doesn't mean that we *have* to do it without thinking. We are thinking beings after all[6]. When we try to solve a problem we can use any method that seems suitable at the moment.

[6] Or at least we think so

EXAMPLE 11. Consider the matrix

$$A = \begin{pmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 6 & -16 & 20 & -14 & 14 & 24 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{pmatrix}$$

The pivot is $a_{21} = 3$. We use it to kill the first entries of the two rows below the row that contains the pivot (notice that the entries above the pivot, a_{11}, is already 0).

$$A \sim \begin{pmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 2 & 4 & -4 & -2 & 6 \\ 0 & 2 & 4 & -4 & -2 & 6 \end{pmatrix}$$

We then interchange the first and second row:

$$A \sim \begin{pmatrix} 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \\ 0 & -2 & 4 & -4 & -2 & 6 \\ 0 & -2 & 4 & -4 & -2 & 6 \end{pmatrix}$$

Now we concentrate on the submatrix (a_{ij}) with $i, j \geq 2$. The new pivot is 3 and we use it to kill the entries below it, that happen to be identical. This is done by adding $-2/3$ times the second row, to the third and fourth rows.

$$A \sim \begin{pmatrix} 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \\ 0 & 0 & 0 & 0 & 2/3 & 8/3 \\ 0 & 0 & 0 & 0 & 2/3 & 8/3 \end{pmatrix}$$

Next we get

$$A \sim \begin{pmatrix} 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

This an echelon matrix. To get the row equivalent reduced echelon matrix we start killing upwards.

$$A \sim \begin{pmatrix} 3 & -7 & 8 & -5 & 0 & -23 \\ 0 & 3 & -6 & 6 & 0 & -21 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$A \sim \begin{pmatrix} 3 & 0 & -6 & 9 & 0 & -72 \\ 0 & 3 & -6 & 6 & 0 & -21 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Finally, we divide first and second row by 3 and we get the reduced echelon form:

$$A \sim \begin{pmatrix} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

There are two kind of columns in a reduced echelon form, those that contain a leading entry and those that don't. Columns of the first kind are called *basic* and those of the second type are called *free*. When we solve systems that have coefficient matrix A, the free columns correspond to free variables.

So any system that has coefficient matrix A will, as long as it is consistent of course, have a solution set with three free parameters, i.e. the solution set will be 3-dimensional.

But wait a minute, what do I mean by "the solution set is three-dimensional"? Using this particular set of row operations we got a reduced echelon matrix with three free columns and this indeed will give a parametrization with three parameters. But maybe if we use an other sequence of row operations we will get a parametrization with two, or four, parameters.

That's a valid objection but it turns out that this can never happen. In fact every matrix is row equivalent to a *unique* matrix in reduced row echelon form. Therefore, the "dimension" of the solution set is well defined. We will prove that in the next section where we turn our attention to the special case of *homogeneous* systems, that is systems where all constants $c_1, c_2, \dots, c_n = 0$.

1.2.1. Homogeneous systems. Consider then the general $m \times n$ homogeneous system

$$\left\{\begin{array}{lcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & 0 \\ \quad \vdots \qquad\qquad \vdots & & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & 0 \end{array}\right. \tag{1.12}$$

Notice that $x_1 = x_2 = \cdots = x_n = 0$ is a solution of (1.12). Therefore homogeneous systems are always consistent, the interesting question then is whether there are other solution besides that obvious one.

DEFINITION 4 (**Trivial solution of a homogeneous system.**). The solution

$$x_1 = 0, \dots, x_n = 0$$

is called the *trivial solution*[7]. A solution with at least one of the variables assigned a non-zero value is called a *non-trivial* solution.

REMARK 5. For a homogeneous system the last column of the augmented matrix is redundant, it will always be the zero-column. So for homogeneous systems we work with the coefficient matrix, not the augmented matrix.

EXAMPLE 12. Consider the homogeneous system:

$$\left\{\begin{array}{rrrrl} x_1 + & x_2 + & x_3 - & 2x_4 & = 0 \\ 2x_1 & & - 2x_3 & & = 0 \\ & x_2 + & x_3 + & 4x_4 & = 0 \end{array}\right. .$$

To solve the system we bring its matrix to reduced echelon form. We first get an echelon form:

[7] The term *zero solution* is also occasionally used.

$$\begin{pmatrix}1&1&2&-2\\2&0&-2&0\\0&1&1&4\end{pmatrix}\sim\begin{pmatrix}1&1&2&-2\\0&-2&-6&4\\0&1&1&4\end{pmatrix}\sim\begin{pmatrix}1&1&2&-2\\0&1&1&4\\0&-2&-6&4\end{pmatrix}\sim\begin{pmatrix}1&1&2&-2\\0&1&1&4\\0&0&-4&12\end{pmatrix}\sim\begin{pmatrix}1&1&2&-2\\0&1&1&4\\0&0&1&3\end{pmatrix}.$$

And the reduced echelon form:

$$\begin{pmatrix}1&1&2&-2\\0&1&1&4\\0&0&1&3\end{pmatrix}\sim\begin{pmatrix}1&1&0&-8\\0&1&0&1\\0&0&1&3\end{pmatrix}\sim\begin{pmatrix}1&0&0&-9\\0&1&0&1\\0&0&1&3\end{pmatrix}.$$

We have one free column, and so the corresponding variable x_4 is free. So we have a one parameter solution set:

$$\begin{cases} x_1 &= 9t\\ x_2 &= -t\\ x_3 &= -3t\\ x_4 &= t\end{cases}\qquad t\in\mathbb{R}.$$

And using "column vectors":

$$\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=t\begin{pmatrix}9\\-1\\-3\\-1\end{pmatrix}.$$

We have two notions of equivalence for $m\times n$ systems:

- **Semantic Equivalence:** Two linear systems are considered equivalent if they have the same solution sets[8].
- **Syntactic Equivalence:** Two systems are considered equivalent if their (augmented) matrices are row equivalent[9].

As is usual the case, syntactic equivalence implies semantic equivalence, and the proof is rather easy. The converse is also true, that is, if two systems have the same solution set then their augmented matrices are row equivalent.

We will first prove this implication for homogeneous systems.

Let's start with the rather trivial case of a homogeneous system with one variable. The matrix of such a system is an $m\times 1$ matrix, i.e. a *column vector*. An echelon form of such a matrix is either the zero column of has all rows after the first 0 and the first non-zero. All matrices that have non-zero first row are row equivalent to the column vector that has first row 1 and all other rows 0.

Now the only possible solution sets of a homogeneous system with one variable are $\mathbb{R}$ and $\{0\}$. This follows because, as we observed in the first section, if x is a solution of a homogeneous equation so is then λx for all numbers λ. If the solution set is $\{0\}$ then at least one coefficient is non-zero and therefore the echelon form will be a column with first row non-zero and all such column vectors are row equivalent. If the solution set is $\mathbb{R}$ then all coefficients are 0 and so the column vector is the zero column.

Now, using induction, we can prove the following theorem.

[8] The term *semantic* is used for concepts related to *meaning*. Two systems with the same solutions have the same meaning in the sense that they describe the same set.

[9] The term *syntactic* is used for concepts related to *syntax*, that is the formal properties of a language, in contrast with the meaning. Row equivalence relates to the form of the system, we defined it without any reference to the solution sets of the system.

THEOREM 1.2.3. *Two reduced echelon $m \times n$ matrices whose homogeneous systems have the same solution set are equal.*

PROOF. We have seen that this is the case for systems with one variable. Assuming that the theorem is true for systems with n variables we will prove that it is also true for systems with $n+1$ variables.

Let then A and B be two $m \times (n+1)$ reduced echelon matrices with the same solution set S, and let A_0 and B_0 be the matrices obtained from A and B, respectively, by removing the last column. Consider the subset S_0 of those solutions that have the last coordinate 0, that is

$$S_0 = \{(x_1, \dots, x_n, x_{n+1}) \in S : x_{n+1} = 0\}\,.$$

Then S_0 is the solution set of A_0 and B_0, and by the inductive step it follows that $A_0 = B_0$. Therefore A and B can differ only on the last column.

The last columns have also to be the same though. To see this let k be the first row that the last columns of A and B differ, and let $a_k \neq b_k$ be the corresponding entries. Consider now the system $A - B$ obtained by subtracting the corresponding equations of A and B. This is a homogeneous system with only the last column non-zero and all elements of S are also solutions of $A - B$. For any such solution the k-th equation of $A - B$ is $(a_k - b_k)x_{n+1} = 0$. By our choice of k this means that $x_{n+1} = 0$. Therefore $S = S_0$, and so the last columns of A and B are both the zero column otherwise there would be solutions of A (respectively B) that are not solutions of A_0 (respectively B_0).

So if the last columns of A and B differ, they are both the zero-column, a contradiction. Therefore the last columns of A and B are the same. □

Since row equivalent systems have the same solution set, we have the following immediate corollaries of Theorem 1.2.3.

COROLLARY 1. *We have:*

(a) *Two reduced echelon matrices are row equivalent if and only if they are equal.*
(b) *The reduced echelon form of any matrix is unique.*
(c) *Two homogeneous systems with the same solution set are row equivalent.*

Let's now consider the question of uniqueness. When does a homogeneous system have a unique solution? The unique solution will be of course the trivial one. Let's consider systems with 3 variables for example. What homogeneous systems with three variables, say x, y, z, admit only the trivial solution $x = y = z = 0$?

Let A be the reduced echelon form of the matrix of the system. If A has free columns, then the system has non-trivial solutions: for example we can just give a non-zero value to one of the free parameters, and set the remaining free variables (if any) to zero. Therefore in order to have only the trivial solution all the columns have of A need to be basic, i.e., all the columns have to contain a leading 1. Since the leading 1s appear in different rows A needs to have at least three rows, i.e. the system needs to have at least three equations. This means that the first three rows of the system have to be[10]

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and the remaining rows (if any) have to be zero rows.

[10] This 3×3 matrix is very special, it will play an important role in the following lectures.

More generally, the number of basic variables, is always equal to the number of non-zero rows of A. In Example 12 we have three non-zero rows and the solution has three basic variables. You should go back through all the examples we have seen so far and verify that this is always the case.

The columns that are not basic are free and so we have the following theorem, a first version of the *Rank Theorem*.

THEOREM 1.2.4 (**The Rank Theorem**). *The number of non-zero rows plus the number of free columns in the reduced echelon form of A equals the numbers of variables of the system.*

1.2.2. Vector subspaces. What kind of subsets of $\mathbb{R}^n$ arise as solutions of homogeneous linear systems? Well, vector subspaces of course! That means that the sum of two solutions is again a solution, and a scalar multiple of a solution is again a solution. Vector subspaces of $\mathbb{R}^n$ are examples of vector spaces, one of the main objects of study of Linear Algebra.

DEFINITION 5 (**Column Vectors, vector addition, scalar multiplication**). An n-dimensional *column vector* is an $n \times 1$ matrix. We identify the n-tuple $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ with the column vector with entries $x_1, \dots, x_n$, that is we set

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

If $\mathbf{x}, \mathbf{y}$ are two column vectors and λ is a scalar (i.e. a real number) then we define the sum $\mathbf{x} + \mathbf{y}$ and the product $\lambda\, \mathbf{x}$ *component-wise*: if

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

then

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}, \text{ and } \lambda\, \mathbf{x} = \begin{pmatrix} \lambda\, x_1 \\ \lambda\, x_2 \\ \vdots \\ \lambda\, x_n \end{pmatrix}.$$

We also define $\mathbf{x} - \mathbf{y} = \mathbf{x} + (-1)\, \mathbf{y}$, so that

$$\mathbf{x} - \mathbf{y} = \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{pmatrix}.$$

There are several equivalent ways to define what a vector subspace is. The one we chose below is convenient for the purposes of this section. For the rest of this section, *vector* means *column vector*.

DEFINITION 6 (**Vector subspace**). A subset $V \subseteq \mathbb{R}^n$ is called a *vector subspace* if the following three conditions hold:

(a) V contains the zero vector, that is $\mathbf{0} \in V$.

(b) V is closed under vector addition, that is

$$\mathbf{x}, \mathbf{y} \in V \implies \mathbf{x} + \mathbf{y} \in V.$$

(c) V is closed under scalar multiplication, that is

$$\lambda \in \mathbb{R}, \mathbf{x} \in V \implies \lambda \mathbf{x} \in V.$$

On route to proving that the solution set of a homogeneous is a vector subspace we prove the following important result.

THEOREM 1.2.5 (**Matrix multiplication is linear**). *Let A be an $m \times n$ matrix, $\mathbf{x}, \mathbf{y}$ two n-vectors, and λ a real number. Then*

(a) $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$.
(b) $A(\lambda \mathbf{x}) = \lambda (A\mathbf{x})$.

PROOF. The k-th entry of $A(\mathbf{x} + \mathbf{y})$ is

$$\begin{aligned} a_{k1}(x_1 + y_1) + \cdots + a_{kn}(x_n + y_n) &= a_{k1}x_1 + a_{k1}y_1 + \cdots + a_{kn}x_n + a_{kn}y_n \\ &= (a_{k1}x_1 + \cdots + a_{kn}x_n) + (a_{k1}y_1 + \cdots + a_{kn}y_n). \end{aligned}$$

The sum in the first parenthesis is the k-th row of $A\mathbf{x}$ and the sum in the second parenthesis is the k-th row of $A\mathbf{y}$. Since this is true for all k the first item has been proved.

Similarly, the k-th row of $A(\lambda \mathbf{x})$ is

$$\begin{aligned} a_{k1}(\lambda x_1) + \cdots + a_{kn}(\lambda x_n) &= \lambda (a_{k1}x_1) + \cdots + \lambda (a_{kn}x_n) \\ &= \lambda (a_{k1}x_1 + \cdots + a_{kn}x_n). \end{aligned}$$

Now the last expression is the k-th row of $\lambda (A\mathbf{x})$ and so the second item has also been proven. □

Now let $\mathbf{x}, \mathbf{y}$ be two solutions a homogeneous system with matrix A. Then $A\mathbf{x} = A\mathbf{y} = \mathbf{0}$. Then,

$$A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0}.$$

Thus, $\mathbf{x} + \mathbf{y}$ is also a solution.

The proof that any scalar multiple of $\mathbf{x}$ is also a solution is entirely similar and we leave as an exercise[11].

We have then, as promised, the following theorem.

THEOREM 1.2.6. *The solution set of a linear homogeneous system with n variables is a vector subspace of $\mathbb{R}^n$.*

REMARK 6. We will see later in the course that every vector subspace of $\mathbb{R}$ is the solution set of some homogeneous linear system.

1.2.3. Solution sets of non-homogeneous systems. If we think of the solution set of a homogeneous systems as a space of vectors, then we should think of the solution set of a non-homogeneous system as a space of points. This is more than an analogy, the solution set of a non-homogeneous system is an *affine subspace* of $\mathbb{R}^n$. We are not going to define what that means precisely, we give some examples instead. A one dimensional affine subspace is the set of points in a line, a two dimensional affine subspace is the set of points in a plane, and so on.

Two points P, Q in $\mathbb{R}^n$ determine a vector $\mathbf{v} = \overrightarrow{PQ}$, that we can think geometrically as the *directed segment* from P to Q. Of course the same vector is defined by many different pairs of points, in fact given any point P' there is a unique point Q' such that $\mathbf{v} = \overrightarrow{P'Q'}$. See Figure 3 for examples,

[11] Do this.

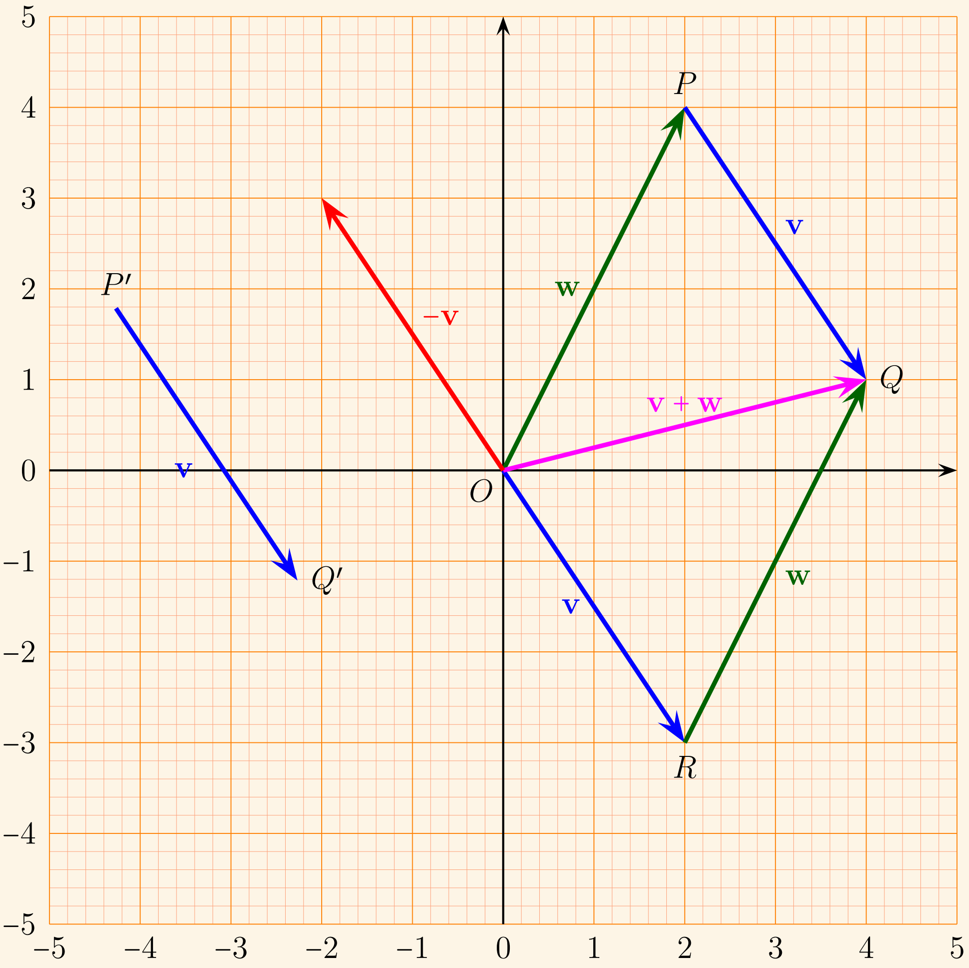


FIGURE 3. Points and vectors in $\mathbb{R}^2$.

If the coordinates of P are (p_1, p_2) and those of Q are (q_1, q_2) then the components of the vector $\mathbf{v}$ are $(q_1 - p_1, q_2 - p_2)$, in particular if we chose the starting point of $\mathbf{v}$ to be the origin $O(0,0)$ then the coordinates of the endpoint of $\mathbf{v}$ are exactly the components of $\mathbf{v}$.

We could write then $Q - P = \mathbf{v}$ and $P + \mathbf{v} = Q$, and say that "the difference of two points is a vector and the sum of a point and a vector is an other point".

Returning to the solution sets of non-homogeneous systems (refer also to Figure 2 and recall the surrounding discussion) we have the following theorem.

THEOREM 1.2.7 (**Solution sets of non-homogeneous systems**). *Let A be any matrix, S the solution set of a non-homogeneous system $A\mathbf{x} = \mathbf{c}$ and V the solution set of the homogeneous system $A\mathbf{x} = \mathbf{0}$. Then*

- *The difference of two solutions of the non-homogeneous system is a solution of the homogeneous system. That is*

$$\mathbf{a}, \mathbf{b} \in S \implies \mathbf{b} - \mathbf{a} \in V.$$

- *The sum of a solution of the non-homogeneous system and a solution of the homogeneous system is again a solution of the non-homogeneous system.*
- *For any solution $\mathbf{a}_0$ of the non-homogeneous system we can express any other solution of the homogeneous system as the sum of $\mathbf{a}_0$ and a unique solution of the homogeneous system. That is*

$$S = \{\mathbf{a}_0 + \mathbf{v} : \mathbf{v} \in V\}.$$

REMARK 7. The third item is sometimes expressed as "The general solution of a non-homogeneous system is the sum of the general solution of the homogeneous system and a particular solution (of the non-homogeneous system)".

SKETCH. [12] The first item follows from Theorem 1.2.5. The second is just a reformulation of the first. To prove the third item use the equation $\mathbf{a} = \mathbf{a}_0 + (\mathbf{a} - \mathbf{a}_0)$. □

EXAMPLE 13. Consider the system

$$\begin{cases} x + 2y - 3z + 2s - 4t &= 2 \\ 2x + 4y - 5z + s - 6t &= 1 \\ 5x + 10y - 13z + 4s - 16t &= 4 \end{cases}.$$

We work with the coefficient matrix. The reduced echelon form is:

$$A = \begin{pmatrix} 1 & 2 & -3 & 2 & -4 \\ 2 & 4 & -5 & 1 & -6 \\ 5 & 10 & -13 & 4 & -16 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & -3 & 2 & -4 \\ 0 & 0 & 1 & -3 & 2 \\ 0 & 0 & 2 & -6 & 4 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & -3 & 2 & -4 \\ 0 & 0 & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 0 & 11 & -10 \\ 0 & 0 & 1 & -3 & 2 \end{pmatrix}$$

So we have two basic variables x, z and three free variables y, s, t. This means that the solution of the homogeneous system, in vector form is

$$\begin{pmatrix} x \\ y \\ z \\ s \\ t \end{pmatrix} = a \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} 11 \\ 0 \\ -3 \\ 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} -10 \\ 0 \\ 2 \\ 0 \\ 1 \end{pmatrix}.$$

To solve the original non-homogeneous system then, we need to find only one particular solution. This is rather easy to do just by substituting values. For example, for $x = y = z = z = 0$ we find $t = 1$. So the solution of the non homogeneous system is

$$\begin{pmatrix} x \\ y \\ z \\ s \\ t \end{pmatrix} = a \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} 11 \\ 0 \\ -3 \\ 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} -10 \\ 0 \\ 2 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

We can interpret the solutions geometrically as follows: the solution set V of the homogeneous system is a 3-dimensional vector subspace of the standard 5-dimensional real vector space $\mathbb{R}^5$. A *basis* of V consists of $\mathbf{v} = 2\,\mathbf{e}_1 + \mathbf{e}_2$, $\mathbf{u} = 11\,\mathbf{e}_1 - 3\,\mathbf{e}_3 + \mathbf{e}_4$, and $\mathbf{w} = -10\,\mathbf{e}_1 + 2\,\mathbf{e}_3 + \mathbf{e}_5$. The solution S of the non-homogeneous system is the translation of V by the vector $\mathbf{e}_4$[13].

The following theorem summarizes our results.

THEOREM 1.2.8 (General solution of linear systems). *We have:*

- *A linear system is consistent if and only if the echelon form of its augmented matrix contains no rows of the form*

$$\begin{pmatrix} 0 & 0 & \dots & 0 & c \end{pmatrix}$$

with $c \neq 0$.

[12] Fill the details.

[13] By the end of the class all of the above will be making sense.

- *The solution set of a consisted system has as many parameters as the number of free columns in its reduced echelon form. In particular a consisted system has a unique solution if and only the reduced echelon form of its matrix*[14] *has ones along the diagonal and zeros everywhere else. For example a, consistent 4×4 system has a unique solution if and only if the reduced echelon form of its matrix is*
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$
- *Two consistent $m \times n$ systems are equivalent (i.e. have the same solution set) if and only if their augmented matrices are row equivalent.*
- *Two consistent $m \times n$ systems are equivalent (i.e. have the same solution set) if and only if their augmented matrices have the same reduced row echelon form.*
- *If the homogeneous system $Ax = \mathbf{0}$ has only the trivial solution then if the system $A\mathbf{x} = \mathbf{c}$ is consistent it has a unique solution.*

PROOF. The proof is left as an exercise. All the ingredients are already present in these notes. □

1.3. The 2×2 case

Let's analyze the case of a 2×2 linear system. Consider the system

$$\begin{cases} a_1 x + b_1 y & = c_1 \\ a_2 x + b_2 y & = c_2 \end{cases}$$

The augmented matrix is

$$\left(\begin{array}{cc|c} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{array}\right).$$

The case where all the coefficients are zero is rather trivial: in that case if both constants are also zero the solution set is $\mathbb{R}^2$, if at least one constant is non-zero the solution set is $\varnothing$.

Let's assume then that one of the coefficients is non-zero. Without loss of generality we can assume that $a_1 \neq 0$. For, if $a_1 = 0$ and $a_2 \neq 0$ then we can interchange the equations and get an equivalent system with the coefficient of x in the first equation non-zero. If both a_1 and a_2 are zero then we can interchange the variables, get a system of two equations where at least one of the coefficients of x is non-zero, solve that system, and then interchange the variables, again.

Since we assumed $a_1 \neq 0$ we can multiply the first equation with $-a_2/a_1$ and add it to the second:

$$\left(\begin{array}{cc|c} a_1 & b_1 & c_1 \\ 0 & b_2 - \dfrac{a_2 b_1}{a_1} & c_2 - \dfrac{a_2 c_1}{a_1} \end{array}\right) = \left(\begin{array}{cc|c} a_1 & b_1 & c_1 \\ 0 & \dfrac{a_1 b_2 - a_2 b_1}{a_1} & \dfrac{a_1 c_2 - a_2 c_1}{a_1} \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & \dfrac{b_1}{a_1} & \dfrac{c_1}{a_1} \\ 0 & a_1 b_2 - a_2 b_1 & a_1 c_2 - a_2 c_1 \end{array}\right).$$

We now look at the second row. Set $D = a_1 b_2 - a_2 b_1$[15], and consider two cases: whether D is zero or not.

[14] The matrix of coefficients **not** its augmented matrix.

[15] Later in the class we will see that this is the *determinant* of the coefficient matrix.

Non-zero Determinant. If $D \neq 0$ then we can divide the second row by D to get

$$\left(\begin{array}{cc|c} 1 & \frac{b_1}{a_1} & \frac{c_1}{a_1} \\ 0 & 1 & \frac{a_1c_2-a_2c_1}{D} \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & \frac{c_1b_2-c_2b_1}{D} \\ 0 & 1 & \frac{a_1c_2-a_2c_1}{D} \end{array}\right).$$

The expression of the third entry of the first column is the result of simplifying the following

$$\frac{c_1}{a_1} - \frac{b_1}{a_1} \cdot \frac{a_1c_2-a_2c_1}{D} = \frac{c_1\left(a_1b_2-a_2b_1\right) - b_1\left(a_1c_2-a_2c_1\right)}{a_1D}.$$

If we set $D_x = c_1b_2 - c_2b_1$ and $D_y = a_1c_2 - a_2c_1$ we have formulas that give the solution of a linear 2×2 system. These formulas are a special case of *Crammer's rule*, that we'll prove later.

2×2 Crammer's rule

When $a_1b_2 - a_2b_1 \neq 0$, the system

$$\begin{cases} a_1x + b_1y &= c_1 \\ a_2x + b_2y &= c_2 \end{cases}$$

has a unique solution given by

$$x = \frac{D_x}{D}, \qquad y = \frac{D_y}{D},$$

where $D = a_1b_2 - a_2b_1$, $D_x = c_1b_2 - c_2b_1$ and $D_y = a_1c_2 - a_2c_1$.

Zero Determinant. If $D = 0$ we have two cases: if $D_y \neq 0$ the system is inconsistent. If $D_y = 0$ then the system reduces in a single equation with, as we saw at the beginning of the previous section, a one parameter solution set.

1.3.1. Geometric interpretation. The condition $D = 0$ (or $D \neq 0$) has a nice geometric interpretation in terms of the graphs of the equations that make up our system. We only consider the nontrivial case where each equation has at least one non-zero coefficient, and therefore its graph is a line.

THEOREM 1.3.1. *The lines with equations*

$$a_1x + b_1y = c_1, \qquad a_2x + b_2y = c_2$$

are parallel if and only if $D = 0$.

PROOF. The condition $D = 0$ is equivalent to

$$a_1b_2 = a_2b_1. \tag{1.13}$$

- **Case I:** $a_1 = 0$. Then the first line is horizontal and the two lines are parallel if and only if $a_2 = 0$. On the other hand, since b_1 has to be non-zero Equation (1.3.1) also holds if and only if $a_2 = 0$.
- **Case II:** $a_1 \neq 0$. We have two cases:
  - **Case IIa:** $b_1 = 0$. Then the first line is vertical and the RHS of Equation (1.3.1) is 0. Since $a_1 \neq 0$ Equation (1.3.1) holds if and only if $b_2 = 0$, i.e. if and only if the second line is also vertical.

- **Case IIb:** $b_1 \neq 0$. Then if $a_2 = 0$, since our equations are non-trivial, $b_2 \neq 0$ and so Equation (1.3.1) cannot hold. The lines are not parallel either since the second line is horizontal and the first isn't.
  Finally if $a_2 \neq 0$ then Equation (1.3.1) is equivalent to
  $$\frac{b2}{a_2} = \frac{b_1}{a_1}$$
  which holds if and only if the lines are parallel.

□

This explains our results geometrically, if $D \neq 0$ the two lines are not parallel and therefore they intersect in a point. The coordinates of that point give us the unique solution of the system. If on the other hand $D = 0$, the two lines are parallel so they don't intersect and th system has no solution.

But what about the case $D = 0$ and $D_x = 0$, where we have a one-parameter solution set? Well, notice that, assuming $a_1 \neq 0$ we have

$$D = 0 \iff b_2 = \frac{a_2}{a_1} b_1$$

and

$$D_x = 0 \iff c_2 = \frac{a_2}{a_1} c_1.$$

So in that case we can write the second equation as

$$a_2 x + \frac{a_2}{a_1} b_1 y = \frac{a_2}{a_1} c_1.$$

which is the first equation multiplied by a_2/a_1. So the two equations are equivalent, and the system has as many solutions as the first equation.

Consider as an example the following three systems:

$$\begin{cases} x - y = 0 \\ x - y = -2 \end{cases}, \qquad \begin{cases} x - y = -2 \\ 2x - 2y = -4 \end{cases}, \qquad \begin{cases} x - y = 0 \\ 2x + 3y = 5 \end{cases}.$$

The first system is inconsistent, while in the second system the second equation equation is twice the first. The third system has the unique solution $x = y = 1$. The graphs of the equations $x - y = 0$, $x - y = -2$ and $2x + 3y = 5$ are shown in Figure 4. The lines of the equations in the first system don't intersect, both equations in the second system represent the same line, while the graphs of the equations in the third system intersect at the point with coordinates $(1, 1)$.

1.3.2. Another Geometric interpretation. Systems of linear equations arise also when we want to express a vector as a *linear combination* of a given set of *basic vectors*. In $\mathbb{R}^2$ we have the *standard basis* consisting of the vectors (written as columns)

Every other vector can be *uniquely* expressed as a sum of multiples of these two basic vectors. Indeed the components of the vector are the coefficients of such an expression since

$$\begin{pmatrix} a \\ b \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Let's give a few definitions. In the following *vector* means an element of some $\mathbb{R}^n$.

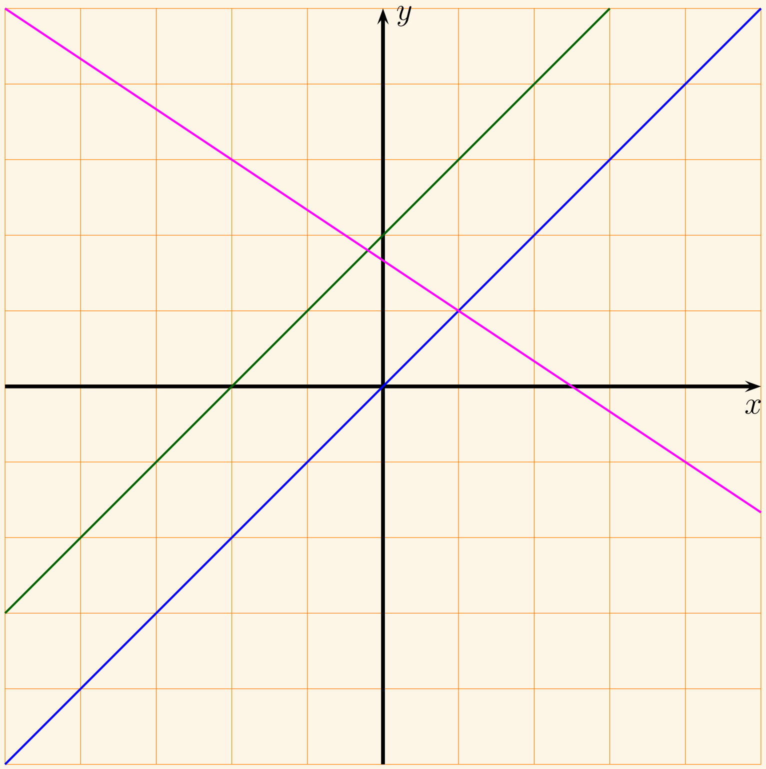


FIGURE 4. Parallel and intersecting lines.

Linear combinations, span, basis

A *linear combination* of m not necessarily distinct vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is a vector of the form

$$\lambda_1 \, \mathbf{v}_1 + \cdots + \lambda_n \, \mathbf{v}_m$$

where $\lambda_1, \ldots, \lambda_m$ are some scalars, called the *coefficients* of the combination.

The set of all linear combinations is called the *linear span* of $\mathbf{v}_1, \ldots, \mathbf{v}_m$ and is denoted by $\langle \mathbf{v}_1, \ldots, \mathbf{v}_m \rangle$,

$$\langle \mathbf{v}_1, \ldots, \mathbf{v}_m \rangle = \{ \lambda_1 \, \mathbf{v}_1 + \cdots + \lambda_m \, \mathbf{v}_m : \lambda_1, \ldots, \lambda_m \in \mathbb{R} \} \, .$$

If $V = \langle \mathbf{v}_1, \ldots, \mathbf{v}_m \rangle$ then we say that V is *spanned* by the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$. That means that every element of V is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m$, if that linear combination is unique we say that $\mathbf{v}_1, \ldots, \mathbf{v}_m$ form a *basis* of V.

The above discussion can then summarized by saying that $\mathbf{e}_1, \mathbf{e}_2$ form a basis of $\mathbb{R}^2$. The term *standard basis* suggests that there are other non-standard bases as well. And indeed there are tons of them!.

EXAMPLE 14 (**An other basis of $\mathbb{R}^2$**). The vectors $\mathbf{v} = 3 \, \mathbf{e}_1 - 2 \, \mathbf{e}_2$ and $\mathbf{w} = -2 \, \mathbf{e}_1 + 3 \, \mathbf{e}_2$ also form a basis of $\mathbb{R}^2$.

The phrase above claims two things.

(a) It claims that $\mathbf{v}, \mathbf{w}$ span $\mathbb{R}^2$, i.e. that any vector $\mathbf{c} \in \mathbb{R}^2$ is a linear combination of $\mathbf{v}, \mathbf{w}$. Unpacking this further the claim is that given $\mathbf{c} \in \mathbb{R}^2$ we can find $x, y \in \mathbb{R}$ so that

$$x\,\mathbf{v} + y\,\mathbf{w} = \mathbf{c}. \tag{1.14}$$

(b) Furthermore it claims that only one such pair of real numbers exist.

In other words, to say "$\mathbf{v}, \mathbf{w}$ is a basis of $\mathbb{R}^2$" is equivalent to saying "Equation (1.14) has a unique solution for all $\mathbf{c} \in \mathbb{R}^2$".

Let's then proceed and prove the claim. Let $\mathbf{c} = c_{1\,1} + c_2\, e_2$ be an arbitrary vector, then using column vector notation Equation (1.14) becomes

$$x \begin{pmatrix} 3 \\ -2 \end{pmatrix} + y \begin{pmatrix} -2 \\ 3 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

Performing the operations in LHS we get equivalently

$$\begin{pmatrix} 3x \\ -2x \end{pmatrix} + \begin{pmatrix} -2y \\ 3y \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \iff \begin{pmatrix} 3x - 2y \\ -2x + 3y \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

Two vectors are equal if and only if their corresponding components are equal, so the last equation is equivalent to the system

$$\begin{cases} 3x - 2y = c_1 \\ -2x + 3y = c_2 \end{cases}.$$

Using Crammer's rule, we get

$$x = \frac{3c_1 + 2c_2}{5}, \qquad y = \frac{2c_1 + 3c_2}{5}.$$

Thus, as claimed we have a unique solution, and $\mathbf{v}, \mathbf{w}$ form a basis of $\mathbb{R}^2$.

This example demonstrates the general procedure that we'll use to find whether a vector is in the linear span of a given list of vectors. That question reduces to solving a linear system.

Vector equations as systems

The vector equation

$$x_1\,\mathbf{v}_1 + \cdots + x_n\,\mathbf{v}_m = \mathbf{c}$$

is equivalent to the system

$$A\,\mathbf{x} = \mathbf{c}$$

where A is the matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_m$.

Consider now two arbitrary vectors $\mathbf{a} = a_1\,\mathbf{e}_1 + a_2\,\mathbf{e}_2$ and $\mathbf{b} = b_1\,\mathbf{e}_1 + b_2\,\mathbf{e}_2$. The question of whether $\mathbf{c} = c_{1\,1} + c_2\, e_2$ is in the linear span $\langle \mathbf{a}, \mathbf{b} \rangle$ reduces to whether the system

$$\begin{cases} a_1 x + b_1 y & = c_1 \\ a_2 x + b_2 y & = c_2 \end{cases}$$

has solutions, and we have a complete answer to that question.

(a) If the determinant $D = a_1 b_2 - a_2 b_1$ is non-zero then $\mathbf{a}, \mathbf{b}$ form a basis. Every vector can be written as a linear combination of $\mathbf{a}, \mathbf{b}$ in exactly one way.

(b) If the determinant is 0, whether $\mathbf{c}$ is in linear span of $\mathbf{a}, \mathbf{b}$ depends on the value of the determinants D_x and D_y.

In the previous section we interpreted the condition $D = 0$ in terms of points. Let's now interpret it in terms of vectors. Lets start with the case where one of the vectors is the zero vector $\mathbf{0}$.

One of the vectors is the zero vector. If $\mathbf{a} = \mathbf{0}$ then $D = 0$ and the answer depends on whether $\mathbf{b}$ is also zero or not.

Case I: Both vectors are zero. Then the system has solutions only if $\mathbf{c} = \mathbf{0}$. Any x, y are actually solutions.

$$\langle \mathbf{0} \rangle = \{\mathbf{0}\}\,.$$

Case II: If $\mathbf{b} \neq \mathbf{0}$ then we have solutions if and only if $\mathbf{c}$ is a scalar multiple of $\mathbf{b}$, in other words if and only if $\mathbf{c} \in \langle \mathbf{b} \rangle$. Since we can give arbitrary values to x the solution is not unique. So we have

$$\langle \mathbf{0}, \mathbf{b} \rangle = \langle \mathbf{b} \rangle\,.$$

Even though $\mathbf{0}, \mathbf{b}$ is not a basis of the linear span, $\mathbf{b}$ by itself constitute a basis.

Both vectors are non-zero. In this case each vector has at least one non-zero component. Let's assume that $a_1 \neq 0$. Then

$$D = 0 \iff b_2 = \frac{a_2}{a_1} b_1 \iff \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \frac{b1}{a_1} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}.$$

Therefore the condition $D = 0$ holds if and only if $\mathbf{b}$ is a multiple of $\mathbf{a}$. If that is the case then any linear combination of $\mathbf{a}$ and $\mathbf{b}$ can be written in terms of only $\mathbf{a}$ or only $\mathbf{b}$.

To see this assume that $\mathbf{b} = \lambda\,\mathbf{a}$ then

$$\lambda_1\,\mathbf{a} + \lambda_2\,\mathbf{b} = \lambda_1\,\mathbf{a} + \lambda_2\,(\lambda\,\mathbf{a}) = (\lambda_1 + \lambda_2\,\lambda)\,\mathbf{a}.$$

We have assumed that both $\mathbf{a}$ and $\mathbf{b}$ are non-zero, so $\lambda \neq 0$ and we can write $\mathbf{a} = \lambda^{-1}\mathbf{b}$ so the roles of $\mathbf{a}$ and $\mathbf{b}$ can be reversed, and we can write any linear combination in terms of $\mathbf{b}$ alone.

In summary, if $D = 0$ the two vectors are multiples of each other and we have

$$\langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{a} \rangle = \langle \mathbf{b} \rangle\,.$$

$\mathbf{a}, \mathbf{b}$ does not constitute a basis of $\langle \mathbf{a}, \mathbf{b} \rangle$[16]. However, by $\mathbf{a}$ (or) by itself forms a basis[17].

If $D \neq 0$ then the system has unique solution for all $\mathbf{c} \in \mathbb{R}^2$. In that case

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbb{R}^2$$

and $\mathbf{a}, \mathbf{b}$ form a basis.

The geometric reason that any two vectors of the plane form a basis, as long as they are not multiples of each other, is the same reason that our familiar Cartesian coordinates work.

[16] Why?

[17] Why?

Let $\mathbf{v} \in \mathbb{R}^2$ be an arbitrary vector. Take the starting point of $\mathbf{v}$ to be the origin $O(0,0)$ and let its endpoint be $P(x,y)$, where we are using the familiar Cartesian coordinate system (see the left side of Figure 5). Then if we draw a line from P, parallel to the y-axis, it will intersect the x-axis at a point P_x with coordinates $(x,0)$. Then

$$\overrightarrow{O\,P_x} = x\,\mathbf{e}_1.$$

Similarly, a line from P parallel to the x-axis intersects the y-axis at a point with coordinates $(0,y)$, and

$$\overrightarrow{O\,P_y} = y\,\mathbf{e}_2.$$

Now $O\,P_x\,P\,P_y$ is a rectangle and therefore we have

$$\overrightarrow{O\,P_y} = \overrightarrow{P_x\,P}.$$

Then we have,

$$\begin{aligned}\mathbf{v} &= \overrightarrow{O\,P}\\ &= \overrightarrow{O\,P_x} + \overrightarrow{P_x\,P}\\ &= \overrightarrow{O\,P_x} + \overrightarrow{O\,P_y}\\ &= x\,\mathbf{e}_1 + y\,\mathbf{e}_2.\end{aligned}$$

Thus every vector is a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$. Furtermore the coefficients x and y are unique, since P_x and P_y are uniquely determined by P.

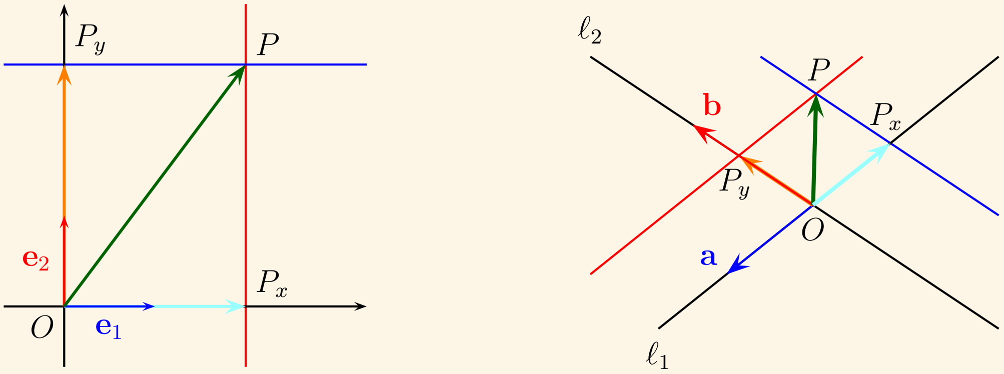


FIGURE 5. Why coordinates work.

The same idea works for any two vectors $\mathbf{a}$, , as long as they are not multiples of each other. Indedd let, $\mathbf{a}$, $\mathbf{b}$ be two vectors in $\mathbb{R}^2$. Chose an arbitrary point O in the plane, and let ℓ_1 be the line through O in the direction of $\mathbf{a}$, and ℓ_2 the line through O in the direction of $\mathbf{b}$. If ℓ_1 and ℓ_2 are not the same line, then for any point P we can find points P_x in ℓ_1 and $P_y \in \ell_2$ such that

$$\overrightarrow{O\,P} = \overrightarrow{O\,P_x} + \overrightarrow{O\,P_y}.$$

Indeed P_x (respectively P_y) is the intersection of a line that passes through P and is parallel to ℓ_2 (respectively ℓ_1), with ℓ_1 (respectively ℓ_2), as shown in the right side of Figure 5. Now, for some $x, y \in \mathbb{R}$ we have

$$\overrightarrow{O\,P_x} = x\,\mathbf{a}, \quad \overrightarrow{O\,P_y} = y\,\mathbf{b}$$

and since for any $\mathbf{v} \in \mathbb{R}^2$ we can find a P such that $\mathbf{v} = \overrightarrow{OP}$ we conclude that,

$$\forall \mathbf{v} \in \mathbb{R}^2,\ \exists x, y \in \mathbb{R},\ \mathbf{v} = x\,\mathbf{a} + y\,\mathbf{b}.$$

1.4. Exercises

(a) Solve each of the following systems:

(a)

$$\begin{cases} x + 2y + 3z = 0 \\ 3x + y + 2z = 0 \\ 2x + 3y + z = 0 \end{cases}.$$

(b)

$$\begin{cases} x - y + z = 0 \\ -x + 3y + z = 5 \\ 3x + y + 7z = 2 \end{cases}.$$

(c)

$$\begin{cases} x_1 + 3x_2 - 2x_3 + 2x_5 = 0 \\ 2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 = -1 \\ 5x_3 + 10x_4 + 15x_6 = 5 \\ 2x_1 + 6x_2 + 8x_4 + 4x_5 + 18x_6 = 6 \end{cases}.$$

(b) Find the real number k so that the following system is consistent

$$\begin{cases} x - 2y + 3z = 2 \\ x + y + z = k \\ 2x - y + 4z = k^2 \end{cases}.$$

(c) Find conditions on the real numbers a, b, c, if any, so that the system

$$\begin{cases} x & + & y & & & = 0 \\ & & y & + & z & = 0 \\ x & & & - & z & = 0 \\ ax & + & by & + & cz & = 0 \end{cases}$$

(a) is inconsistent.
(b) Has a unique solution.
(c) Has more than one solution.

(d) Consider the 2×2 matrix

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

where $a, b, c, d \in \mathbb{R}$.

(a) Prove that if $ad - bc \neq 0$ then the reduced row echelon form of A is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

(b) Prove that if $ad - bc \neq 0$ then the system

$$\begin{cases} ax + by & = k \\ cx + dy & = l \end{cases}$$

has a unique solution, for all real numbers k, l.

(e) Prove that there is a unique line passing through any two *distinct* points of the plane.

Hint 1. Work as in Example 4. Show that the system we obtain has non-trivial solutions and all the non-trivial equations differ by a multiplicative constant.

(f) Find the cubic polynomial

$$p(x) = a\,x^3 + b\,x^2 + c\,x + d$$

given that $p(1) = 0$, $p(2) = 3$, $p(-1) = -6$, and $p(-2) = -21$.

(g) Look at Examples 5 and 6, there is a geometric reason why in Example 6 the polynomial we got was not quadratic. The graph of a quadratic polynomial is a parabola so in these examples we were trying to find a parabola that passes through three distinct points. But the points in Example 6 are *colinear*[18] and so there is no parabola that passes through all three of them.

  (a) Prove that given any three *distinct* real numbers x_1, x_2, x_3 and any three real numbers y_1, y_2, y_3 we can always find a polynomial $p(x) = a\,x^2 + b\,x + c$ such that $p(x_1) = y_1$, $p(x_2) = y_2$, and $p(x_3) = y_3$.

  (b) The polynomial in part (a) is quadratic (i.e. $a \neq 0$) if and only if the points (x_1, y_1), (x_2, y_2), and (x_3, y_3) are not colinear.

The following is more of an invitation to think than an exercise. A puzzle if you will. See whether you can figure it out, but don't feel bad if you

(h) **What's going on with "free" and "basic" variables?** In a reduced echelon matrix the free variables are determined, they are those that correspond to the free columns. Since the reduced echelon form of a matrix is *unique* this means that which variables are free and which are basic are determined in advance for any system.

But how can this be true? Can't we just choose which variables to solve for? Haven't we done that already?

[18] This means that they lie in a line.

CHAPTER 2

Standard Real vector spaces

In this chapter we officially introduce the n-dimensional vector spaces $\mathbb{R}^n$, for all natural numbers n. We have already encounter them as the spaces where solutions of linear systems live: a solution of an $m \times n$ system is an n-tuple of real numbers, i.e. an element of $\mathbb{R}^n$. We call $\mathbb{R}^n$ the *standard n-dimensional real vector space* and its elements *standard n-dimensional real vector s*.

The theory we develop in this chapter will be abstracted into two directions later on. We will consider vector spaces that are not necessarily real, for example $\mathbb{C}^n$ is a *complex n-dimensional vector space*, and we will consider vector spaces whose elements are not standard vectors, for example we will encounter vector spaces whose elements are matrices, polynomials, functions, and so on.

However, as we will see, in the finite dimension case at least, every vector space "looks exactly like" a standard vector space. This means that the concepts we develop in this chapter apply to all (finite dimensional) vector spaces.

2.1. The standard real vector spaces and their subspaces

The *standard (real) n-dimensional vector space* is the set $\mathbb{R}^n$ endowed with the operations of vector addition and scalar multiplication that we will formally introduce below[1]. We call element of $\mathbb{R}^n$, *n-vectors* or simply vectors when n is understood or irrelevant. Thus an n-vector is an ordered tuples of real numbers $\mathbf{a} = (a_1, a_2, \ldots, a_n)$. We often identify n-vectors with $n \times 1$ matrices and call them *column* vectors, and sometimes we identify vectors with $1 \times n$ matrices and call them *row vectors*. So we have three notations for the same vector:

$$\mathbf{a} = (a_1, a_2, \ldots, a_n), \qquad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \qquad \mathbf{a} = \begin{pmatrix} a_1 & a_2 & \ldots & a_n \end{pmatrix}.$$

When $n = 1$ we identify $\mathbb{R}^1$ with $\mathbb{R}$ and write for example 3 instead of (3). The case $n = 0$ is also included, $\mathbb{R}^0$ has a single element, the empty tuple $()$ which we denote by $\mathbf{0}$, and call it the (0-dimensional) *zero vector*. Thus, $\mathbf{R} = \{\mathbf{0}\}$.

For $n \geq 1$ we call the n-tuple with all components 0 the (n-dimensional) *zero vector* and denote it also by $\mathbf{0}$. So

$$\mathbf{0} = (0, 0, \ldots, 0).$$

This abuse of notation doesn't cause confusion because the context makes it clear what $\mathbf{0}$ stands for if we write "Consider $\mathbf{a} \in \mathbb{R}^4$ with $\mathbf{a} \neq \mathbf{0}$" then we clearly mean $(0, 0, 0, 0)$, while in "Two non zero vectors of $\mathbf{R}^2$" we refer to $(0, 0)$.

For $n \geq 1$ we say that the n-vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$, where $\mathbf{e}_k$ has 1 at the k-th slot and 0 everywhere else, form the *standard basis* of $\mathbf{R}^n$. For example the standard basis of $\mathbb{R}^4$ consists of the four vectors

[1] We have already see these operations, but in this section we make it official.

$$\begin{aligned}\mathbf{e}_1 &= (1,0,0,0)\\ \mathbf{e}_2 &= (0,1,0,0)\\ \mathbf{e}_3 &= (0,0,1,0)\\ \mathbf{e}_4 &= (0,0,0,1).\end{aligned}$$

Again the use of the same symbol for different things doesn't usually cause confusion.

DEFINITION 7 (**Vector addition and scalar multiplication**). Let $\mathbf{a} = (a_1,\dots,a_n)$ and $\mathbf{b} = (b_1,\dots,b_n)$ be two n-vectors and λ a real number. We define

$$\lambda\,\mathbf{a} = (\lambda\, a_1,\dots,\lambda a_n)$$

and

$$\mathbf{a}+\mathbf{b} = (a_1+b_1,\dots,a_n+b_n).$$

The *opposite* of $\mathbf{a}$, denoted by $-\mathbf{a}$, is the vector

$$-\mathbf{a} = (-a_1,\dots,-a_n),$$

and we denote $\mathbf{a}+(-\mathbf{b})$ by $\mathbf{a}-\mathbf{b}$. So,

$$\mathbf{a}-\mathbf{b} = (a_1-b_1,\dots,a_n-b_n).$$

EXAMPLE 15 (**Two dimensional vectors**). Let's see some examples of two dimensional vectors. If $\mathbf{a} = (2,-1)$ and $\mathbf{b} = (3,2)$.

$$5\,\mathbf{a} = (5\cdot 2, 5\,(-1)) = (10,-5),$$

$$\mathbf{a}+\mathbf{b} = (2+3,-1+2) = (5,1),$$

$$\mathbf{a}-\mathbf{b} = (2-3,-1-2) = (-1,-3),$$

$$-2\,\mathbf{a}+7\,\mathbf{b} = (-2\cdot 2,-2\,(-1)) + (7\cdot 3, 7\cdot 2) = (-4,2)+(21,14) = (17,16).$$

Now, let $x,y\in\mathbb{R}$ and consider the linear combination

$$x\,\mathbf{e}_1 + y\,\mathbf{e}_2 = x\,(1,0) + y\,(0,1) = (x,0)+(0,y) = (x,y).$$

So any vector in $\mathbb{R}^2$ can be written as a linear combination of the vectors of the standard basis, and actually the components of the vector are the coefficients.

In general, if $\mathbf{a} = (a_1,\dots,a_n)$ then we have,

$$\mathbf{a} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n. \tag{2.1}$$

THEOREM 2.1.1 (**Vector Space Axioms**). *The operations of vector addition and scalar multiplication enjoy the following properties:*

(a) Vector addition is commutative. *This means that for any two vectors* $\mathbf{a}$, $\mathbf{b}$ *we have*

$$\mathbf{a}+\mathbf{b} = \mathbf{b}+\mathbf{a}.$$

(b) Vector addition is associative. *This means that for any three vectors* $\mathbf{a}$, $\mathbf{b}$, *and* $\mathbf{c}$ *we have*

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

(c) $\mathbf{0}$ *is* neutral *for addition. This means that for any vector* $\mathbf{a}$ *we have*

$$\mathbf{0}+\mathbf{a} = \mathbf{a}.$$

(d) *For every vector* $\mathbf{a}$ *we have*

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

(e) *The number* 1 *is neutral for scalar multiplication. This means that for every vector* $\mathbf{a}$ *we have*

$$1\,\mathbf{a} = \mathbf{a}.$$

(f) *Scalar multiplication distributes over vector addition. This means that if* λ *is a scalar and* $\mathbf{a}$, $\mathbf{b}$ *are vectors we have*

$$\lambda\,(\mathbf{a} + \mathbf{b}) = \lambda\,\mathbf{a} + \lambda\,\mathbf{b}.$$

(g) *Addition of scalars distributes over scalar multiplication. This means that*

$$(\lambda + \mu)\,\mathbf{a} = \lambda\,\mathbf{a} + \mu\,\mathbf{a}.$$

(h) *Multiplication of scalars and scalar multiplication are compatible in the following sense: if* λ, μ *are scalars and* $\mathbf{a}$ *is a vector, we have*

$$\lambda\,(\mu\,\mathbf{a}) = (\lambda\,\mu)\,\mathbf{a}.$$

The proofs of all of these properties are straightforward, they follow from the analogous properties of real numbers. For example for (6), we have

$$\begin{aligned}\lambda\,(\mathbf{a} + \mathbf{b}) &= \lambda\,((a_1, \dots, a_n) + (b_1, \dots, b_n))\\ &= \lambda\,(a_1 + b_1, \dots, a_n + b_n))\\ &= (\lambda\,(a_1 + b_1), \dots, \lambda\,(a_n + b_n))\\ &= (\lambda\,a_1 + \lambda\,b_1, \dots, \lambda\,a_n + \lambda\,b_n)\\ &= (\lambda\,a_1, \dots \lambda\,a_n) + (\lambda\,b_1, \dots, \lambda\,b_n)\\ &= \lambda\,(a_1, \dots a_n) + \lambda\,(b_1, \dots, b_n)\\ &= \lambda\,\mathbf{a} + \lambda\,\mathbf{b}.\end{aligned}$$

There are many other properties that we could have listed. The importance of these particular eight is that they are sufficient to prove any algebraic property of vectors that we'll ever need. If we knew nothing else about vectors except that there are two operations that satisfy these eight properties we still would be able to prove anything we need to develop our theory.

We list now some useful properties that follow from these "axioms".

THEOREM 2.1.2 (Some consequences of the axioms). *We have:*

- *For all vectors* $\mathbf{a}$, $\mathbf{b}$ *the equation*

$$\mathbf{a} + \mathbf{x} = \mathbf{b}$$

has a unique solution.

- *For any vector* $\mathbf{a}$

$$-1\,\mathbf{a} = -\mathbf{a}$$

- *For any scalar* λ *we have*

$$\lambda\,\mathbf{0} = \mathbf{0}.$$

- *For any vector* $\mathbf{a}$

$$0\,\mathbf{a} = \mathbf{0}.$$

- *For scalar* λ *and vector* $\mathbf{a}$

$$\lambda\,\mathbf{a} = \mathbf{0} \iff \lambda = 0 \text{ or } \mathbf{a} = \mathbf{0}.$$

All of these properties are straightforward to prove directly from the definitions of vector addition and scalar multiplication and we will be using them freely. We will see proofs from the axioms when we introduce abstract vector spaces.

Recall the definitions of *vector subspace* (Definition 6 in Section 1.2.2), *linear combination*, *linear span*, and *basis* (Section 1.3.2).

The following gives an alternative characterization of vector subspaces. It could be used as the definition instead. For brevity from now on we will simply say *subspace* instead of *vector subspace*.

THEOREM 2.1.3 (**Alternative definition of Vector subspace**). *A subset $V \subseteq \mathbf{R}^n$ is a subspace if and only if the following two properties hold:*

- *$V \neq \varnothing$.*
- *For all $\lambda, \mu \in \mathbb{R}$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$*

$$\mathbf{a}, \mathbf{b} \in V \implies \lambda\,\mathbf{a} + \mu\,\mathbf{b} \in V.$$

PROOF. A subspace V satisfies (1) since $\mathbf{0} \in V$.

Also, if $\lambda, \mu \in \mathbb{R}$ and $\mathbf{a}, \mathbf{b} \in V$ then by the third property listed in Definition 6 we have $\lambda\,\mathbf{a} \in V$ and $\mu\,\mathbf{b} \in V$ and therefore by the second property in Definition 6 we have $\lambda\,\mathbf{a} + \mu\,\mathbf{b} \in V$. Thus V satisfies (2) as well and the only if part has been proved.

Conversely, if V satisfies the two conditions listed in the theorem then it contains the zero vector. Indeed take any $\mathbf{a} \in V$[2], then by the second property we have

$$1\,\mathbf{a} + (-1)\,\mathbf{a} \in V \implies \mathbf{0} \in V.$$

Condition (2) of Definition 6 follows from the second property if we take $\lambda = \mu = 1$ and Condition (3) if we take $\lambda = 1$ and $\mu = 0$. Thus V is a subspace and the if part is also proved. □

By induction we can generalize the second property as follows.

PROPOSITION 1. *If V is a subspace then all linear combinations of elements of V are elements of V. That is,*

$$\lambda_1, \ldots, \lambda_m \in \mathbb{R}, \mathbf{v}_1, \ldots, \mathbf{v}_m \in V \implies \lambda_1\,\mathbf{v}_1 + \cdots + \lambda_m\,\mathbf{v}_m \in V.$$

Before proceeding let's observe that there are are two "trivial" subspaces. The whole $\mathbb{R}^n$ and the set $\{\mathbf{0}\}$ that contains only the zero vector, and every subspace is between those two subspaces, in the sense that

$$\{\mathbf{0}\} \subseteq V \subseteq \mathbb{R}^n.$$

Let's also prove the following important fact.

THEOREM 2.1.4 (**Intersection of subspaces is a subspace**). *If V and W are subspaces of $\mathbf{R}^n$ then their intersection $V \cap W$ is also a subspace of $\mathbb{R}^n$.*

PROOF. We will prove that $V \cap W$ has the two properties described in Theorem 2.1.3.

For the first, notice that the zero vector is in the intersection because it is in both V and W. The intersection therefore is not empty.

For the second, if $\mathbf{a}, \mathbf{b} \in V \cap W$ then $\mathbf{a}, \mathbf{b} \in V$ and therefore $\lambda, \mathbf{a} + \mu\,\mathbf{b} \in V$. But we also have $\mathbf{a}, \mathbf{b} \in W$ and therefore $\lambda, \mathbf{a} + \mu\,\mathbf{b} \in W$ as well. It follows that $\lambda\,\mathbf{a} + \mu\,\mathbf{b} \in V \cap W$. □

A linear combination of one vector $\mathbf{a}$ is just a multiple of that vector. By convention we set that a linear combination of zero n-vectors to be the zero vector of $\widetilde{R}^n$.

THEOREM 2.1.5 (**Linear Spans are subspaces**). *For any $S \subseteq \mathbb{R}^n$ the linear span $\langle S \rangle$ is a subspace of $\mathbb{R}^n$.*

[2] We can do this because V is not empty.

PROOF. **Sketch**[3] For the trivial case $S = \varnothing$ we have $\langle S \rangle = \{\mathbf{0}\}$ which is a subspace. For non-empty S the three conditions of Definition 6 are satisfied because:

(a) $\mathbf{0} = 0\,\mathbf{a}$ for any $\mathbf{a} \in S$.

(b) The sum of two sums of multiples of elements S is obviously also a sum of multiples of elements of S.

(c) We have

$$\lambda\,(\lambda_1\,\mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n) = (\lambda\,\lambda_1)\,\mathbf{v}_1 + \cdots + (\lambda\,\lambda_n)\mathbf{v}_n.$$

□

DEFINITION 8 (**Basis of a subspace**). We say that a set of vectors $B \subseteq \mathbf{R}^n$ is a *basis* of the subspace V if any $\mathbf{v} \in V$ can be expressed as a linear combination of vectors of B in a unique way.

We should clarify what we mean by *unique* in the definition above. For example we don't consider

$$2\,\mathbf{v}_1 + 3\,\mathbf{v}_1 - \mathbf{v}_2, \quad 5\,\mathbf{v}_1 - \mathbf{v}_2$$

different ways of expressing the the same vector as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. We also don't consider

$$-3\,\mathbf{v}_1 + 2\,\mathbf{v}_2 + 0\,\mathbf{v}_3, \quad -3\,\mathbf{v}_1 + 2\,\mathbf{v}_2 + 0\,\mathbf{v}_4$$

to be different.

Two linear combinations are considered different if after we rewrite them so that every vector appears only once (i.e. after we combine "like terms") then there is at least one vector that appears with different coefficients.

EXAMPLE 16. The fundamental example of a basis is the standard basis of $\mathbb{R}^n$. To see that it is indeed a basis notice that if $\mathbf{c} = (c_1, \ldots, c_n)$ then

$$\mathbf{c} = c_1\,\mathbf{e}_1 + \cdots + c_n\,\mathbf{e}_n$$

so the components of $\mathbf{c}$ are the coefficients of an expression of $\mathbf{c}$ as a linear combination of elements of $\{\mathrm{e}_1, \ldots, \mathrm{e}_n\}$. This is the only way to get $\mathbf{c}$ as a linear combination, because

$$\lambda_1\,\mathbf{e}_1 + \cdots + \lambda_n\,\mathbf{e}_n = (\lambda, \ldots, \lambda_n)$$

and therefore

$$\mathbf{c} = \lambda_1\,\mathbf{e}_1 + \cdots + \lambda_n\,\mathbf{e}_n \implies (c_1, \ldots, c_n) = (\lambda, \ldots, \lambda_n).$$

In general to prove that a set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ forms a basis of a subspace V we have to prove that the vector equation

$$x_1\,\mathbf{v}_1 + \cdots x_m\,\mathbf{v}_m = \mathbf{c}$$

has a unique solution for all $\mathbf{c} \in V$. As we have seen this vector equation is equivalent to the system

$$A\,\mathbf{x} = \mathbf{c}$$

where A is the matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_m$.

In the case of the standard basis we have the $n \times n$ matrix

$$\begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

[3] Fill the details.

and by Theorem 1.2.8 we conclude that the system has a unique solution for all $\mathbf{c}$.

EXAMPLE 17. The vectors

$$\begin{aligned}\mathbf{v}_1 &= (1,0,2,3),\\ \mathbf{v}_2 &= (-1,2,3,1),\\ \mathbf{v}_3 &= (1,4,-5,0),\\ \mathbf{v}_4 &= (0,1,-2,1).\end{aligned}$$

form a basis of $\mathbb{R}^4$. Indeed the matrix with columns these vectors is

$$\begin{pmatrix}1 & -1 & 1 & 0\\ 0 & 2 & 4 & 1\\ 2 & 3 & -5 & -2\\ 3 & 1 & 0 & 1\end{pmatrix}$$

We now obtain an echelon form. We first add -2 times the first row to the third, and -3 the first row to the fourth. Then we add 5 times the second row to -2 times the second, and 2 times the second row to the fourth. Then we add 3 times the fourth row to the third.

$$\begin{pmatrix}1 & -1 & 1 & 0\\ 0 & 2 & 4 & 1\\ 0 & 5 & -7 & -2\\ 0 & 4 & -3 & 1\end{pmatrix} \sim \begin{pmatrix}1 & -1 & 1 & 0\\ 0 & 2 & 4 & 1\\ 0 & 0 & 34 & 1\\ 0 & 0 & -11 & -1\end{pmatrix} \sim \begin{pmatrix}1 & -1 & 1 & 0\\ 0 & 2 & 4 & 1\\ 0 & 0 & 1 & -2\\ 0 & 0 & -11 & -1\end{pmatrix}$$

We finally add 11 times the third row to the fourth.

$$\begin{pmatrix}1 & -1 & 1 & 0\\ 0 & 2 & 4 & 1\\ 0 & 0 & 1 & -2\\ 0 & 0 & 0 & -23\end{pmatrix}$$

Since there is no zero rows we know that the system and no free columns we conclude that the system has a unique solution for all $\mathbf{c}$. Therefore $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is a basis of $\mathbf{R}^4$.

All the bases of $\mathbb{R}^n$ we have encountered so far have exactly n vectors. The systems we obtain when we try to express an n-vector as a linear combination of a set with m elements have n equations and m variables. Thus if we have a set with more than n vectors the system will have free variables so it's impossible to have unique solution. If on the other hand, there are less than n vectors the echelon form of the matrix will have zero rows and therefore it won't be consistent for all $\mathbf{c}$.

In other words if we have more than n vectors we can't have uniqueness of solutions, and if we have less than n vectors we can't always have existence of solutions.

So we proved the following theorem, that as we will see, says that the dimension of $\mathbb{R}^n$ is n.

THEOREM 2.1.6. *All bases of $\mathbb{R}^n$ have exactly n elements.*

Of course, not all sets with n elements are bases of $\mathbb{R}^n$. In Section 1.3.2 we show that if two vectors are colinear then they don't form a basis.

***Question* 1.** How about subspaces though? How can we find a basis of a subspace? Does any subspace of $\mathbb{R}^n$ have a basis? If so do all bases of a subspace have the same cardinality?

We'll answer these questions in the next class. As a preparation work through the following example.

EXAMPLE 18. Consider the vectors $\mathbf{v} = (1,0,-1)$, $\mathbf{u} = (2,1,0)$, and $\mathbf{w} = (-1,1,3)$. When is a vector $\mathbf{c} = (c_1,c_2,c_3)$ in the linear span of these three vectors?

The question again reduces to solving the vector equation

$$x\,\mathbf{v} + y\,\mathbf{u} + z\,\mathbf{w} = \mathbf{c},$$

or equivalently, the system with augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & c_1 \\ 0 & 1 & 1 & c_2 \\ -1 & 0 & 3 & c_3 \end{array}\right)$$

Adding the first row to the third, and then subtracting twice the second row from the third we get the following echelon form:

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & c_1 \\ 0 & 1 & 1 & c_2 \\ 0 & 0 & 0 & c_1 - 2\,c_2 + c_3 \end{array}\right).$$

So in order for the system to have solutions it is necessary to have

$$c_1 - 2\,c_2 + c_3 = 0 \iff c_3 = -c_1 + 2\,c_2. \tag{2.2}$$

When that condition is satisfied we can discard the third row, and then subtract twice the second row from the first we get:

$$\left(\begin{array}{ccc|c} 1 & 0 & -3 & -c_1 + 2\,c_2 \\ 0 & 1 & 1 & c_2 \end{array}\right).$$

So the condition (2.2) is also sufficient.

We conclude then that

$$\langle \mathbf{v}, \mathbf{u}, \mathbf{w} \rangle = \{(c_1, c_2, -c_1 + 2\,c_2) : c_1, c_2 \in \mathbb{R}\}.$$

Observe now that,

$$(c_1, c_2, -c_1 + 2c_2) = c_1(1,0,-1) + c_2(0,1,2) = c_1\,\mathbf{v} + c_2\,\mathbf{a}.$$

So

$$\langle \mathbf{v}, \mathbf{u}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{a} \rangle.$$

Let's express $\mathbf{a}$ as a linear combination of $\mathbf{v}, \mathbf{u}, \mathbf{w}$. The reduced echelon form tells us how to do so.

$$\mathbf{a} = (3\,z - 2)\mathbf{v} + (-z + 1)\mathbf{u} + z\mathbf{w}$$

where z is any real number. Taking $z = 0$ we get

$$\mathbf{a} = -2\,\mathbf{v} + \mathbf{u}$$

while taking $z = 1$ we get

$$\mathbf{a} = \mathbf{v} + \mathbf{w}.$$

2.2. Linear dependence, Dimension

Let's take a closer look at Example 18. Let

$$V = \langle \mathbf{v}, \mathbf{u}, \mathbf{w} \rangle$$

be the linear span of the vectors defined there. We'll look for a basis of V.

Before proceeding we introduce the term *spanning subset*.

DEFINITION 9. Let $V \subseteq \mathbb{R}^n$ be a vector subspace. We say that a subset $S \subseteq V$ is a *spanning subset of* V (or simply, when V is understood, *spanning*) if

$$V = \langle S \rangle ,$$

i.e. every vector in V is a linear combination of vectors from S.

A spanning subset B of V is said to be *a basis of* V if every vector of V can be written as a linear combination of vectors from B in a *unique* way.

So if $S = \{\mathbf{v}, \mathbf{u}, \mathbf{w}\}$ then S is a spanning subset of V. However S is not a basis, because as we saw in Example 18 the vector $\mathbf{a} = (0, 1, 2)$ is equal to two different linear combinations of vectors of S, namely

$$\mathbf{a} = -2\,\mathbf{v} + \mathbf{u} \text{ and } \mathbf{a} = \mathbf{v} + \mathbf{w}. \tag{2.3}$$

An other spanning set of V is $B = \{\mathbf{v}, \mathbf{a}\}$ so let's check if this set is a basis. We want to check whether the vector equation

$$x\,\mathbf{v} + y\,\mathbf{a} = \mathbf{c}$$

has a unique solution for all $\mathbf{c} \in V$. Equivalently, we want to check whether the linear system

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix},$$

has a unique solution for all $\mathbf{c} \in V$. By Theorem 1.2.8 this happens when the homogeneous system

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

has a unique solution. The reduced echelon form of the matrix is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix},$$

and therefore the homogeneous system indeed has only the trivial solution. We conclude then that the set $B = \{\mathbf{v}, \mathbf{a}\}$ is a basis of V.

Once we find a basis we can find many more. For example, the set $B' = \{\mathbf{v}, \mathbf{w}\}$ is also a basis. This follows from the fact that B is a basis and the second equation in (2.3).

CLAIM 1. *B' is a basis of V.*

PROOF. The proof consists of two steps.

Step 1: B' is a spanning subset of V. Let $\mathbf{c} \in V$ then since B is a basis there are $x, y \in \mathbb{R}$ such that

$$\mathbf{c} = x\,\mathbf{v} + y\,\mathbf{a}.$$

But since $\mathbf{a} = \mathbf{v} + \mathbf{w}$ we have

$$\begin{aligned}\mathbf{c} &= x\,\mathbf{v} + y\,\mathbf{a}\\ &= x\,\mathbf{v} + y\,(\mathbf{v} + \mathbf{w})\\ &= x\,\mathbf{v} + y\,\mathbf{v} + y\,\mathbf{w}\\ &= (x + y)\,\mathbf{v} + y\,\mathbf{w}.\end{aligned}$$

So $\mathbf{c}$ can be expressed as a linear combination of vectors from B'.

Step 2: We now need to prove that any $\mathbf{c} \in V$ can be expressed as a linear combination of elements of B' in a unique way. So we have to prove that if two linear combinations of $\mathbf{v}, \mathbf{w}$ are equal then they are the same linear combination. In other words, we need to prove that if

$$x_1\,\mathbf{v} + y_1\,\mathbf{w} = x_2\,\mathbf{v} + y_2\,\mathbf{w} \tag{2.4}$$

then

$$x_1 = x_2 \text{ and } y_1 = y_2.$$

We will again use the fact that the uniqueness property holds for B. Now since $\mathbf{w} = \mathbf{a} - \mathbf{v}$ we have

$$\begin{aligned}x_1\,\mathbf{v} + y_1\,\mathbf{w} &= x_1\,\mathbf{v} + y_1\,(\mathbf{a} - \mathbf{v})\\ &= (x_1 - y_1)\,\mathbf{v} + y_1\,\mathbf{a}.\end{aligned}$$

and similarly

$$x_2\,\mathbf{v} + y_2\,\mathbf{w} = (x_2 - y_2)\,\mathbf{v} + y_2\,\mathbf{a}.$$

So if Equation (2.4) holds we have

$$(x_1 - y_1)\,\mathbf{v} + y_1\,\mathbf{a} = (x_2 - y_2)\,\mathbf{v} + y_2\,\mathbf{a}.$$

So we have two linear combinations of $\mathbf{v}, \mathbf{a}$ that represent the same vector. Since B is a basis this implies that the coefficients of these two linear combinations have to be equal. So we have

$$x_1 - y_1 = x_2 - y_2 \text{ and } y_1 = y_2 \implies x_1 = x_2 \text{ and } y_1 = y_2.$$

□

Notice that the above will work for any $\mathbf{v}, \mathbf{a}, \mathbf{w}$. If $\{\mathbf{v}, \mathbf{a}\}$ is a basis of a subspace V and $\mathbf{a} = \mathbf{v} + \mathbf{w}$ then $\{\mathbf{v}, \mathbf{w}\}$ is also a basis of V.

Exercise 1. Let $\mathbf{v}, \mathbf{u}, \mathbf{w}, \mathbf{a} \in \mathbb{R}^n$ and V a vector subspace of $\mathbb{R}^n$ such that the following hold:

(a) $\{\mathbf{v}, \mathbf{a}\}$ is a basis of V.
(b) $\mathbf{a} = \mathbf{v} + \mathbf{w}$.
(c) $\mathbf{a} = -2\,\mathbf{v} + \mathbf{u}$.

Prove that any two of those four vectors form a basis. That is, prove that each one of

$$\{\mathbf{v}, \mathbf{w}\}, \quad \{\mathbf{v}, \mathbf{u}\}, \quad \{\mathbf{u}, \mathbf{w}\}, \quad \{\mathbf{a}, \mathbf{w}\}, \quad \{\mathbf{a}, \mathbf{u}\}$$

is also a basis.

Consider again a general vector subspace of $V \subseteq \mathbb{R}^n$, and let B be a spanning set of V. In order for B to be a basis every vector of V has to have a unique expression as a linear combination of elements of B. In particular, the zero vector which is an element of V, has to have only one representation as a linear combination of elements of B. But we can easily find a linear combination that represents **0**, namely the one where all coefficients are 0. Therefore if there is a non-trivial linear combination

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_m \mathbf{v}_m = \mathbf{0},$$

with $\mathbf{v}_1, \ldots, \mathbf{v}_m \in B$ and coefficients $\lambda_1, \lambda_2, \ldots, \lambda_m$ not all 0, then B is not a basis.

It turns out that that's the only way to prevent a spanning set from being a basis. If the zero vector can be expressed as a linear combination of vectors from B in only one way, then every other vector of V has also a unique expression. To see this let's assume that for some $\mathbf{v} \in V$ we have two different expressions

$$\mathbf{v} = \lambda_1 \mathbf{v}_1 + \ldots + \lambda_k \mathbf{v}_k$$

and

$$\mathbf{v} = \mu_1 \mathbf{u}_1 + \ldots + \mu_m \mathbf{u}_m,$$

where $\lambda_i \in \mathbb{R}$, $\mathbf{v}_i \in B$ for $i = 1, \ldots, k$ and $\mu_j \in \mathbb{R}$, $\mathbf{u}_j \in B$ for $j = 1, \ldots, m$. Then, by adding terms of the form $0 \cdot \mathbf{u}_j$ to the first expression and terms of the form $0\,\mathbf{v}_i$ to the second if necessary, we can get two linear combinations where exactly the same vectors from B occur. Let's then assume that we have two linear combinations

$$\mathbf{v} = \lambda_1 \mathbf{w}_1 + \ldots + \lambda_\ell \mathbf{w}_\ell,$$

and

$$\mathbf{v} = \mu_1 \mathbf{w}_1 + \ldots + \mu_\ell \mathbf{u}_\ell,$$

where for some k, $\lambda_k \neq \mu_k$. But then subtracting we have

$$\mathbf{0} = \sum_{i=1}^{\ell} (\lambda_i - \mu_i)\, \mathbf{w}_i$$

and the k-th term $\lambda_k - \mu_k \neq 0$. So we got a non-trivial linear combination representing the zero vector.

We have thus proved the following Lemma.

LEMMA 1. *Let $S \subseteq \mathbb{R}^n$ be any set. If there are two linear combinations of elements from S with different coefficients represent the same vector then the zero vector is represented by a non-trivial linear combination of elements from S.*

DEFINITION 10 (**Linearly dependent and linearly independent sets**). A non-trivial linear combination that represents the zero vector, that is an equation of the form

$$\sum_{i=1}^{m} \lambda_i \mathbf{v}_i = \mathbf{0}$$

with $\lambda_i \neq 0$ for some $i \in \{1, \ldots, m\}$, is called a *linear dependency condition among* $\mathbf{v}_1, \ldots, \mathbf{v}_m$.

If there is a linear dependency condition among some elements of a subset $S \subseteq \mathbb{R}^n$ we say that S is *linearly dependent*.

If S is not linearly dependent we say that it is *linearly independent*.

With this terminology in place we can summarize the results of our discussion so far in the following theorem.

THEOREM 2.2.1. *Let V be a vector subspace of $\mathbb{R}^n$ and $B \subseteq V$. Then B is a basis of V if and only if it is spanning and linearly independent.*

THEOREM 2.2.2. *The following hold.*

(a) If $\mathbf{0} \in S$ then S is linearly dependent.
(b) If $S = \{\mathbf{v}\}$ then S is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}$.

(c) *If $S = \{\mathbf{v}, \mathbf{w}\}$ then if and only if $\mathbf{v} = \lambda\,\mathbf{w}$ or $\mathbf{w} = \lambda\,\mathbf{v}$ for some scalar λ.*
(d) *If S is linearly independent and $\mathbf{v}_1, \dots, \mathbf{v}_m$ are distinct elements of S then $\mathbf{v}_1$ cannot be expressed as a linear combination of $\mathbf{v}_2, \dots, \mathbf{v}_m$.*
(e) *If $S \subseteq S'$ and S is linearly dependent then S' is linearly dependent as well.*
(f) *If $S \subseteq S'$ and S' is linearly independent then S is linearly independent as well.*

PROOF. (a) We have $42 \cdot \mathbf{0} = \mathbf{0}$ an expression of the zero vector as a non-trivial linear combination of vectors from S.
(b) By Item 1 $\{\mathbf{0}\}$ is linearly dependent. Conversely, if $\mathbf{v} \neq \mathbf{0}$ then

$$\lambda\,\mathbf{v} = \mathbf{0} \iff \lambda = 0.$$

Thus if $\mathbf{v} \neq \mathbf{0}$ only the trivial linear combination is equal to the zero vector.
(c) Since $\mathbf{v}_1 = 1\,\mathbf{v}_1$ expresses $\mathbf{v}_1$ as a linear combinations of elements of S, there is no other linear combination.
(d) A linear dependency among elements of S is also a linear dependency among elements of S' because all elements of S are also elements of S'.
(e) This is the contra-positive of the previous item.

□

THEOREM 2.2.3. *If V has a basis B with cardinality d then any linearly independent subset of V with d elements is also a basis of V.*

The idea of the proof is contained in the proof of Claim 1. If B' is linearly independent subset of V with d elements we will construct a sequence of sets $B_0, B_1, B_2, \dots, B_d$, where $B_0 = B$ and $B_d = B'$, and prove that all of them are bases. B_1 is obtained from B by replacing one element, say $\mathbf{v}_1$ with an element from B'. B_2 is obtained by B_1 by replacing one more element of B by an element of B'. At every step we get a basis B_i that has i elements from B' and the remaining $d - i$ from B. At the next step to get B_{i+1} we replace one of those elements of B_i that are in B with a new element of B'. Eventually all the elements of B have been replaced by the elements of B' and since at every step we still get a basis, we conclude that B' is a basis.

We first prove the following Lemma.

LEMMA 2. *If $B = \{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ is a basis of V and $\mathbf{w}_1 \in V$ is such that*

$$\mathbf{w}_1 = \lambda_1\,\mathbf{v}_1 + \lambda_2\,\mathbf{v}_2 + \cdots + \mathbf{v}_d$$

with $\lambda_1 \neq 0$ then $B' = \{\mathbf{w}_1, \mathbf{v}_2 \dots, \mathbf{v}_d\}$ is also a basis.

PROOF. Since $\lambda_1 \neq 0$ we can express $\mathbf{v}_1$ as a linear combination of $\mathbf{w}, \mathbf{v}_2, \dots, \mathbf{v}_n$:

$$\mathbf{v}_1 = \frac{1}{\lambda_1}\mathbf{w}_1 - \frac{\lambda_2}{\lambda_1}\,\mathbf{v}_2 - \cdots - \frac{\lambda_d}{\lambda_1}\,\mathbf{v}_d. \tag{2.5}$$

Let be an arbitrary element of V. Then we can write $\mathbf{c}$ as a linear combination

$$\mathbf{c} = \mu_1\,\mathbf{v}_1 + \mu_2\,\mathbf{v}_2 + \cdots + \mu_d\,v_d.$$

Substituting the RHS of Equation (2.5) for $\mathbf{v}_1$ and collecting terms gives

$$\mathbf{c} = \frac{\mu_1}{\lambda_1}\,\mathbf{w}_1 + \left(\mu_2 - \frac{\lambda_2}{\lambda_1}\right)\,\mathbf{v}_2 + \cdots + \left(\mu_d - \frac{\lambda_d}{\lambda_1}\right)\,v_d.$$

Therefore B' is spanning.

To prove that B' is also linearly independent, consider a linear dependency

$$\mu_1\,\mathbf{w}_1 + \mu_2\,\mathbf{v}_2 + \cdots + \mu_d\,\mathbf{v}_d = \mathbf{0}.$$

Substituting $\mathbf{w}_1$ with its expression in terms of B we have

$$\mu_1\,\lambda_1\,\mathbf{v}_1 + (\mu_1\,\lambda_2 + \mu_2)\,\mathbf{v}_2 + \cdots + (\mu_1\,\lambda_d + \mu_d)\,\mathbf{v}_d = \mathbf{0}.$$

Since B is a basis all the coefficients in this linear dependency have to be 0. Since $\lambda_1 \neq 0$ we get from the coefficient of $\mathbf{v}_1$ that $\mu_1 = 0$. Substituting in the other coefficients then gives $\mu_i = 0$ for $i = 2, \ldots, d$ as well. □

REMARK 8. Lemma 2 says that we can replace *any* element $\mathbf{v} \in B$ by $\mathbf{w}$ as long as $\mathbf{v}$ appears with non-zero combination in the expression of $\mathbf{w}$ as linear combination of elements of B. For, we can order the elements of B so that $\mathbf{v}$ comes first.

PROOF OF THEOREM 2.2.3. Let B' be a linear independent subset of V with d elements. Chose an arbitrary order of B', say $B' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_d\}$.

Now express $\mathbf{w}_1$ as a linear combination of elements of B. Since B' is linearly independent, $\mathbf{w}_1 \neq \mathbf{0}$, and so at least one element of B will appear with non-zero coefficient in that linear combination, call that element $\mathbf{v}_1$. By Lemma 2 the set

$$B_1 = (B \setminus \{\mathbf{v}_1\}) \cup \{\mathbf{w}_1\},$$

i.e. the set obtained from B by replacing $\mathbf{v}_1$ with $\mathbf{w}_1$, is a basis.

Next express $\mathbf{w}_2$ as a linear combination of the elements of B_1. In that linear combination at least one element of B appears with non-zero coefficient, because otherwise $\mathbf{w}_2$ would be a multiple of $\mathbf{w}_1$, impossible since B' is linearly independent. Choose one such element, say $\mathbf{v}_2$, and let B_2 be the set obtained by B_1 by replacing $\mathbf{v}_2$ with $\mathbf{w}_2$, i.e.

$$B_2 = (B_1 \setminus \{\mathbf{v}_2\}) \cup \{\mathbf{w}_2\} = (B \setminus \{\mathbf{v}_1, \mathbf{v}_2\}) \cup \{\mathbf{w}_1, \mathbf{w}_2\}.$$

Again by Lemma 2, B_2 is a basis.

Next, assuming $d > 2$, we express $\mathbf{w}_3$ as a linear combination of elements of B_2. In that linear combination at least one element of B appears with non-zero coefficient, otherwise $\mathbf{w}_3$ is a linear combination of $\mathbf{w}_1$ and $\mathbf{w}_2$, impossible since B' is linearly independent. Then, again by Lemma 2,

$$B_3 = (B_2 \setminus \{\mathbf{v}_3\}) \cup \{\mathbf{w}_3\}$$

is a basis.

We continue this procedure until all the elements of B have been replaced. At the k-th step we choose one of the remaining elements of B, say $\mathbf{v}_k$, that appears with non-zero coefficient in the expression of $\mathbf{w}_k$ as a linear combination of elements of B_{k-1}. Since B' is linearly independent, such $\mathbf{v}_k$ must exist. We then define B_k via

$$B_k = (B_{k-1} \setminus \{\mathbf{v}_k\}) \cup \{\mathbf{w}_k\}.$$

By Lemma 2, B_k is a basis.

After d steps we will get $B_d = B'$ and therefore B' is a basis. □

As a corollary we have the following fundamental theorem.

THEOREM 2.2.4 (**Subspaces have well-defined dimension**). *All bases of a vector subspace have the same cardinality.*

PROOF. Let B and B' be two bases of V. We first remark that both B and B' are finite sets. Indeed a subset of $\mathbb{R}^n$ with more than n elements is linearly dependent[4].

If the cardinality of B is smaller than the cardinality of B', say B has d elements while B' has $d + k$ elements with $k > 0$, by Theorem 2.2.3, any subset S of B' with d elements would be

[4] Why?

a basis of V, and thus each of the remaining k elements of B' would be a linear combination of elements of B', contradicting Item (4) of Theorem 2.2.2.

Similarly the cardinality of B' cannot be smaller than the cardinality of B. Therefore B and B' have the same cardinality. □

One final question remains though: Does any subspace have a basis? The answer is yes. To see why let's prove the following theorem.

THEOREM 2.2.5 (**A maximal independent subset is a basis**). *A linearly independent subset B of V is a basis if and only if every subset of V that is a proper superset of B is linearly dependent. In other words, a linearly independent subset of V is a basis of V if and only if, for any S we have*

$$B \subsetneq S \subseteq V \implies S \text{ is linearly dependent.} \tag{2.6}$$

PROOF. Let $B = \{\mathbf{b}_1, \dots, \mathbf{b}_d\}$ be a basis of V, and $\mathbf{v} \in V \setminus B$, i.e. an element of V not in B. Then $B \cup \{\mathbf{v}\}$ is linearly dependent. Indeed there are scalars $\lambda_1, \dots, \lambda_d$ such that

$$\mathbf{v} = \lambda_1 \mathbf{b}_1 + \cdots + \lambda_d \mathbf{b}_d.$$

But then

$$-1\,\mathbf{v} + \lambda_1 \mathbf{b}_1 + \cdots + \lambda_d \mathbf{b}_d = \mathbf{0}.$$

So $\mathbf{0}$ can be expressed as a non-trivial linear combination of $B \cup \{\mathbf{v}\}$, and thus $B \cup \{\mathbf{v}\}$ is linearly dependent. Now if

$$B \subsetneq S \subseteq V$$

then there is an element $\mathbf{v} \in S \setminus B$ and for such a $\mathbf{v}$

$$B \cup \{\mathbf{v}\} \subseteq S$$

and thus S has a linearly dependent subset. By Item (5) of Theorem 2.2.2 we conclude that S is linearly dependent.

Conversely, assume that (2.6) holds. To prove that B is a basis we need to prove that it is spanning. Consider then $\mathbf{v} \in V$, if $\mathbf{v} \in B$ then clearly $\mathbf{v}$ is a linear combination of elements of B. Assume then that $\mathbf{v} \notin B$, in which case $B \cup \{\mathbf{v}\}$ is linearly independent. Therefore there are $\lambda, \lambda_1, \dots, \lambda_d \in \mathbb{R}$ such that

$$\lambda\,\mathbf{v} + \lambda_1 \mathbf{b}_1 + \cdots + \lambda_d \mathbf{b}_d = \mathbf{0}$$

with $\lambda, \lambda_1, \dots, \lambda_d$ not all 0. Then $\lambda \neq 0$ because otherwise we would have a linear dependency among the elements of B, and therefore

$$\mathbf{v} = -\frac{\lambda_1}{\lambda}\mathbf{b}_1 - \cdots - \frac{\lambda_d}{\lambda}\mathbf{b}_d,$$

and we expressed $\mathbf{v}$ as a linear combination of the elements of B. Thus, all elements of V can be expressed as linear combinations of the elements of B. □

We now can prove that every vector subspace has a basis.

THEOREM 2.2.6 (**Every subspace has a basis**). *We first consider $V = \{\mathbf{0}\}$. Then $B = \varnothing$, the empty set, is a basis of V. Indeed $\varnothing$ is linearly independent, vacuously. The only set S that satisfies the hypothesis of (2.6) is V itself, and is linearly dependent.*

If $V \neq \{\mathbf{0}\}$ we can find a basis as follows. Chose any $\mathbf{v}_1 \in V$ with $\mathbf{v}_1 \neq \mathbf{0}$. Then $S_1 := \{\mathbf{v}_1\}$ is linearly independent. If $\langle S_1 \rangle = V$ then S_1 is a basis. If not chose a second vector $\mathbf{v}_2 \in V$ not in $\langle S \rangle$ and consider the set $S_2 := \{v_1, v_2\}$. Then S_2 is linearly independent otherwise $\mathbf{v}_2$ would be in $\langle S_1 \rangle$. If $\langle S_2 \rangle = V$ then S_2 is a basis of V.

We continue this way until we get a linearly independent set S_d with $\langle S_d \rangle = V$. This process cannot continue for ever because we know that we can't choose more than n linearly independent vectors, so *we can continue for at most n steps. This means that after a finite number of steps, say d, we won't be able to find any vectors in V that are not in the linear span of S_d. The set S_d then will be a basis of V.*

We end this section with the definition of the very important concept of dimension.

DEFINITION 11. Let V be a subspace of $\mathbb{R}^n$. The common cardinality of all the bases of V is called the *dimension of V* and is denoted by $\dim V$. If the dimension of V is d we also say that V is a *d-dimensional* subspace of of $\mathbb{R}^n$.

A one-dimensional subspace is sometimes called a *line* and a two dimensional subspace a *plane*.

2.2.1. How to find a basis. We give a few examples that illustrate the concepts we've described so far, and develop a method for finding a basis of a subspace if we have a finite spanning set.

EXAMPLE 19. Which of the following subsets of $\mathbb{R}^4$ are vector subspaces?

(a) $V = \{(a,0,b,0) : a,b \in \mathbb{R}\}$.
(b) $V = \{(a,1,b,0) : a,b \in \mathbb{R}\}$.
(c) $V = \{(a-2b,3c,b-a,d) : a,b,c,d \in \mathbb{R}\}$.
(d) $V = \{(a,b,c,d) : a,b,c,d \in \mathbb{R}$ with $d > 0\}$.

ANSWER. (a) This set is a subspace. To prove this we will prove that the two conditions in Theorem 2.1.3 are satisfied.

(a) $V \neq \varnothing$ because by setting, for example $a = 0, b = 0$ we have that $(0,0,0,0) \in V$.
(b) Let $\mathbf{v},\mathbf{w} \in V$ and $\lambda,\mu \in \mathbb{R}$. Then for some $a_1,b_1,a_2,b_2 \in \mathbb{R}$ we have

$$\mathbf{v} = (a_1,0,b_1,0), \quad \mathbf{w} = (a_2,0,b_2,0).$$

Then

$$\begin{aligned}\lambda\mathbf{v} + \mu\mathbf{w} &= \lambda(a_1,0,b_1,0) + \mu(a_2,0,b_2,0)\\ &= (\lambda a_1,0,\lambda b_1,0) + (\mu a_2,0,\mu b_2,0)\\ &= (\lambda a_1 + \mu a_2, 0, \lambda b_1 + \mu b_2, 0).\end{aligned}$$

Therefore $\lambda\mathbf{v} + \mu\mathbf{w} = (a,0,b,0)$ where $a = \lambda a_1 + \mu a_2$, and $b = \lambda b_1 + \mu b_2$ are real numbers. It follows that

$$\lambda\mathbf{v} + \mu\mathbf{w} \in V.$$

(b) V is not a vector subspace since $\mathbf{0} \notin V$.
(c) V is a vector subspace. We can proceed as in Item (1) and show that the two properties of Theorem 2.1.3 are satisfied[5]. An other method is to show that V is the linear span of a subset of $\mathbb{R}^4$. Then by Theorem 2.1.5 V is a subspace[6].

For all real numbers a,b,c,d we have

$$\begin{aligned}(a-2b,3c,b-a,d) &= (a,0,-a,0) + (-2b,0,b,0) + (0,3c,0,0) + (0,0,0,d)\\ &= a(1,0,-1,0) + b(-2,0,1,0) + c(0,3,0,0) + d(0,0,0,1).\end{aligned}$$

Thus V consists of all linear combinations of the vectors

$$(1,0,-1,0),(-2,0,1,0),(0,3,0,0),(0,0,0,1)$$

and is therefore the linear span of these vectors.

It follows by Theorem 2.1.5 that V is a subspace of $\mathbb{R}^4$.

[5] Do this

[6] Use this method for Item (1). That is prove that the set in Item 1 is the linear span of a certain set of vectors.

(d) V is not a vector subspace because it is not closed under scalar multiplication. For example $(0,0,0,1) \in V$ but $-1\,(0,0,0,1) = (0,0,0,-1) \notin V$.

□

EXAMPLE 20. Find a basis for each of the sets in Example 19 that is a subspace.

SOLUTION. I will do Item (3), and leave Item (1) as an exercise.

Let $\mathbf{v}_1 = (1,0,-1,0)$, $\mathbf{v}_2 = (-2,0,1,0)$, $\mathbf{v}_2 = (0,3,0,0)$, and $\mathbf{v}_4 = (0,0,0,1)$. Since $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is a spanning set we check if S is linearly independent. If it is then it forms a basis.

S is linearly independent if and only if the homogeneous system $A\mathbf{x} = \mathbf{0}$, where A is the matrix with columns $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, and $\mathbf{v}_4$, has a unique solution. We therefore have to find an echelon form of A.

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

Since there are no free columns it follows that the homogeneous system has only the trivial solution and therefore S is linearly independent. Thus S is a basis of V. □

EXAMPLE 21. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} \subseteq \mathbb{R}^5$, where

$$\begin{aligned} \mathbf{v}_1 &= (1,1,1,2,3), \\ \mathbf{v}_2 &= (1,2,-1,-2,1) \\ \mathbf{v}_3 &= (3,5,-1,-2,5) \\ \mathbf{v}_4 &= (1,2,1,-1,4). \end{aligned}$$

Find a basis for $V = \langle S \rangle$. What is $\dim V$?

SOLUTION. We again consider the matrix with columns the vectors of S.

$$A = \begin{pmatrix} 1 & 1 & 3 & 1 \\ 1 & 2 & 5 & 2 \\ 1 & -1 & -1 & 1 \\ 2 & -2 & -2 & -1 \\ 3 & 1 & 5 & 4 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 3 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & -2 & -4 & 0 \\ 0 & -4 & -8 & -3 \\ 0 & -2 & -4 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 3 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 3 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Since the reduced echelon form has free columns the homogeneous system $A\mathbf{x} = \mathbf{0}$ has non-trivial solutions. Each non trivial solution gives a non-trivial linear combination of S that is equal to $\mathbf{0}$.

The solution set is $\{(-t, -2t, t, 0) : t \in \mathbb{R}\}$ so by setting $t = -1$ we get $x_1 = 1, x_2 = 2, x_3 = -1, x_4 = 0$. Thus we have the following non-trivial linear dependency

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

and it follows that

$$\mathbf{v}_3 = \mathbf{v}_1 + 2\,\mathbf{v}_2.$$

We can then throw away v_3 and still have a spanning set. That is,

$$V = \langle \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4 \rangle\,.$$

Now, $B := \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$ is linearly independent. Indeed the the first,second, and fourth columns, of the reduced echelon form of A give the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This matrix is therefore the reduced echelon form of the matrix with columns the elements of B. Since there are no free columns only the trivial linear combination of B gives the zero vector.

Since B is a basis of V, and B has three elements we have $\dim V = 3$. □

Notice that in the previous example it turned out that the vectors that correspond to the basic columns actually form a basis of the linear span. This is always the case, and the reason that we call non-free columns basic.

Let's see one more example.

EXAMPLE 22. Find a basis and the dimension of the linear span of the vectors

$$\begin{aligned} \mathbf{v}_1 &= (3, 0, 6, 3), & \mathbf{v}_2 &= (-7, 3, -16, -9), \\ \mathbf{v}_3 &= (8, -6, 20, 12), & \mathbf{v}_4 &= (-5, 6, -14, -9), \\ \mathbf{v}_5 &= (8, 4, 14, 6), & \mathbf{v}_6 &= (9, -5, 24, 15). \end{aligned}$$

SOLUTION. The matrix with columns these vectors is:

$$A = \begin{pmatrix} 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \\ 6 & -16 & 20 & -14 & 14 & 24 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{pmatrix}.$$

To get an echelon form of A we start by adding -2 times the first row to the third, subtracting the first row from the fourth. Then we subtract the third row from the fourth and that turns the fourth row in to a zero row and we discard it. Then we add 2 times the second row to 3 times the third, and divide the last row by 2

$$A \sim \begin{pmatrix} 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \\ 0 & -2 & 4 & -4 & -2 & 6 \\ 0 & -2 & 4 & -4 & -2 & 6 \end{pmatrix} \sim \begin{pmatrix} 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \\ 0 & -2 & 4 & -4 & -2 & 6 \end{pmatrix} \sim \begin{pmatrix} 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{pmatrix}.$$

From the echelon form we see that the basic columns are the first, second and fifth. From the discussion above it follows that a basis of the linear span is $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_5\}$. Since there are three vectors in the basis we have that the dimension of the linear span is 3. □

EXAMPLE 23. We use the same notation as in Example 22.

(a) Express each of the "free" vectors $\mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_6$ as a linear combination of the elements of B.

(b) Find a fourth vector $\mathbf{w}$ to complete B to a basis of $\mathbb{R}^4$. In other words, the set $\{\mathbf{w}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_5\}$ should be a basis of $\mathbb{R}^4$.

SOLUTION. (a) We need to solve the systems

$$B\,\mathbf{x} = \mathbf{v}_3, \quad B\,\mathbf{x} = \mathbf{v}_4, \quad B\,\mathbf{x} = \mathbf{v}_6,$$

where B is the matrix with columns the basic vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_5$, that is

$$B = \begin{pmatrix} 3 & -7 & 8 \\ 0 & 3 & 4 \\ 6 & -16 & 14 \\ 3 & -9 & 6 \end{pmatrix}.$$

Since all these systems have the same coefficients to solve the we will apply the same row operations to A. Instead of considering three different augmented matrices, we augment A with three columns and operate at all of them at once. So we'll get the *reduced* echelon of the following matrix:

$$\left(\begin{array}{ccc|ccc} 3 & -7 & 8 & 8 & -5 & 9 \\ 0 & 3 & 4 & -6 & 6 & -5 \\ 6 & -16 & 14 & 20 & -14 & 24 \\ 3 & -9 & 6 & 12 & -9 & 15 \end{array}\right).$$

Notice that this matrix has the same columns as A of Example 22, but permuted namely the fifth column has been moved to the third place, and the third and fourth to the fourth and fifth place, respectively. So if we apply the row operations of Example 22 we'll get the reduced form of A with columns permuted the same way, that is the following matrix

$$\left(\begin{array}{ccc|ccc} 3 & -7 & 8 & 8 & -5 & 9 \\ 0 & 3 & 4 & -6 & 6 & -5 \\ 0 & 0 & 1 & 0 & 0 & 4 \end{array}\right).$$

The reduced echelon form of the last matrix is[7]

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & 3 & -24 \\ 0 & 1 & 0 & -2 & 2 & -7 \\ 0 & 0 & 1 & 0 & 0 & 4 \end{array}\right).$$

Therefore,

$$\begin{aligned} \mathbf{v}_3 &= -2\,\mathbf{v}_1 - 2\,\mathbf{v}_2 \\ \mathbf{v}_4 &= 3\,\mathbf{v}_1 + 2\,\mathbf{v}_2 \\ \mathbf{v}_6 &= -24\,\mathbf{v}_1 - 7\,\mathbf{v}_2 + 4\,\mathbf{v}_5. \end{aligned}$$

(b) Any linearly independent subset of $\mathbb{R}^4$ forms a basis. Therefore the set $\{\mathbf{w}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_5\}$ will be a basis if (and only if) it is linearly independent, that is if and only if $\mathbf{w} \notin \langle \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_5 \rangle$. So we need to find a vector $\mathbf{w}$ so that the system

$$B\,\mathbf{x} = \mathbf{w}$$

has no solutions. By Theorem 1.2.8 this happens if and only if the echelon form of its augmented matrix contains a row of the form

$$\left(0 \quad 0 \quad 0 \quad | \quad c\right)$$

with $c \neq 0$. Now recall that in the process of obtaining the echelon form of A we discarded a zero row. This happened after applying the following row operations:

(a) Add -2 times the first row to the third.

(b) Add the first row to the fourth.

[7] Do the calculations and verify this.

(c) Subtract the third row from the fourth.

After those operations the matrix B becomes

$$\begin{pmatrix} 3 & -7 & 8 \\ 0 & 3 & 4 \\ 0 & -2 & -2 \\ 0 & 0 & 0 \end{pmatrix}.$$

So we have to choose a vector $\mathbf{w}$ that the row operations listed above transform it to a vector $\mathbf{w}'$ with non zero fourth coordinate. The simplest choice for such a w' is $\mathbf{e}_4$. Assume then that $\mathbf{w}$ is such that after these three row operations the augmented matrix of the system $B\,\mathbf{x} = \mathbf{w}$ is

$$\left(\begin{array}{ccc|c} 3 & -7 & 8 & 0 \\ 0 & 3 & 4 & 0 \\ 0 & -2 & -2 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right).$$

To recover $\mathbf{w}$ we have to reverse the effect of the rows operations. In other words, we need to apply to $\mathbf{e}_4$ the following operations:

(a) Add the third row to the fourth.

(b) Add the first row the the fourth.

(c) Add 2 times the first row to the third.

None of these reverse operations change $\mathbf{e}_4$ though. Thus $\mathbf{w} = \mathbf{e}_4$. So the set

$$\{\mathbf{w}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_5\}$$

is a basis of $\mathbb{R}^4$.

□

Basis of linear span

To find a base of the linear span of k vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in \mathbb{R}^n$

(a) Create an $n \times k$ matrix A that has the given vectors as columns

$$A = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots \mathbf{v}_k \end{pmatrix}.$$

(b) Find an echelon form for A.

(c) A basis consists of the columns of A that correspond to the basic columns of the echelon form.

CHAPTER 3

Matrices and their algebra

3.1. Matrices as transformations

We have already introduced the notation $A\mathbf{x}$ where A is an $m \times n$ matrix andx is an n-vector. We were writing a system of m equations with n variables as

$$A\mathbf{x} = \mathbf{c}, \tag{3.1}$$

where A is the matrix with entries the coefficients of the equations, $\mathbf{x}$ is the column vector of the variables, and $\mathbf{c}$ is the column vector of constants.

If we expand the LHS we get an equation of two m-vectors namely,

$$\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}. \tag{3.2}$$

Before proceeding, let's officially define the product of a matrix and a column vector.

DEFINITION 12. If A is an $m \times n$ matrix and $\mathbf{x}$ an $n \times 1$ column vector the *product* $A\mathbf{x}$ is defined to be the LHS of Equation (3.2). The result is thus an $m \times 1$ column vector whose k-th row consists of the element

$$a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n = \sum_{i=1}^{n} a_{ki}x_i.$$

REMARK 9. Notice: in order for the product $A\mathbf{x}$ to be defined the dimensions have to match, the number of columns of A has to be equal to the number of rows of $\mathbf{x}$.

When the dimensions match, every row of A has as many entries as $\mathbf{x}$ and the result has as many rows as A. Furthermore each row of $A\mathbf{x}$ is the product of the corresponding row of A with $\mathbf{x}$.

We can think of $A\mathbf{x}$ as a generalization of *dot product* of two vectors as defined in *Vector Calculus*.

EXAMPLE 24. If we compute the product of a 1×3 matrix (a 3-dimensional row vector) and a 3×1 column vector, the result will be a 1×1 column matrix.

$$\begin{pmatrix} 2 & 5 & -1 \end{pmatrix} \begin{pmatrix} -3 \\ 5 \\ 7 \end{pmatrix} = \begin{pmatrix} 2 \cdot (-3) + 5 \cdot 5 + (-1) \cdot 7 \end{pmatrix} = \begin{pmatrix} 12 \end{pmatrix}.$$

In calculus classes the standard basic vectors of $\mathbb{R}^3$ are often denoted by $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Now if $\mathbf{v} = 2\,\mathbf{i} + 5\,\mathbf{j} - \mathbf{k}$ and $\mathbf{u} = -3\,\mathbf{i} + 5\,\mathbf{j} + 7\,\mathbf{k}$ then

$$\mathbf{v} \cdot \mathbf{u} = 12.$$

So the matrix product of a row vector with a column vector of the same dimension is their dot product considered as a 1×1 matrix.

$A\,\mathbf{x}$ via dot product

If $\mathbf{r}_1,\dots,\mathbf{r}_m$ are the rows of the matrix A then $A\,\mathbf{x}$ has rows $\mathbf{r}_1\,\mathbf{x},\dots,\mathbf{r}_m\,\mathbf{x}$. If we write A as a column vector of row vectors then we have

$$\begin{pmatrix}\mathbf{r}_1\\ \mathbf{r}_2\\ \vdots\\ \mathbf{r}_m\end{pmatrix}\mathbf{x}=\begin{pmatrix}\mathbf{r}_1\,\mathbf{x}\\ \mathbf{r}_2\,\mathbf{x}\\ \vdots\\ \mathbf{r}_m\,\mathbf{x}\end{pmatrix}$$

EXAMPLE 25. Calculate the product $A\,\mathbf{x}$ if defined.

(a) $A=\begin{pmatrix}1&2&0\\-2&5&1\\0&6&1\end{pmatrix}$, $\mathbf{x}=\begin{pmatrix}3\\-1\\2\end{pmatrix}$

ANSWER. A is a 3×3 matrix and $\mathbf{x}$ is a 3×1 column vector so the product is defined. We calculate the result row by row:

$$A\,\mathbf{x}=\begin{pmatrix}1\cdot3+2\cdot(-1)+0\cdot2\\-2\cdot3+5\cdot(-1)+1\cdot2\\0\cdot3+6\cdot(-1)+2\cdot2\end{pmatrix}=\begin{pmatrix}1\\-9\\-2\end{pmatrix}$$

□

(b) $A=\begin{pmatrix}-2&5&0&-7\\3&4&-1&0\end{pmatrix}$, $\mathbf{x}=\begin{pmatrix}0\\\pi\\-2\end{pmatrix}$

ANSWER. The product is not defined because the number of columns of A is different than the number of rows of $\mathbf{x}$: A is 2×4 and $\mathbf{x}$ is 3×1. □

(c) $A=\begin{pmatrix}-2&5&0&-7\\3&4&-1&0\end{pmatrix}$, $\mathbf{x}=\begin{pmatrix}0\\\pi\\-2\\\sqrt{3}\end{pmatrix}$

ANSWER. The dimensions now match and we have

$$\begin{pmatrix}-2&5&0&-7\\3&4&-1&0\end{pmatrix}\begin{pmatrix}0\\\pi\\-2\\\sqrt{3}\end{pmatrix}=\begin{pmatrix}-2\cdot0+5\,\pi+0\cdot(-2)-7\sqrt{3}\\3\cdot0+4\,\pi+(-1)\cdot(-2)+0\sqrt{3}\end{pmatrix}=\begin{pmatrix}5\pi-7\sqrt{3}\\2+4\,\pi\end{pmatrix}.$$

□

The product $A\,\mathbf{x}$ can be also calculated column by column. In Section 2.1, when we wanted to find the linear span of a set of vectors we saw that the vector equation $x_1\,\mathbf{a}_1+\cdots+x_n\mathbf{a}_n=\mathbf{c}$ is equivalent to the system $A\,\mathbf{x}=\mathbf{c}$, where the columns of the matrix A are the vectors $\mathbf{a}_1,\dots,\mathbf{a}_n$.

$A\,\mathbf{x}$ as linear combination of columns

If $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the columns of A, and $\mathbf{x} = (x_1, \ldots, x_n)$ then

$$A\,\mathbf{x} = x_1\,\mathbf{a}_1 + \cdots + x_n\,\mathbf{a}_n.$$

Or, if we write A as a row of column vectors

$$\begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1\,\mathbf{a}_1 + x_2\,\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n \end{pmatrix}.$$

EXAMPLE 26. Here is an example of how to compute $A\,\mathbf{x}$ column by column. To compute

$$\begin{pmatrix} 1 & 3 & -2 & 5 & 4 \\ 1 & 4 & 1 & 3 & 5 \\ 1 & 4 & 2 & 4 & 3 \\ 2 & 7 & -3 & 6 & 12 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 3 \\ -2 \\ 5 \end{pmatrix}$$

we compute the linear combination of the columns of the matrix with coefficients the components of the vector:

$$-1\begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \end{pmatrix} + 0\begin{pmatrix} 3 \\ 4 \\ 4 \\ 7 \end{pmatrix} + 3\begin{pmatrix} -2 \\ 1 \\ 2 \\ -3 \end{pmatrix} - 2\begin{pmatrix} 5 \\ 3 \\ 4 \\ 6 \end{pmatrix} + 5\begin{pmatrix} 4 \\ 5 \\ 3 \\ 12 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ -1 \\ -2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -6 \\ 3 \\ 6 \\ -9 \end{pmatrix} + \begin{pmatrix} -10 \\ -6 \\ -8 \\ -12 \end{pmatrix} + \begin{pmatrix} 20 \\ 25 \\ 15 \\ 60 \end{pmatrix} = \begin{pmatrix} 3 \\ 21 \\ 12 \\ 37 \end{pmatrix}.$$

In this section, and for the remaining of the class, we view Equation (3.1) from a different vantage point. We think of it as defining a *function* with domain $\mathbb{R}^n$ and codomain $\mathbb{R}^m$. To emphasize this new point of view let us rewrite it as

$$\mathbf{y} = A\,\mathbf{x}. \tag{3.3}$$

and consider $\mathbf{y}$ the *dependent* and $\mathbf{x}$ the *independent* variable.

If we expand Equation (3.3) as a vector equation we get

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

Finally, denoting the column vectors by $\mathbf{a}_1, \cdots, \mathbf{a}_n$ we can rewrite Equation (3.3) as

$$\mathbf{y} = x_1\,\mathbf{a}_1 + \cdots + x_n\,\mathbf{a}_n. \tag{3.4}$$

If f is a function we use the notation $f(x)$ to denote the image of x under the application of the function f. So for example if f is the function

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}, \qquad x \longmapsto x^2 + 3,$$

then $f(2) = 7$ because f maps 2 to 7.

We can think of the notation $A\mathbf{x}$ as a shorthand of $A(\mathbf{x})$, it's the image of $\mathbf{x}$ under the function A, we just omit the parenthesis. This may seem strange at first, but this is what we usually do with functions of several variables, for example we write

$$f(x,y,z)=x^2+y^2-3xz$$

for a function from $\mathbb{R}^3$ to $\mathbb{R}$. But elements of $\mathbb{R}^3$ are triples (x,y,z) so if we were really using the functional notation $f(\cdot)$ we would have written

$$f\left((x,y,z)\right)=x^2+y^2-3xz.$$

Nobody does that!

Definition 13 (**Matrices as linear transformations**). An $m\times n$ matrix with real numbers as entries determines a function

$$A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m, \qquad \mathbf{x}\longmapsto A\mathbf{x},$$

that we call the *linear function associated with A*, or the *linear function induced by A*
We use the same symbol for the matrix and the associated linear function.

The concept of a function plays a central role in mathematics and there are several names used to signify a function, for example *function*, *map*, *mapping*, *correspondence*, *transformation*, *operator*, ... There are different connotations for each of these terms but we will consider them as synonyms. In these notes besides the term "linear function" we will often use the terms "linear transformation" and "linear map".

Example 27 (**The zero matrix**). The $m\times n$ matrix with all entries 0 is called the zero $m\times n$ matrix and is denoted by O_{mn}, or when no confusion is likely, O. It induces the *zero linear function*, for all vectors $\mathbf{x}$

$$O\mathbf{x}=\mathbf{0}.$$

Example 28. Consider the 2×3 matrix

$$M=\begin{pmatrix}1 & -2 & 4\\ 2 & 0 & -1\end{pmatrix}.$$

Let's find formulas for the the function $M\colon \mathbb{R}^3\longrightarrow\mathbb{R}^2$. We have:

$$\begin{pmatrix}1 & -2 & 4\\ 2 & 0 & -1\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=\begin{pmatrix}x-2y+4z\\ 2x-3z\end{pmatrix}.$$

So we have

$$M(x,y,z)=(x-2y+4z,2x-3z).$$

In Section 1.2.2 we proved (see Theorem 1.2.5) that the function associated with a matrix has two important properties, it maps the sum of two vectors to the sum of their images and the the product of a scalar λ and a vector to the product of λ and the image of the vector. We call functions with those properties *linear functions* so Theorem 1.2.5 says that the functions defined by matrices are linear.

Definition 14 (**Linear function**). A function

$$T\colon \mathbb{R}^n\longrightarrow\mathbb{R}^m$$

is said to be *linear* if it enjoys the following two properties.

(a) *It respects vector addition*. This means that for any two vectors $\mathbf{v},\mathbf{w}\in\mathbf{R}^n$ we have

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

(b) *It respects scalar multiplication*. This means that for all $\lambda \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$ we have

$$T(\lambda\mathbf{v}) = \lambda T(\mathbf{v}).$$

EXAMPLE 29 (**The identity function is linear**). The identity function of $\mathbb{R}^n$ is denoted by I_n, or when no confusion is likely, simply by I. Thus

$$I\,\mathbf{x} = \mathbf{x}.$$

The two properties of Definition 14 are satisfied by I_n. Indeed,

(a) For $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ we have

$$I\,(\mathbf{v}+\mathbf{w}) = \mathbf{v}+\mathbf{w} = I\,\mathbf{v} + I\,\mathbf{w}.$$

(b) For $\lambda \in \mathbb{R}$ and $\mathbf{v} \in \mathbf{R}^n$ we have

$$I\,(\lambda) = \lambda\mathbf{v} = \lambda I\,\mathbf{v}.$$

EXAMPLE 30 (**Template for proving linearity or lack thereof**). Let's see a linear and a non-linear function from $\mathbb{R}^3$ to $\mathbb{R}^4$. You should use this example as a template for proving that a function is linear or not linear.

(a) The function $T\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^4$ given by the formula

$$T(x,y,z) = (3x-2y, x-2y+3z, y+z, 2x+3y-z)$$

is linear.

PROOF. To prove that the function is linear we have to prove that it satisfies the two properties in Definition 14. To prove the first property we proceed as follows:

Let $\mathbf{v} = (v_1, v_2, v_3)$ and $\mathbf{w} = (w_1, w_2, w_3)$ be two arbitrary vectors in $\mathbb{R}^3$. Then

$$\begin{aligned}\mathbf{v}+\mathbf{w} &= (v_1,v_2,v_3) + (w_1,w_2,w_3)\\ &= (v_1+w_1, v_2+w_2, v_3+w_3).\end{aligned}$$

We now will compute $T(\mathbf{v}+\mathbf{w})$. To make the calculations easier to read we use column vectors. We have

$$T(\mathbf{v}+\mathbf{w}) = \begin{pmatrix} 3\,(v_1+w_1) - 2\,(v_2+w_2) \\ (v_1+w_1) - 2\,(v_2+w_2) + 3\,(v_2+w_2) \\ (v_2+w_2) + (v_3+w_3) \\ 2\,(v_1+w_1) + 3\,(v_2+w_2) - (v_3+w_3) \end{pmatrix}$$

On the other hand,

$$T(\mathbf{v}) = \begin{pmatrix} 3v_1 - 2v_2 \\ v_1 - 2v_2 + 3v_3 \\ v_2 + v_3 \\ 2v_1 + 3v_2 - v_3 \end{pmatrix}, \qquad T(\mathbf{w}) = \begin{pmatrix} 3w_1 - 2w_2 \\ w_1 - 2w_2 + 3w_3 \\ w_2 + w_3 \\ 2w_1 + 3w_2 - w_3 \end{pmatrix}.$$

and so

$$T(\mathbf{v}) + T(\mathbf{w}) = \begin{pmatrix} (3v_1 - 2v_2) + (3w_1 - 2w_2) \\ (v_1 - 2v_2 + 3v_3) + (w_1 - 2w_2 + 3w_3) \\ (v_2 + v_3) + (w_2 + w_3) \\ (2v_1 + 3v_2 - v_3) + (2w_1 + 3w_2 - w_3) \end{pmatrix}.$$

Rearranging the terms in each component we get

$$T(\mathbf{v}) + T(\mathbf{w}) = \begin{pmatrix} (3v_1 + 3w_1) + (-2v_2 - 2w_2) \\ (v_1 + w_1) + (-2v_2 - 2w_2) + (3v_3 + 3w_3) \\ (v_2 + v_3) + (w_2 + w_3) \\ (2v_1 + 2w_1) + (3v_2 + 3w_2) + (-v_3 - w_3) \end{pmatrix}.$$

Finally taking common factors we have

$$T(\mathbf{v}) + T(\mathbf{w}) = \begin{pmatrix} 3\,(v_1 + w_1) - 2\,(v_2 + w_2) \\ (v_1 + w_1) - 2\,(v_2 + w_2) + 3\,(v_2 + w_2) \\ (v_2 + w_2) + (v_3 + w_3) \\ 2\,(v_1 + w_1) + 3\,(v_2 + w_2) - (v_3 + w_3) \end{pmatrix} = T(\mathbf{v} + \mathbf{w}).$$

Thus T respects vector addition.

To prove that T also preserves scalar multiplication we proceed similarly. Let $\lambda \in \mathbb{R}$ be an arbitrary scalar, and $\mathbf{v}$ and arbitrary vector as above. Then $\lambda\,\mathbf{v} = (\lambda\,v_1, \lambda\,v_2, \lambda\,v_3)$ and we have:

$$T(\lambda\,\mathbf{v}) = \begin{pmatrix} 3(\lambda\,v_1) - 2(\lambda\,v_2) \\ (\lambda\,v_1) - 2\,(\lambda\,v_2) + 3\,(\lambda\,v_3) \\ (\lambda\,v_2) + (\lambda\,v_3) \\ 2\,(\lambda\,v_1) + 3\,(\lambda\,v_2) - (\lambda\,v_3) \end{pmatrix} = \begin{pmatrix} \lambda\,(3\,v_1 - 2\,v_2) \\ \lambda\,(v_1 - 2\,v_2 + 3\,v_3) \\ \lambda\,(v_2 + v_3) \\ \lambda\,(2\,v_1 + 3\,v_2 - v_3) \end{pmatrix} = \lambda \begin{pmatrix} 3\,v_1 - 2\,v_2 \\ v_1 - 2\,v_2 + 3\,v_3 \\ v_2 + v_3 \\ 2\,v_1 + 3\,v_2 - v_3 \end{pmatrix} = \lambda\,T(\mathbf{v}).$$

Therefore T respects scalar multiplication as well. Thus, T is linear. □

(b) The function $T\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^4$ given by the formula

$$T(x, y, z) = (xy, x^2 + 3y - 1, y, x^3 + 3y - z^2)$$

is *not* linear.

PROOF. To prove that a function is not linear we need to prove that (at least) one of the conditions is not satisfied. To prove that a condition that is defined with universal quantifiers (i.e. it starts with *for all*) we only need to find one counterexample. I will prove that this function does not have property (2). If I choose $\lambda = 2$ and $\mathbf{v} = (0, 0, 1)$ then

$$T(\lambda\,\mathbf{v}) = T(0, 0, 2) = (0, 0, 0, -4)$$

while

$$\lambda\,T(\mathbf{v}) = 2\,(0, 0, 0, -1) = (0, 0, 0, -2).$$

Since for this particular $\lambda \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^3$ we have $T(\lambda\,\mathbf{v}) \neq \lambda\,T(\mathbf{v})$, the function is not linear. □

In the example above we could have proved that T in the first item is linear by showing that it is the linear function of a matrix. To do this we separate the terms in each row according to their variable, putting 0 if a variable is missing:

$$T(\mathbf{v}) = \begin{pmatrix} 3\,v_1 - 2\,v_2 \\ v_1 - 2\,v_2 + 3\,v_3 \\ v_2 + v_3 \\ 2\,v_1 + 3\,v_2 - v_3 \end{pmatrix} = \begin{pmatrix} 3\,v_1 \\ v_1 \\ 0 \\ 2\,v_1 \end{pmatrix} + \begin{pmatrix} -2\,v_2 \\ -2\,v_2 \\ v_2 \\ 3\,v_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 3\,v_3 \\ v_3 \\ -v_3 \end{pmatrix} = v_1 \begin{pmatrix} 3 \\ 1 \\ 0 \\ 2 \end{pmatrix} + v_2 \begin{pmatrix} -2 \\ -2 \\ 1 \\ 3 \end{pmatrix} + v_3 \begin{pmatrix} 0 \\ 3 \\ 1 \\ -1 \end{pmatrix}.$$

Therefore the function is given by the matrix

$$T = \begin{pmatrix} 3 & -2 & 0 \\ 1 & -2 & 3 \\ 0 & 1 & 1 \\ 2 & 3 & -1 \end{pmatrix}.$$

It turns out that all linear functions come from matrices. If T is a linear function there is a matrix A such that for all $\mathbf{x}$ we have $T(\mathbf{x}) = A\mathbf{x}$. We will prove this fundamental fact after proving an important feature of linear functions: they are determined by the values they take in a basis.

As we did with the definition of vector subspace we can combine the two properties that define a linear function into one.

THEOREM 3.1.1 (**Alternative definition of linear function**). *A function is linear if and only if it respects linear combinations. In other words, a function $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is linear if and only if for any k scalars $\lambda_1, \dots, \lambda_k$ and any k vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ we have:*

$$T(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_k \mathbf{v}_k) = \lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_k T(\mathbf{v}_k).$$

PROOF. Exercise. See the proof of Theorem 2.1.3 and proceed similarly. □

When checking if a function is linear we only need to check that it respects linear combinations of two vectors.

THEOREM 3.1.2 (**Alternative statement of Alternative definition of linear function**). *A function is linear if and only if it respects linear combinations of two vectors. In other words, a map $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is linear if and only if for every $\lambda, \mu \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ we have:*

$$T(\lambda \mathbf{v} + \mu \mathbf{w}) = \lambda T(\mathbf{v}) + \mu T(\mathbf{w}).$$

PROOF. Exercise. □

COROLLARY 2 (**Linear maps send zero to zero**). *Let $T\colon \mathbf{R}^n \to \mathbf{R}^m$ be a linear map. Then*

$$T\,\mathbf{0} = \mathbf{0}.$$

Equivalently,

$$T\,\mathbf{0} \neq \mathbf{0} \implies T \text{ is not linear.}$$

PROOF. We have

$$T\,\mathbf{0} = T\,(\mathbf{0} + \mathbf{0}) = T\,\mathbf{0} + T\,\mathbf{0}.$$

Subtracting $T\,\mathbf{0}$ from both sides of this equation yields the result. □

EXAMPLE 31. None of the following functions is linear:

$$\begin{aligned} f\colon \mathbb{R} \longrightarrow \mathbb{R}, &\quad x \longmapsto 2x - 3 \\ T\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^3, &\quad (x, y) \longmapsto (x - y + 2, 2x + 3y, 42x) \\ S\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2, &\quad (x, y, z) \longmapsto (2x - 3y + z, 42). \end{aligned}$$

A very useful consequence of Theorem 3.1.1 is that if we know the values of a linear function at a basis then we can compute its value at any vector. We illustrate this with an example.

EXAMPLE 32. For a linear function $T\colon \mathbb{R}^4 \longrightarrow \mathbb{R}$ we have

$$T\,\mathbf{e}_1 = -5, \quad T\,\mathbf{e}_2 = 3, \quad T\,\mathbf{e}_3 = 1, T\,\mathbf{e}_4 = -2.$$

Find $T(-2, 1, 3, 4)$.

SOLUTION. Let $\mathbf{v} = (-2, 1, 3, 4)$ then $\mathbf{v} = -2\,\mathbf{e}_1 + \mathbf{e}_2 + 3\,\mathbf{e}_3 + 4\,\mathbf{e}_4$. It follows that

$$\begin{aligned} T(\mathbf{v}) &= -2\,\mathbf{e}_1 + \mathbf{e}_2 + 3\,\mathbf{e}_3 + 4\,\mathbf{e}_4 \\ &= -2\,(-5) + 3 + 3 \cdot 1 + 4\,(-2) \\ &= -10 + 3 + 3 - 8 \\ &= -12. \end{aligned}$$

□

So if two linear functions agree on a basis they agree everywhere and are therefore equal.

THEOREM 3.1.3. *Let $T, S\colon \mathbb{R} \longrightarrow \mathbb{R}^m$ be linear functions and let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis of $\mathbb{R}^n$. If*

$$T\,\mathbf{v}_1 = S\,\mathbf{v}_1, \dots, T\,\mathbf{v}_n = S\,\mathbf{v}_n$$

then we have

$$\forall \mathbf{v} \in \mathbb{R}^n, T\,\mathbf{v} = S\,\mathbf{v},$$

in other words

$$T = S.$$

PROOF. Let $\mathbf{v} \in \mathbf{R}^n$ then there are unique $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ so that

$$\mathbf{v} = \lambda_1 \mathbf{v}_1 + \dots + \lambda_n \mathbf{v}_n.$$

Then we have

$$\begin{aligned} T\,\mathbf{v} &= \lambda_1\,T\,\mathbf{v}_1 + \dots + \lambda_n\,T\,\mathbf{v}_n \\ &= \lambda_1\,S\,\mathbf{v}_1 + \dots + \lambda_n\,S\,\mathbf{v}_n \\ &= S\,\mathbf{v}. \end{aligned}$$

□

Now, let's remember that the linear function defined by a matrix A with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$ is given by the formula

$$A\,\mathbf{x} = x_1\,\mathbf{a}_1 + \dots + x_n\,\mathbf{a}_n$$

where $\mathbf{x} = x_1\,\mathbf{e}_1 + \dots + x_n\,\mathbf{e}_n$. In particular we have that the columns of A are the images of the standard basis, in other words

$$\mathbf{a}_i = A\,\mathbf{e}_i, \quad i = 1, \dots, n.$$

As a consequence we have that any linear function is equal to the linear function that has columns the images of the standard basis under under that function. So we have the following theorem.

THEOREM 3.1.4. *Let $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear function. Then T is equal to the linear function associated with the matrix with columns $T\,\mathbf{e}_1, \dots, T\mathbf{e}_n$.*

EXAMPLE 33 (**The identity matrix**). For the identity function we have $I\,\mathbf{e}_n = \mathbf{e}_n$. Therefore the identity function is induced by the $n \times n$ matrix with columns $\mathbf{e}_1, \dots, \mathbf{e}_n$. That is,

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

EXAMPLE 34. Consider again the linear function $T\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^4$ given by

$$T(x,y,z) = (3x-2y, x-2y+3z, y+z, 2x+3y-z)$$

from Example 30. We have

$$\begin{aligned} T(1,0,0) &= (3,1,0,2) \\ T(0,1,0) &= (-2,-2,1,3) \\ T(0,0,1) &= (0,3,1,-1) \end{aligned}$$

and we again get that T is given by the matrix

$$T = \begin{pmatrix} 3 & -2 & 0 \\ 1 & -2 & 3 \\ 0 & 1 & 1 \\ 2 & 3 & -1 \end{pmatrix}.$$

So if we know the values of a linear transformation in the standard basis of $\mathbb{R}^n$ it's straightforward to find its matrix. What about other bases though? A linear transformation is uniquely determined by its values in *any* basis, is there a method to find the matrix if the basis is not the standard one?

Indeed there is! We illustrate with an example.

EXAMPLE 35. Consider the basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ of $\mathbb{R}^3$ where

$$\mathbf{v}_1 = (1,-1,0), \quad \mathbf{v}_2 = (0,2,-1), \quad \mathbf{v}_3 = (1,0,2).$$

For a 3×4 matrix A we have that

$$A\mathbf{v}_1 = (2,3,0,1), \quad A\mathbf{v}_2 = (1,1,1,1), \quad A\mathbf{v}_4 = (0,4,2,0).$$

Determine the matrix A.

SOLUTION. The solution has two steps. We first express the standard basis in terms of the new basis, and then we calculate the images of the standard basis i.e. the columns of A.

First Step: So we have to express each $\mathbf{e}_i$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. So we have to solve three systems

$$B\mathbf{x} = \mathbf{e}_i, \quad i = 1,2,3$$

where

$$B = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 2 & 0 \\ 0 & -1 & 2 \end{pmatrix}.$$

Rather than doing essentially the same calculations with three different augmented matrices we augment B with all three vectors at once. At the end of our calculations the first column of the augmented part will be the coefficients to express $\mathbf{e}_1$, the second $\mathbf{e}_2$, and the third $\mathbf{e}_3$ in terms of the basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ -1 & 2 & 0 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 1 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 1 & 1 & 0 \\ 0 & 0 & 5 & 1 & 1 & 2 \end{array}\right).$$

In the first step I added the first row to the second row. In the second step I added the second row to twice the third row. Next I'll add the third row to -5 times the second, and the first to get a diagonal matrix. The final step is then to divide each row by the the corresponding diagonal element.

$$\left(\begin{array}{ccc|ccc} -5 & 0 & 0 & -4 & 1 & 2 \\ 0 & -10 & 0 & -4 & -4 & 2 \\ 0 & 0 & 5 & 1 & 1 & 2 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 4/5 & -1/5 & -2/5 \\ 0 & 1 & 0 & 2/5 & 2/5 & -1/5 \\ 0 & 0 & 1 & 1/5 & 1/5 & 2/5 \end{array}\right).$$

So all three systems have been solved and we have

$$\begin{aligned} \mathbf{e}_1 &= \frac{4}{5}\mathbf{v}_1 + \frac{2}{5}\mathbf{v}_2 + \frac{1}{5}\mathbf{v}_3 \\ \mathbf{e}_2 &= -\frac{1}{5}\mathbf{v}_1 + \frac{2}{5}\mathbf{v}_2 + \frac{1}{5}\mathbf{v}_3 \\ \mathbf{e}_3 &= -\frac{2}{5}\mathbf{v}_1 - \frac{1}{5}\mathbf{v}_2 + \frac{2}{5}\mathbf{v}_3. \end{aligned}$$

Second Step: By the linearity of A we have

$$A\,\mathbf{e}_1 = \frac{4}{5}A\,\mathbf{v}_1 + \frac{2}{5}A\,\mathbf{v}_2 + \frac{1}{5}A\,\mathbf{v}_3.$$

Therefore,

$$A\,\mathbf{e}_1 = \frac{4}{5}\begin{pmatrix}2\\3\\0\\1\end{pmatrix} + \frac{2}{5}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} + \frac{1}{5}\begin{pmatrix}0\\4\\2\\0\end{pmatrix} = \begin{pmatrix}2\\18/5\\4/5\\6/5\end{pmatrix}$$

Entirely similar calculations[1] give

$$A\,\mathbf{e}_2 = \begin{pmatrix}0\\3/5\\4/5\\1/5\end{pmatrix}, \quad A\,\mathbf{e}_3 = \begin{pmatrix}-1\\1/5\\3/5\\-3/5\end{pmatrix}.$$

Therefore

$$A = \begin{pmatrix} 2 & 0 & -1 \\ 18/5 & 3/5 & 1/5 \\ 4/5 & 4/5 & 3/5 \\ 6/5 & 1/5 & -3/5 \end{pmatrix}.$$

□

EXAMPLE 36 (3×3 **Permutation matrices)**. There are 6 ways to order a set with 3 elements. For example for the set $\{1,2,3\}$ we have the following possibilities:

$$1\ 2\ 3, \quad 1\ 3\ 2, \quad 2\ 1\ 3, \quad 2\ 3\ 1, \quad 3\ 1\ 2, \quad 3\ 2\ 1.$$

Each of these orders is determines a *permutation* of $\{1,2,3\}$, i.e. a *one-to-one* and *onto* function $\sigma\colon \{1,2,3\} \longrightarrow \{1,2,3\}$, namely the function that maps i to the element that appears in the i-th position. So the third ordering is determined the function with values $\sigma(1) = 2, \sigma(2) = 1$, and $\sigma(3) = 3$. Conversely, a permutation σ gives the ordering $\sigma(1)\ \sigma(2)\ \sigma(3)$. For example the permutation with values $\sigma(1) = 2$, $\sigma(2) = 3$, and $\sigma(3) = 1$ determines the fourth ordering.

[1] Verify all calculations yourself.

Now, given any such permutation we can define a linear transformation $\mathbb{R}^3 \longrightarrow \mathbb{R}^3$, by permuting the standard basis accordingly. What I mean is the following: take for example the last ordering 3 2 1, that corresponds to the permutation $\sigma(1) = 3, \sigma(2) = 2$, $\sigma(3) = 1$, and set

$$T\,\mathbf{e}_1 = \mathbf{e}_3, \quad T\,\mathbf{e}_2 = \mathbf{e}_2, \quad T\,\mathbf{e}_3 = \mathbf{e}_1.$$

There is one and only one linear transformation that satisfies these conditions, namely (see Theorem 3.1.4), the linear transformation associated with the matrix that has columns (listed in order) $\mathbf{e}_3, \mathbf{e}_2, \mathbf{e}_1$.

In other words, for any ordering of $\{1,2,3\}$ we order the vectors of the standard basis the same way and then take the matrix with those columns. We obtain the following 3×3 matrices

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}, \quad \begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}, \quad \begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}, \quad \begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}, \quad \begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}, \quad \begin{pmatrix}0&0&1\\0&1&0\\1&0&0\end{pmatrix}.$$

The first of these *permutation matrices* is I_3 the 3×3 identity matrix, and is obtained by the identity permutation. Let's see what P_{132} does to a vector (x, y, z).

$$\begin{pmatrix}1&0&0\\0&0&1\\0&1&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}1\cdot x+0\cdot y+0\cdot z\\0\cdot x+0\cdot y+1\cdot z\\0\cdot x+1\cdot y+0\cdot z\end{pmatrix} = \begin{pmatrix}x\\z\\y\end{pmatrix}.$$

So $P_{132}(x, yz) = (x, z, y)$. Let's also find $P_{312}(x, y, z)$, where P_{312} is the permutation matrix that corresponds to the ordering 3 1 2.

$$\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}0\cdot x+1\cdot y+0\cdot z\\0\cdot x+0\cdot y+1\cdot z\\1\cdot x+0\cdot y+0\cdot z\end{pmatrix} = \begin{pmatrix}y\\z\\x\end{pmatrix}.$$

So $P_{312}(x, y, z) = (y, z, x)$. Notice that we looked at three permutation matrices P_{123}, P_{132}, and P_{312} and for all three, the image of (x, y, z) was a vector with coordinates some permutation of (x, y, z).

We will see later in the class that the rows of a permutation matrix are also given by a permutation of the rows of the identity matrix. Furthermore, if P is a permutation matrix and $\mathbf{x}$ a column vector, then the rows of $P\,\mathbf{x}$ are given from the columns of $\mathbf{x}$ by the same permutation.

We can verify that this is the case for the three permutation matrices we checked. The rows of I_3 are given by the identity permutation and the rows of I_3, and the coordinates of $I_3(x, y, z)$ are also given by the identity permutation of the coordinates of (x, y, z).

The rows of P_{132} are obtained by interchanging the second and third row of I_3, and in $P_1(x, y, z) = (x, z, y)$ the second and third coordinate of (x, y, z) are interchanged.

Finally, the first row of P_{132} is the second row of I_3 and the first coordinate of $P_{132}(x, y, z)$ is y. The second row of P_{132} is the third row of I_3 and the second coordinate of $P_{132}(x, y, z)$ is z. The third row of P_{132} is the first row of I_3 and the third coordinate of $P_{132}(x, y, z)$ is x.

Exercise 2. Compute $P(x, y, z)$ for the remaining three permutation matrices and verify that the coordinates are permuted the same way that the rows of P have been permuted.

3.1.1. New linear functions from old. Let $A, B \colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be two functions, and $\lambda \in \mathbb{R}$. We define a new function $A + B$, called the *sum* of A and B, and a new function $\lambda\,A$, called the *scalar product* of λ and A, as follows

$$A + B \colon \mathbb{R}^n \longrightarrow \mathbb{R}^m, \quad (A + B)\,\mathbf{x} = A\,\mathbf{x} + B\,\mathbf{x},$$

and

$$\lambda\, A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m, \quad (\lambda\, A)\ \mathbf{x} = \lambda\ (A\,\mathbf{x})\,.$$

In other words, to find the image of $\mathbf{x}$ under $A + B$ we add its images under A and B. To find the image of x under $\lambda\, A$ we multiply its image under A by λ.

We also define the *opposite of* A to be the function

$$-A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m, \quad (A)\ \mathbf{x} = -\,(A\,\mathbf{x})\,.$$

Clearly $-A = -1\, A$.

THEOREM 3.1.5 (**Function Spaces are Vector Spaces**). *Addition and scalar multiplication of functions satisfy all the axioms listed in Theorem 2.1.1, where the role of the zero vector is played by the zero function O. In other words we have the following properties:*

(a) *Function addition is* commutative. *This means that for any two functions A, B we have*

$$A + B = B + A$$

(b) *Function addition is* associative. *This means that for any three functions A, B, and C we have*

$$(A + B) + C = A + (B + C)\,.$$

(c) *O is* neutral *for addition. This means that for any function A we have*

$$O + A = A.$$

(d) *For every function A we have*

$$A + (-A) = O.$$

(e) *The number 1 is neutral for scalar multiplication. This means that for every function A we have*

$$1\, A = A.$$

(f) *Scalar multiplication distributes over function addition. This means that if λ is a scalar and A, B are functions we have*

$$\lambda\,(A + B) = \lambda\, A + \lambda\, B.$$

(g) *Addition of scalars distributes over scalar multiplication. This means that*

$$(\lambda + \mu)\, A = \lambda\, A + \mu\, A.$$

(h) *Multiplication of scalars and scalar multiplication are compatible in the following sense: if λ, μ are scalars and A is a function, we have*

$$\lambda\ (\mu\, A) = (\lambda\, \mu)\ A.$$

PROOF. To prove that two functions are equal we need to prove that they give the same result when applied to the same argument. For example to prove that addition is commutative we need to prove that for all $x \in \mathbb{R}^n$ we have

$$(A + B)\,\mathbf{x} = (B + A)\,\mathbf{x}.$$

Indeed,

$$\begin{aligned}(A + B)\,\mathbf{x} &= A\,\mathbf{x} + B\,\mathbf{x}\\ &= B\,\mathbf{x} + A\,\mathbf{x}\\ &= (B + A)\,\mathbf{x}.\end{aligned}$$

Let’s also prove that addition of scalars distributes over scalar multiplication.

Let $\mathbf{x} \in \mathbb{R}^n$, and $\lambda, \mu \in \mathbb{R}$. Then

$$\begin{aligned} ((\lambda + \mu)\, A)\, \mathbf{x} &= \lambda + \mu)\, A\, \mathbf{x} \\ &= \lambda\, (A\, \mathbf{x}) + \mu\, (A\, \mathbf{x}) \\ &= (\lambda\, A\,)\, \mathbf{x} + (\mu\, A)\, \mathbf{x} \\ &= ((\lambda + \mu)\;\, A)\, \mathbf{x}. \end{aligned}$$

The proofs of the remaining properties are similar and left as an exercise. □

We will visit these operations later in the class. At the moment we concentrate in the case of linear functions. We have the following two theorems.

THEOREM 3.1.6 (**The sum of two linear functions is a linear function**). *If A, B are linear functions then $A + B$ is also linear. Furthermore, if the matrix of A is (a_{ij}) and the matrix of B is (b_{ij}) then the matrix of $A + B$ is $(a_{ij} + b_{ij})$. In other words*

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

PROOF. Let $\lambda, \mu \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. We have,

$$\begin{aligned} (A + B)\, (\lambda\mathbf{x} + \mu\, \mathbf{y}) &= A\, (\lambda\mathbf{x} + \mu\, \mathbf{y}) + B\, (\lambda\mathbf{x} + \mu\, \mathbf{y}) \\ &= \lambda\, A\, \mathbf{x} + \mu\, A\, \mathbf{y} + \lambda\, B\, \mathbf{x} + \mu\, B\, \mathbf{y} \\ &= \lambda\, A\, \mathbf{x} + \lambda\, B\, \mathbf{x} + \mu\, B\, \mathbf{y} + \mu\, A\, \mathbf{y} \\ &= \lambda\; (A\, \mathbf{x} + B\, \mathbf{x}) + \mu\; (A\, \mathbf{y} + B\, \mathbf{y}) \\ &= \lambda\; ((A\; + B)\;\, \mathbf{x}) + \mu\; ((A\; + B)\;\, \mathbf{y}) \\ &= (\lambda\; (A\; + B))\;\, \mathbf{x} + (\mu\; (A\; + B))\;\, \mathbf{y}. \end{aligned}$$

Therefore $A + B$ is linear. Now, recall that the j-th column of the matrix of $A + B$ is $(A + B)\, \mathbf{e}_j$, but by the definition of $A + B$ we have

$$(A + B)\, \mathbf{e}_j = A\, \mathbf{e}_j + B\, \mathbf{e}_j.$$

Therefore the j-th column of $A + B$ is the sum of the j-th column of A and the j-th column of B. □

THEOREM 3.1.7 (**A multiple of a linear function is a linear function**). *If A is a linear function then $\lambda\, A$ is also linear for every $\lambda \in \mathbb{R}$. Furthermore, if the matrix of A is (a_{ij}) then the matrix of $\lambda\, A$ is $(\lambda\, a_{ij})$. In other words,*

$$\lambda \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} \lambda\, a_{11} & \lambda\, a_{12} & \cdots & \lambda\, a_{1n} \\ \lambda\, a_{21} & \lambda\, a_{22} & \cdots & \lambda\, a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda\, a_{m1} & \lambda\, a_{m2} & \cdots & \lambda\, a_{mn} \end{pmatrix}.$$

The proof is similar to the proof of Theorem 3.1.6 and is left as an exercise.

We can combine Theorems 3.1.6 and 3.1.7 into a single theorem.

THEOREM 3.1.8 (**Linear combinations of linear functions are linear**). *If A, B are linear function and λ, μ are scalars then $\lambda\, A + \mu\, B$ is linear with matrix $(\lambda\, a_{ij} + \mu\, b_{ij})$.*

EXAMPLE 37. We have

$$3\begin{pmatrix}2 & 1 & 0\\ -1 & 3 & 1\end{pmatrix} - 4\begin{pmatrix}3 & 0 & 1\\ 2 & -1 & 0\end{pmatrix} = \begin{pmatrix}-6 & 3 & -4\\ -11 & 13 & 3\end{pmatrix}.$$

We next look at the operation of composition. Recall that if $g\colon X \longrightarrow Y$ and $f\colon Y \longrightarrow Z$ are two functions then the composition $f \circ g$ is defined as follows:

$$f \circ g\colon X \longrightarrow Z, \quad (f \circ g)(x) = f(g(x)).$$

Let A be an $m \times n$ and B an $n \times k$ matrix. Then

$$A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m, \quad B\colon \mathbb{R}^k \longrightarrow \mathbb{R}^n$$

and so the composition

$$A \circ B\colon \mathbb{R}^k \to \mathbb{R}^m, \qquad \mathbf{x} \longmapsto A(B\mathbf{x}),$$

is defined. We write AB instead of $A \circ B$.

THEOREM 3.1.9 (**Composition of linear maps is linear**). *If A, B are linear maps such that the composition AB is defined, then AB is linear map.*

PROOF. Let $\lambda, \mu \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. We have,

$$\begin{aligned}(AB)(\lambda\mathbf{x} + \mu\mathbf{y}) &= A(B(\lambda\mathbf{x} + \mu\mathbf{y}))\\ &= A(\lambda B\mathbf{x} + \mu B\mathbf{y})\\ &= \lambda A(B\mathbf{x}) + \mu A(B\mathbf{y})\\ &= \lambda((AB)\mathbf{x}) + \mu((AB)\mathbf{y}).\end{aligned}$$

□

We want to find a formula for the matrix of AB. Let's first do this for matrices of relatively low dimensions. Let's take A to be 3×2 and B to be 2×2. Then the composition AB is given by a 3×2 matrix. We want to find

$$A(B\mathbf{x}) = \begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\\ a_{31} & a_{32}\end{pmatrix}\left(\begin{pmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}\right).$$

We know that the image of $\mathbf{x}$ is a linear combination of the column vectors of B with coefficients given by the coordinates of $\mathbf{x}$. We have then, using the linearity of A,

$$A(B\mathbf{x}) = A\left(x_1\begin{pmatrix}b_{12}\\ b_{21}\end{pmatrix} + x_2\begin{pmatrix}b_{21}\\ b_{22}\end{pmatrix}\right) = x_1 A\begin{pmatrix}b_{12}\\ b_{21}\end{pmatrix} + x_2 A\begin{pmatrix}b_{21}\\ b_{22}\end{pmatrix}.$$

Using again the fact that the image of a vector under A is a linear combination of the columns of A with coefficients the coordinates of the vector we have

$$A\begin{pmatrix}b_{11}\\ b_{21}\end{pmatrix} = b_{11}\begin{pmatrix}a_{11}\\ a_{21}\\ a_{31}\end{pmatrix} + b_{21}\begin{pmatrix}a_{11}\\ a_{21}\\ a_{31}\end{pmatrix} = \begin{pmatrix}a_{11}b_{11}\\ a_{21}b_{11}\\ a_{31}b_{11}\end{pmatrix} + \begin{pmatrix}a_{12}b_{21}\\ a_{22}b_{21}\\ a_{32}b_{21}\end{pmatrix} = \begin{pmatrix}a_{11}b_{11} + a_{12}b_{21}\\ a_{21}b_{11} + a_{22}b_{21}\\ a_{31}b_{11} + a_{32}b_{21}\end{pmatrix},$$

and, by entirely similar calculations

$$A\begin{pmatrix}b_{12}\\ b_{22}\end{pmatrix} = \begin{pmatrix}a_{11}b_{12} + a_{12}b_{22}\\ a_{21}b_{12} + a_{22}b_{22}\\ a_{31}b_{12} + a_{32}b_{22}\end{pmatrix}.$$

So,

$$(A\,B)\;\mathbf{x} = x_1 \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} \\ a_{21}b_{11} + a_{22}b_{21} \\ a_{31}b_{11} + a_{32}b_{21} \end{pmatrix} + x_2 \begin{pmatrix} a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{12} + a_{32}b_{22} \end{pmatrix}.$$

Keeping the LHS, the linear combination of columns in the RHS can be expressed as a product of a 2×2 matrix and a $\mathbf{x}$. Therefore we have,

$$(A\,B)\;\mathbf{x} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

So we got that,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{pmatrix}. \tag{3.5}$$

The same ideas can be used to get the formula for the matrix of $A\,B$ in the general case where A has dimensions $m \times k$ for $m, k \geq 1$ and B has dimensions $k \times n$[2] for $n \geq 1$.

Let $\mathbf{a}_i^*$ be the i-th *row* and $\mathbf{b}_j$ the j-th *column* of B. That is we consider A as a column of m row vectors, each of dimension k, while B is considered as a row of n column vectors each of dimension k. Let us also set $C = A\,B$, an $m \times n$ matrix.

The j-th column of C is $C\,\mathbf{e}_j$. But,

$$C\,\mathbf{e}_j = A\,(B\,\mathbf{e}_j) = A\,\mathbf{b}_j.$$

Therefore, by the boxed formula at the bottom of Page 50 the i-th element of the j-th column of C is the "dot product" of the i-th row of A with the j-column of B.

We have thus proved the following theorem.

Theorem 3.1.10. *Let A be an $m \times k$ matrix and B a $k \times n$ matrix. Then the entries of $C = A\,B$ are given by*

$$c_{ij} = \mathbf{a}_i^* \cdot \mathbf{b}_j = \sum_{\ell=1}^{k} a_{i\ell}\, a_{\ell j}. \tag{3.6}$$

Or, if we expand the sum in the RHS,

$$c_{ij} = a_{i1}\, b_{1j} + \cdots + a_{ik}\, b_{kj}.$$

We can express Equation (3.6) as follows:

$$\begin{pmatrix} \mathbf{a}_1^* \\ \vdots \\ \mathbf{a}_m^* \end{pmatrix} \begin{pmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1^* \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_1^* \cdot \mathbf{b}_n \\ \vdots & \ddots & \vdots \\ \mathbf{a}_m^* \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_m^* \cdot \mathbf{b}_n \end{pmatrix}.$$

Example 38. Let's compute $A\,B$ and $B\,A$ where

$$A = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 0 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} -1 & 1 \\ 2 & 0 \\ 4 & 3 \end{pmatrix}.$$

[2] This is the same k. In order for the matrices to be composable the number of rows of B has to equal the number of columns of A.

We have

$$AB = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 0 & 3 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 2 & 0 \\ 4 & 3 \end{pmatrix} = \begin{pmatrix} -1+4+0 & 1+0+0 \\ 1+0+12 & -1+0+9 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 13 & 8 \end{pmatrix},$$

while

$$BA = \begin{pmatrix} -1 & 1 \\ 2 & 0 \\ 4 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ -1 & 0 & 3 \end{pmatrix} = \begin{pmatrix} -1-1 & -2+0 & 0+3 \\ 2+0 & 4+0 & 0+0 \\ 4-3 & 8+0 & 0+9 \end{pmatrix} = \begin{pmatrix} -2 & -2 & 3 \\ 2 & 4 & 0 \\ 1 & 8 & 9 \end{pmatrix}.$$

3.1.2. Some Exercises.

Exercise 3. Let $P\colon \mathbb{R}^4 \longrightarrow \mathbb{R}^4$ be given by

$$P(x, y, z, w) = (y, z, w, x).$$

(a) Prove that P is linear using Theorem 3.1.2.
(b) Find the matrix of P.
(c) Verify that the rows of the matrix of P, listed in order, are $\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_1$.

Exercise 4. An 1×1 matrix has only one entry (a). It's natural to identify this matrix with the real number a. We have also identified $\mathbb{R}^1$ with $\mathbb{R}$. So the linear function defined by a 1×1 matrix has domain and codomain $\mathbb{R}$.

(a) What functions $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ arise as the linear functions associated with 1×1 matrices?
(b) When is a linear function $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ one-to-one?
(c) When is a linear function $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ onto?
(d) When is a linear function $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ invertible?
(e) Let $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ be an invertible linear function. What is f^{-1}?
(f) Let f, g be two linear functions $\mathbb{R} \longrightarrow \mathbb{R}$. Prove that $f \circ g$ is also linear. Give the matrix $f \circ g$ in terms of the matrices of f and g.

Your answers to Questions (2) through (5) should be in terms of the matrices that define the linear functions.

Exercise 5. For each of the following functions T

(a) Prove that it is linear.
(b) Find the matrix that gives T.
(a) The function $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ given by
$$T\mathbf{x} = \lambda \mathbf{x}.$$
(b) The function $T\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ given by
$$T(x_1, x_2, x_3) = (x_1, x_2).$$
(c) The function $T\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ given by
$$T(x_1, x_2) = (x_1, x_2, 0).$$
(d) The function $T\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ given by
$$T(x, y, z) = (x + z, x - z).$$

Exercise 6. Let

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

(a) Prove that the columns of the matrix form a basis of $\mathbb{R}^3$.

(b) Let T be the linear map that interchanges the columns and rows of A. In other words

$$T\,\mathbf{a}_1 = \mathbf{a}_1^*, \quad ,T\,\mathbf{a}_2 = \mathbf{a}_2^*, \quad T\,\mathbf{a}_3 = \mathbf{a}_3^*,$$

where $\mathbf{a}_i$ (respectively $\mathbf{a}_i^*$) are the columns (respectively rows) of A. Explain why T is well defined. Then find the matrix of T.

(c) Find $T\,\mathbf{a}_i^*$, for $i = 1, 2, 3$.

3.2. Range and rank, Kernel and nullity

Let's introduce some terminology and recall some concepts about functions. A function f with *domain* X and *codomain* Y associates to every $x \in X$ *unique* element $y \in Y$, denoted by $f(x)$. We also use the notation $x \mapsto y$ to indicate that $y = f(x)$. The notation,

$$f\colon X \longrightarrow Y$$

means that f is a function with domain X and codomain Y.

DEFINITION 15 (**Range, Image, Preimage**). The set of all elements of Y that are images of elements of X is called the *range* of f and denoted by $\mathcal{R}(f)$. Thus

$$\begin{aligned}\mathcal{R}(f) &= \{f(x) : x \in X\}\\ &= \{y \in Y : \exists x \in X, \quad x \longmapsto y\}.\end{aligned}$$

If $S \subseteq X$ then the *image of S under f*, denoted $f(S)$ is the set of the images of all elements of S. Thus

$$f(S) = \{f(s) : s \in S\} \subseteq Y.$$

Note that $f(X) = \mathcal{R}(f)$.

If $T \subseteq Y$ then the *preimage of T under f*, denoted $f^{-1}(T)$ is the set of all elements of X that are mapped to an element of T. Thus

$$f^{-1}(T) = \{x \in X : f(x) \in T\}.$$

Consider now a linear function with matrix A

$$A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m, \quad \mathbf{x} \longmapsto A\mathbf{x}.$$

What is the range of A? The definition says that

$$\mathcal{R}(A) = \{\mathbf{y} \in \mathbb{R}^m : \exists \mathbf{x} \in \mathbb{R}^n, \quad A\mathbf{x} = \mathbf{y}\},$$

so $\mathbf{y} \in \mathcal{R}(A)$ if and only if the system

$$A\mathbf{x} = \mathbf{y}$$

has solutions. Now if $\mathbf{x} = (x_1, \dots, x_n)$ is such a solution then

$$x_1\,\mathbf{a}_1 + \dots + x_n\,\mathbf{a}_n = \mathbf{y},$$

where $\mathbf{a}_1, \dots, \mathbf{a}_n$ are the columns of A. Thus the range of A is the linear span of its columns. So we have the following theorem.

THEOREM 3.2.1 (**Range is the span of the columns**). *The range of the linear map with matrix A is the linear span of the the columns of A. In other words, if $\mathbf{a}_1, \dots, \mathbf{a}_n$ are the columns of A, then*

$$\mathcal{R}(A) = \langle \mathbf{a}_1, \dots, \mathbf{a}_n \rangle.$$

DEFINITION 16 (**Rank of a matrix**). The *rank* of a linear map is the dimension of its range. The rank of a linear map A is denoted by $\operatorname{rank} A$. Thus

$$\operatorname{rank} A = \dim \mathcal{R}(A).$$

We can summarize the discussion in Section 2.2.1 as follows.

THEOREM 3.2.2. *The basic columns of an $m \times n$ matrix form a basis of $\mathcal{R}(A)$. Therefore* $\operatorname{rank} A$ *is the number of columns in the reduced echelon form of A that contain a leading* 1.

EXAMPLE 39. Find the rank of the following matrix

$$A = \begin{pmatrix} 1 & 3 & -2 & 5 & 4 \\ 1 & 4 & 1 & 3 & 5 \\ 1 & 4 & 2 & 4 & 3 \\ 2 & 7 & -3 & 6 & 13 \end{pmatrix}$$

SOLUTION. The reduced row echelon form of A is[3]

$$A \sim \begin{pmatrix} 1 & 0 & 0 & 22 & -21 \\ 0 & 1 & 0 & -5 & 7 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

There are three basic columns and therefore $\operatorname{rank} A = 3$. □

If $\mathbf{c} \in \mathbb{R}^m$ then the solution set of the linear system $A\mathbf{x} = \mathbf{c}$ is the preimage of $A^{-1}\{\{\mathbf{c}\}\}$. In particular the preimage of the zero vector is the solution set of the homogeneous system $A\mathbf{x} = \mathbf{0}$.

Recall (see Theorem 1.2.6 in Section 1.2.2) that the solution sets of homogeneous systems are subspaces of $\mathbb{R}^m$.

DEFINITION 17 (**Kernel and nullity**). The *kernel* (or *null space*) of A, denoted by $\ker A$ is the preimage of the zero vector. Thus

$$\ker A = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}.$$

The dimension of $\ker A$ is called the *nullity* of A and is denoted by $\operatorname{null} A$. Thus

$$\operatorname{null} A = \dim(\ker A).$$

Throughout Section 1.1 we were referring to the number of free parameters in the solution of a system has as the "dimension" of the solution set. This suggests that the nullity of a matrix is the number of free columns in its reduced echelon form. This is indeed the case. Consider for example the homogeneous system with matrix the matrix A of Example 39. The solution of $A\mathbf{x} = \mathbf{0}$ is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -22s + 21t \\ 5s - 7t \\ -s + 2t \\ s \\ t \end{pmatrix} = s\begin{pmatrix} -22 \\ 5 \\ -1 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} 21 \\ -7 \\ 2 \\ 0 \\ 1 \end{pmatrix}, \quad s, t \in \mathbb{R}. \tag{3.7}$$

Thus,letting $\mathbf{s} = (-22, 5, -1, 1, 0)$ and $\mathbf{t} = (21, -7, 2, 0, 1)$ we have

$$\ker A = \langle \mathbf{s}, \mathbf{t} \rangle.$$

Now the set $\{\mathbf{s}, \mathbf{t}\}$ is linearly independent. This follows immediately from the fact that the if the second vector in (3.7) is $\mathbf{0}$ then $s = t = 0$. Therefore $\{\mathbf{s}, \mathbf{t}\}$ is a basis of $\ker A$, and so $\ker A$ is two-dimensional.

Notice that the first three coordinates of $\mathbf{s}$ form the opposite of fourth column of the reduced echelon form, and the last two extra coordinates are the coordinates of $\mathbf{e}_1$. Similarly the first three coordinates $\mathbf{t}$ are the opposites of the fifth column and its last two coordinates those of $\mathbf{e}_2$.

[3] Do the calculations yourself!

A similar pattern will arise always. In order not to get tangled in over complicated notations let's consider the case that the free variables are the third, fourth, and seventh. Then the parametric solution will be

$$\begin{cases} x_1 &= -s\,b_{13} - t\,b_{14} - w\,b_{17} \\ x_2 &= -s\,b_{23} - t\,b_{24} - w\,b_{27} \\ x_3 &= s \\ x_4 &= t \\ x_5 &= -s\,b_{53} - t\,b_{54} - w\,b_{57} \\ x_6 &= -s\,b_{63} - t\,b_{64} - w\,b_{67} \\ x_7 &= w \end{cases}$$

Then in vector form the solution is

$$\mathbf{x} = s\,\mathbf{b}_3' + t\,\mathbf{b}_4' + w\,\mathbf{b}_7',$$

where

$$\begin{aligned} \mathbf{b}_3' &= (-b_{13}, -b_{23}, 1, 0, -b_{53}, -b_{63}, 0) \\ \mathbf{b}_4' &= (-b_{14}, -b_{24}, 0, 1, -b_{54}, -b_{64}, 0) \\ \mathbf{b}_7' &= (-b_{17}, -b_{27}, 0, 0, -b_{57}, -b_{67}, 1). \end{aligned}$$

The set $B = \{\mathbf{b}_3', \mathbf{b}_4', \mathbf{b}_7'\}$ is linearly independent. We can see this by looking only at the free slots, namely the third, fourth and seventh: we have the coordinates of $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Thus a linear dependency on B would give a linear dependency on the standard basis of $\mathbb{R}^3$, and that's not possible.

So we have the following theorem, that we will see again in a more general and precisely stated form later in the course.

THEOREM 3.2.3. *Let A be an $m \times n$ matrix and with reduced row echelon form B. The nullity of a A is the number of free columns B. Furthermore, a basis of* $\ker A$ *is obtained from the free columns of B by "interpolating" the coordinates of the standard basic vectors at the "free slots".*

As a corollary of Theorems 3.2.2 and 3.2.3 we have the following theorem.

THEOREM 3.2.4 (**Rank-nullity Theorem**). *If A in an $m \times n$ matrix then*

$$\operatorname{rank} A + \operatorname{null} A = n.$$

Bases of Range and Kernel

The basic columns in an echelon form of the matrix of a linear map give a basis for its range, and the free columns give a basis for its kernel.

EXAMPLE 40. Consider the linear function $T\colon \mathbb{R}^4 \longrightarrow \mathbb{R}^2$ with matrix

$$T = \begin{pmatrix} 1 & 2 & 3 & 4 \\ -1 & -2 & -3 & 4 \end{pmatrix}.$$

The reduced echelon form of T is

$$B = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The basic columns of B are the first and fourth. So the first and fourth column of T give a basis for the range of T. So,

$$\mathcal{R}(T) = \langle (1,-1),(4,4) \rangle .$$

The second and third columns of B will give a basis of $\ker T$. We are missing two coordinates to make the (opposites of the) second and third columns of B four dimensional and we fill those with the coordinates of $(1,0)$ and $(0,1)$ interpolated at the second and third slot. Thus the second column of B gives the vector $(2,1,0,0)$ and the third the vector $(3,0,1,0)$. Thus

$$\mathcal{R}(T) = \langle (-2,1,0,0),(-3,0,1,0) \rangle .$$

EXAMPLE 41. Consider the linear function $A\colon \mathbb{R}^5 \longrightarrow \mathbb{R}^4$ with matrix

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

The reduced echelon form of A is

$$A \sim \begin{pmatrix} 1 & 0 & 0 & -2 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

We have four basic columns and one free. Thus the range is four-dimensional and the first, second, third and fifth columns of A form a basis of $\mathcal{R}(A)$.

The kernel is one dimensional so a basis will have only one vector. We obtain that vector from the the opposite of the fourth column, by inserting 1 in the fourth slot. Thus

$$\ker A = \langle (2,-1,-1,1,0) \rangle .$$

3.3. Injective, Surjective, and Invertible linear maps

Recall that we say that a function is *one-to-one* or *injective* if it the images of two different elements is different. Thus a function $f\colon X \longrightarrow Y$ is one-to-one if for all $x_1, x_2 \in X$ we have

$$x_1 \neq x_2 \implies f(x_1) \neq f(x_2).$$

The contra-positive of the above, namely

$$f(x_1) = f(x_2) \implies x_1 = x_2$$

is often useful in proving (or disproving) that a function is injective. So if f is injective and $y \in Y$, then there can be at most one $x \in X$ with $f(x) = y$. We can express this in terms of preimages by saying that f is injective if and only if $f^{-1}(\{y\})$ contains at most one element.

f is called *onto* or *surjective* if every $y \in Y$ is the image of some element in X, i.e. if the range of f is Y.

Solutions of $y = f(x)$

Consider the equation

$$y = f(x). \tag{3.8}$$

(a) A function $f\colon X \longrightarrow Y$ is *injective* if for all $y \in Y$ Equation (3.8) has *at most* one solution.
(b) A function $f\colon X \longrightarrow Y$ is *surjective* if for all $y \in Y$ Equation (3.8) has *at least* one solution.
(c) A function is a *bijection*, i.e. both injective and surjective, if and only if for all $y \in Y$ Equation (3.8) has a unique solution.
(d) If f is a bijection then f has *inverse function* $f^{-1}\colon Y \longrightarrow X$ defined so that

$$x = f^{-1}(y) \iff y = f(x).$$

in other words, $f^{-1}(y)$ is the unique solution of Equation (3.8).

Recall also that if f is invertible then f^{-1} is also invertible and $\left(f^{-1}\right)^{-1} = f$. Finally recall that a pair of inverse functions is characterized by the equations

$$f\left(f^{-1}(y)\right) = y, \quad f^{-1}\left(f(x)\right) = x$$

or equivalently,

$$f \circ f^{-1} = I_Y, \quad f^{-1} \circ f = I_X.$$

Let now

$$A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

be a linear map. Then the equation

$$\mathbf{y} = A\mathbf{x}$$

is a system of m linear equations and n variables, and the nature of the solution set is determined by the reduced echelon form of A. The following theorem summarizes most of what we have seen so far in this class.

THEOREM 3.3.1. *Let A be an $m \times n$ and as usual denote the linear function it defines by the same symbol*

$$A\colon \mathbb{R}^n \to \mathbb{R}^m, \quad \mathbf{x} \mapsto A\mathbf{x}.$$

Let B be the reduced echelon form of A. The following hold.

(a) *A is injective if and only if B has no free columns.*
(b) *A is injective if and only if its kernel contains only the zero vector, i.e.*

$$\ker A = \{\mathbf{0}\}$$

.
(c) *A is injective if and only if* $\operatorname{null} A = 0$.
(d) *If A is injective then $n \leq m$.*
(e) *A is surjective if and only if its columns span $\mathbb{R}^m$.*
(f) *A is surjective if and only if* $\operatorname{rank} A = m$.
(g) *A is surjective if and only if B has m basic columns.*
(h) *If A is surjective then $n \geq m$.*
(i) *A is invertible if and only if $B = I_n$, the $n \times n$ identity matrix.*
(j) *A is invertible if and only if the columns of A form a basis of $\mathbb{R}^n$.*
(k) *If A is invertible then A^{-1} is linear.*
(l) *If A is invertible then $n = m$.*

Proof. Most of the statements are reformulations of things we have already proved. Try to understand why this is the case for each of the statements. I provide some hints to guide you.

(a) This just says that a consistent system has a unique solution if and only if there are no free variables.
(b) This just says that a consistent system has unique solution if and only if the corresponding homogeneous system has only the trivial solution.
(c) The nullity of A is the dimension of its kernel. A subspace has 0 dimension if and only if it equals $\mathbf{0}$.
(d) If there are more unknowns than equations then B will contain free columns. Conversely if there are no free columns the solution has no free variables, thus if it exists it is unique.
(e) The range of A is the linear span of the columns of A. A is surjective if and only if the range of A is $\mathbb{R}^m$.
(f) The rank of A is the dimension of its range. If a subspace of $\mathbb{R}^m$ has dimension m then it is the whole $\mathbb{R}^m$.
(g) The basic columns form a basis of the range of A, so if A is surjective then B has at least m basic columns. The leading ones have to be in different columns and therefore B cannot have more than m free columns.
(h) If there are less variables than equations there are not enough columns to form a basis of $\mathbb{R}^m$.
(i) A is invertible if an only if the system $A\mathbf{x} = \mathbf{c}$ has a unique solution, for all $\mathbf{c} \in \mathbb{R}^m$.
(j) If the system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution then the columns of A are linearly independent. If $A\mathbf{x} = \mathbf{y}$ is consistent for all $\mathbf{y}$ then the columns of A are spanning.
(k) We have

$$\begin{aligned} A\left(\lambda A^{-1}\mathbf{x} + \mu A^{-1}\mathbf{y}\right) &= \lambda A\left(A^{-1}\mathbf{x}\right) + \mu\left(A^{-1}\mathbf{y}\right) \\ &= \lambda\mathbf{x} + \mu\mathbf{y}. \end{aligned}$$

So

$$A^{-1}\left(A\left(\lambda A^{-1}\mathbf{x} + \mu A^{-1}\mathbf{y}\right)\right) = A^{-1}\left(\lambda\mathbf{x} + \mu\mathbf{y}\right).$$

Therefore

$$\lambda A^{-1}\mathbf{x} + \mu A^{-1}\mathbf{y} = A^{-1}\left(\lambda\mathbf{x} + \mu\mathbf{y}\right).$$

(l) It follows from (4) and (8).

□

As a consequence of the Rank-Nullity Theorem (see Theorem 3.2.4), we have the following Theorem.

THEOREM 3.3.2. *Let $A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a linear map. Then the following are equivalent.*

(a) A is injective.
(b) A is surjective.
(c) A is invertible.

PROOF. We will prove that

$$(1) \implies (2) \implies (3) \implies (1).$$

If A is injective then $\operatorname{null} A = 0$ and therefore by Theorem 3.2.4 we have $\operatorname{rank} A = n$. Thus A is surjective.

If A is surjective then $\operatorname{rank} A = n$ and therefore, again by Theorem 3.2.4 we have $\operatorname{null} A = 0$, thus A is also injective. A is therefore invertible.

If A is invertible then, by definition it is injective. □

REMARK 10. We remark that this property is not shared by general maps. If X is an infinite set there are always functions $X \longrightarrow X$ that are injective but not surjective, and functions that are surjective but not injective. For example for the set of natural numbers $\mathbb{N}$ the function

$$f\colon \mathbb{N} \longrightarrow \mathbb{N}, \quad f(n) = 2n,$$

is injective but not surjective. On the other hand,

$$g\colon \mathbb{N} \longrightarrow \mathbb{N}, \quad g(n) = \begin{cases} n/2 & n \text{ even} \\ n & n \text{ odd}, \end{cases}$$

is surjective but not injective.

THEOREM 3.3.3 (**Solving matrix equations**). *If A be an invertible $n \times n$ matrix, and C an $n \times k$ matrix for some positive integer k. Then the equation*

$$A\,X = C$$

has a unique solution, namely the $n \times k$ matrix

$$X = A^{-1}\,C.$$

Similarly, the equation

$$X\,A = C$$

has a unique solution, namely the $n \times k$ matrix

$$X = C\,A^{-1}.$$

REMARK 11. Because composition of functions is generally not commutative, we need to be careful to multiply in the right order.

PROOF. We have:

$$\begin{aligned} A\,X = C &\iff A^{-1}\,(A\,X) = A^{-1}\,C \\ &\iff \left(A^{-1}\,A\right)\,X = A^{-1}\,C \\ &\iff I\,X = A^{-1}\,C \\ &\iff X = A^{-1}\,C. \end{aligned}$$

The proof for the other equation is entirely similar. Just multiply from the right with A^{-1}[4]. □

[4] You should do it!

EXAMPLE 42 (**How to find the inverse of a linear function)**. Let A be an invertible linear function. Then the columns of (the matrix of) A^{-1} are the images of the vectors of the standard basis. That is the j-th column $\mathbf{c}_j$ of A_{-1} is given by

$$\mathbf{c}_j = A^{-1}\mathbf{e}_j,$$

or equivalently,

$$A\,\mathbf{c}_j = \mathbf{e}_j.$$

We solve all of these systems simultaneously by finding the reduced echelon form of the augmented matrix

$$\begin{pmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n & | & \mathbf{e}_1 & \dots & \mathbf{e}_n \end{pmatrix}.$$

So to find the inverse of

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{pmatrix}$$

we proceed as follows.

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 2 & -1 & 3 & 0 & 1 & 0 \\ 4 & 1 & 8 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & -1 & -1 & -2 & 1 & 0 \\ 0 & 1 & 0 & -4 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & -1 & -6 & 1 & 1 \\ 0 & 1 & 0 & -4 & 0 & 1 \end{array}\right)$$
$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & -4 & 0 & 1 \\ 0 & 0 & -1 & -6 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -11 & 2 & 2 \\ 0 & 1 & 0 & -4 & 0 & 1 \\ 0 & 0 & 1 & 6 & -1 & -1 \end{array}\right).$$

Therefore

$$A^{-1} = \begin{pmatrix} -11 & 2 & 2 \\ -4 & 0 & 1 \\ 6 & -1 & -1 \end{pmatrix}.$$

How to find the inverse a matrix

If A is an invertible $n \times n$ matrix then the reduced row echelon form of the block matrix

$$\begin{pmatrix} A & | & I \end{pmatrix} \sim \begin{pmatrix} I & | & A^{-1} \end{pmatrix}.$$

EXAMPLE 43 (2×2 **revisited)**. Let's consider again a 2×2 matrix

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

When is A invertible?

We start with the augmented matrix

$$\left(\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right).$$

If both a, c are 0 then the columns are not linearly independent and thus A is not invertible. Assume then that $a \neq 0$. We add to $-c$ times the first row to a times the second and we get

$$\left(\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & ad - bc & -c & a \end{array}\right).$$

If the determinant $D := ad - bc = 0$ then A is not invertible because $a \neq 0$ and thus the system $A\mathbf{x} = \mathbf{e}_2$ has no solutions.

If $D \neq 0$ we divide the second row by D,

$$\left(\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & 1 & -c/D & a/D \end{array}\right) \sim \left(\begin{array}{cc|cc} a & 0 & 1 + bc/D & -ab/D \\ 0 & 1 & -c/D & a/D \end{array}\right)$$

$$= \left(\begin{array}{cc|cc} a & 0 & ad/D & -ab/D \\ 0 & 1 & -c/D & a/D \end{array}\right) \sim \frac{1}{ad - bc} \left(\begin{array}{cc|cc} 1 & 0 & d & -b \\ 0 & 1 & -c & a \end{array}\right).$$

So when $a \neq 0$ and $D \neq 0$ we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - dc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{3.9}$$

If $a = 0$ and $c \neq 0$ we interchange the rows, and divide the first row by c and the second by b. and then add d times the second row to $-b$ times the first.

$$\left(\begin{array}{cc|cc} c & d & 0 & 1 \\ 0 & b & 1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & d/c & 0 & 1/c \\ 0 & 1 & 1/b & 0 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 0 & -d/bc & 1/c \\ 0 & 1 & 1/b & 0 \end{array}\right) = \frac{1}{-bc} \left(\begin{array}{cc|cc} 1 & 0 & d & -b \\ 0 & 1 & -c & 0 \end{array}\right).$$

Thus Equation (3.9) holds in all cases.

Inverse of a 2×2 matrix

Let $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, and $D = a_{11}a_{22} - a_{12}a_{21}$. A is invertible if and only if $D \neq 0$. Then

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \frac{1}{D} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}. \tag{3.10}$$

Consider now a 2×2 system of linear equations:

$$\begin{cases} a_{11}\,x_1 + a_{12}x_2 &= c_1 \\ a_{12}\,x_1 + a_{22}x_2 &= c_2. \end{cases}.$$

If A is invertible then we have (see Theorem 3.3.3)

$$\begin{aligned} A\mathbf{x} = \mathbf{c} &\iff A^{-1}\left(A\mathbf{x}\right) = A^{-1}\mathbf{c} \\ &\iff \left(A^{-1}A\right)\mathbf{x} = A^{-1}\mathbf{c} \\ &\iff I\,\mathbf{x} = A^{-1}\mathbf{c} \\ &\iff \mathbf{x} = A^{-1}\mathbf{c}. \end{aligned}$$

Thus, we can recover Crammer's rule (see Section 1.3). Indeed, we have that the solution of the system is

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{D} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \frac{1}{D} \begin{pmatrix} c_1a_{22} - c_2a_{12} \\ -c_1a_{21} + c_2a_{11} \end{pmatrix}.$$

3.4. The algebra of matrices

In the previous couple of lectures we studied the linear functions induced by matrices. We now are going to study matrices as algebraic objects of their own right.

If m, n are positive integers we denote by $\mathbf{M}_{m\times n}$the set of all $m \times n$ matrices. The set of $n \times n$ matrices is simply denoted by $\mathbf{M}_n$ and its elements are called *square matrices of size* n. As we have already done, if a matrix is denoted by a capital letter, say X, then the entry at the i-th row and j-th column will be denoted by x_{ij}, and we write $X = (x_{ij})$.

REMARK 12. Be careful to distinguish the notations (a_{ij}) and a_{ij}. The former denotes a matrix while the latter denotes an entry of that matrix.

The operations of addition, scalar multiplication, and composition of linear functions define analogous operations on matrices, that we call *matrix addition*, *scalar multiplication*, and *product*.

DEFINITION 18 (**Matrix addition and scalar multiplication)**. For any positive integers m, n we have the operations of addition

$$\mathbf{M}m \times n \times \mathbf{M}m \times n \longrightarrow \mathbf{M}m \times n, \quad (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij}),$$

and scalar multiplication

$$\mathbb{R} \times \mathbf{M}m \times n \longrightarrow \mathbf{M}m \times n, \quad \lambda\,(a_{ij}) = (\lambda\, a_{ij}).$$

Of course, these are the "same" operations we've seen in Section 3.1.1, the only difference is the point of view. We now view these operations as defined on the set of matrices. In particular all the *vector space axioms*, i.e. the properties listed in Theorem 3.1.5 hold.

Since we have proved[5] we don't really need to prove it again just because we changed our point of view. It is instructive however to give "purely algebraic" proofs, i.e. proofs that don't rely on the fact that matrices induce linear functions, and these properties hold for the corresponding operations of linear functions.

In fact, all these properties can be proved in exactly the same manner as the corresponding properties of vector addition and scalar multiplication, see Theorem 2.1.1. All we need to do is add an extra subscript in the calculations. Here is how to prove property (8) for example.

Let $A = (a_{ij})$ be a matrix and let λ, μ be scalars. Then using the definition of scalar multiplication we get

$$\lambda\,(\mu\, A) = \lambda \left(\mu \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}\right) = \lambda \begin{pmatrix} \mu\, a_{11} & \cdots & \mu\, a_{1n} \\ \vdots & \ddots & \vdots \\ \mu\, a_{m1} & \cdots & \mu\, a_{mn} \end{pmatrix} = \begin{pmatrix} \lambda\,(\mu\, a_{11}) & \cdots & \lambda\,(\mu\, a_{1n}) \\ \vdots & \ddots & \vdots \\ \lambda\,(\mu\, a_{m1}) & \cdots & \lambda\,(\mu\, a_{mn}) \end{pmatrix}.$$

Now we use the fact that multiplication of real numbers is associative, and again the definition of scalar multiplication we have that the last matrix is

$$= \begin{pmatrix} (\lambda\,\mu)\; a_{11} & \cdots & (\lambda\,\mu)\; a_{1n} \\ \vdots & \ddots & \vdots \\ (\lambda\,\mu)\; a_{m1} & \cdots & (\lambda\,\mu)\; a_{mn} \end{pmatrix} = (\lambda\,\mu) \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = (\lambda\,\mu)\; A.$$

Exercise 7. Prove all the *Vector Space Axioms* (i.e. the properties listed in Theorem 3.1.5) for matrices in this manner.

[5] We did do the proofs left as an exercise. Didn't we?

The above discussion suggests that we can think of matrices as vectors. In fact an $m \times n$ matrix consists of mn numbers arranged in a rectangular manner, and if we read them starting with the leftmost element of the top row we get the coordinates of a mn-vector, i.e. and element of $\mathbb{R}^{mn}$. For example,

$$\mathbf{M}_{2\times 3} \ni \begin{pmatrix} 1 & 2 & -1 \\ 3 & -2 & 0 \end{pmatrix} \cong (1, 2, -1, 3, -2, 0) \in \mathbb{R}^6.$$

If we then identify 2×3 matrices with 6-dimensional vectors this way, then we see that matrix addition and scalar multiplication of matrices is just vector addition and scalar multiplication of vectors. No surprise then that these two sets of operations have the same properties, in some sense they are the same operations!

We will further pursue these ideas later in these class, we will say then that the identification of $\mathbf{M}_{2\times 3}$ with $\mathbb{R}^6$ that we just described is an *isomorphism of Vector Spaces*.

With the above identification the standard basis of $\mathbb{R}^{mn}$ translates to matrices that have all entries 0 except one 1.

DEFINITION 19 (**Notation: The Kronecker delta**). The *Kronecker delta* is defined via

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

The two variables are usually natural numbers but in principle they could be any two mathematical objects.

EXAMPLE 44. The *dot product* of two n-vectors $\mathbf{v} = (v_i)$ and $\mathbf{w} = (v_i)$ is given by the formula

$$\mathbf{v} \cdot \mathbf{w} = \sum_{j=1}^{n} v_i \, \delta_{ij} \, w_j.$$

The standard basis of $\mathbf{R}^n$ consists of the vectors

$$\mathbf{e}_i = (\delta_{ij})_{j=1}^{n}.$$

The identity matrix is

$$I_n = (\delta_{ij})_{i,j=1}^{n}.$$

DEFINITION 20 (**The standard basis of $\mathbf{M}_{m\times n}$**). For $i = 1, \ldots, m$, and $j = 1, \ldots, n$ the basic matrix $E_{i,j}$ has the i, j-th entry equal to 1 and all other entries equal to 0. In other words, if $e_{k\ell}$ is the entry at the k-th row and ℓ-th column then

$$e_{k\ell} = \delta_{ik}\delta_{j\ell}.$$

For example here are the four basic 2×2 matrices:

$$E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Now any 2×2 matrix can is a linear combination of these four basic matrices. Indeed we have

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + a_{12} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + a_{21} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + a_{22} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

PROPOSITION 2. *An $m \times n$ matrix can be written as a linear combination of the basic matrices in a unique way. In fact the i, j-th entry is the coefficient of E_{ij}.*

PROOF. Exercise. The general case is very similar to the 2×2 case proved above. □

If A is an $m \times k$ and B a $k \times n$ matrix then A and B define linear maps that can be composed and the composition is a linear map. From an algebraic point of view, we call the matrix of the composition $A\,B$ the *product* of A and B. Let's recall the definition.

Matrix Multiplication

If $A = (a_{ij}) \in \mathbf{M}_{m\times k}$ and $B = (b_{ij}) \in \mathbf{M}_{k\times n}$ then their product $C := A\,B \in \mathbf{M}_{m\times n}$ is defined, and if $C = (c_{ij})$ then,

$$c_{ij} = \sum_{\ell=1}^{k} a_{i\ell} b_{\ell j}.$$

Equivalently, if $a_1^*, \ldots, a_m^*$ are the rows of A, and $b_1, \ldots, b_n$ are the columns of B we have

$$\begin{pmatrix} a_1^* \\ \vdots \\ a_m^* \end{pmatrix} \begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix} = \begin{pmatrix} a_1^* \cdot b_1 & \cdots & a_1^* \cdot b_n \\ \vdots & \ddots & \vdots \\ a_m^* \cdot b_1 & \cdots & a_m^* \cdot b_n \end{pmatrix}$$

The following theorem states some fundamental algebraic properties of matrix multiplication, and its interactions with matrix addition and matrix multiplication. If we think as matrices as linear maps then these properties are straightforward to verify. Furthermore they hold for all maps, not only linear maps. For example, composition of functions is associative. To see this let $h\colon X \longrightarrow Y$, $g\colon Y \longrightarrow Z$, and $f\colon Z \longrightarrow W$ be three functions. Then

$$(f \circ g) \circ h = f \circ (g \circ h)\,.$$

We first note that the compositions are defined and they have the same domain, namely X, and the same codomain, namely W. To prove that they are equal we need to prove that for all $x \in X$ we have

$$((f \circ g) \circ h)\,(x) = (f \circ (g \circ h)\,(x))\,.$$

This is straightforward:

$$\begin{aligned} ((f \circ g) \circ h)\,(x) &= (f \circ g)\,(h(x)) \\ &= f\,(g\,(h(x))) \\ &= f\,((g \circ h)\,(x)) \\ &= (f \circ (g \circ h)\,(x))\,. \end{aligned}$$

However, this is an algebraic section. So we will be giving mostly algebraic proofs.

THEOREM 3.4.1 (**Matrices form an algebra**). *The following properties hold for all matrices A, B, C and all scalars λ, μ provided that the operations are defined*[6].

(a) Matrix multiplication is associative.

$$A\,(B\,C) = (A\,B)\,C.$$

(b) Matrix multiplication distributes over matrix addition on both sides.

$$(A + B)\,C = A\,C + A\,B, \qquad A\,(B + C) = A\,B + A\,C.$$

[6] When is that the case? For each property, find what conditions must hold for the dimensions of A, B, and C for the operations in each side to be defined.

(c) Scalar multiplication is compatible with matrix multiplication.

$$(\lambda A)\, B = A\,(\lambda B) = \lambda\,(A\,B)\,.$$

(d) Multiplication with the identity matrix

$$I\,A = A, \qquad A\,I = A.$$

PROOF. We prove (1) leaving the remaining as an exercise. Let $A\,B = T$, and $B\,C = S$. Then

$$t_{i\ell} = \sum_{j=1}^{m} a_{ij} b_{j\ell}, \qquad s_{ij} = \sum_{k=1}^{n} b_{jk} c_{kj}.$$

Then the i,p entry of $A\;(B\,C)\,S = A\,S$ is

$$a_{i1}t_{1p} + a_{i2}t_{1p} + \cdots + a_{in}t_{np} = \sum_{k=1}^{n}\sum_{j=1}^{m} (a_{ij}b_{jk})c_{kp}.$$

Similarly, the i,p entry of $(A\,B)\,C = T\,S$ is

$$\sum_{k=1}^{n}\sum_{j=1}^{m} a_{ij}(b_{jk}c_{kp}).$$

Multiplication of real numbers is associative and therefore for all k,p the k,p entries of $A\,(B\,C)$ and $(A\,B)\,C$ are equal. Therefore the matrices are equal. □

Many other properties follow the from the properties listed in Theorem 3.4.1. A very important is stated in the following proposition. This proposition if obvious if actually use the definition of the matrix product, multiplying any number with zero gives zero and adding a bunch of zeros also gives zero. However we provide a proof using only the four properties listed in Theorem 3.4.1, the benefit of this being that the proposition will be true whenever those properties (as well as the *vector space axioms*) hold.

PROPOSITION 3. *If O is the $m \times n$ zero matrix then for any $n \times k$ matrix A we have*

$$O\,A = O,$$

where O in the RHS stands for the $m \times k$ zero matrix.

Similarly, if B is an $k \times m$ matrix then

$$B\,O = O,$$

where O in the RHS stands for the $k \times n$ zero matrix.

PROOF. We have

$$O\,A = (O + O)\,A = O\,A + O\,A.$$

Subtracting $O\,A$ from both sides yields the result.

The proof of the second statement is entirely similar and is left as an exercise. □

Notice that a property we usually expect for multiplication, namely the *commutative property* is not listed. The reason is, of course, that it is not true, that is **it is not true that** for all A, B

$$A\,B = B\,A. \tag{3.11}$$

First of all, if $A\,B$ is defined, $B\,A$ is not necessarily defined. In order for both products to be defined we need to have that if A is an $m \times n$ matrix then B is $n \times m$. And even in that case, $A\,B$ and $B\,A$ have different dimensions in general, the first is $m \times m$ and the second $n \times n$. So the only case that we could have that (3.11) has a chance of holding is when $m = n$. But even then it is not generally true. As an example consider that standard basis of $\mathbf{M}_{3\times3}$. We can easily verify that

$$E_{12}\,E_{23} = E_{13},\ \text{while } E_{23}\,E_{12} = O.$$

The last example exhibits an other surprising property of matrix multiplication. Sometimes the product of two non-zero matrices may be zero. In other words, for matrices A, B **it is not true that**

$$A\,B = O \implies A = O \text{ or } B = O.$$

3.4.1. The algebra of Square Matrices. We now concentrate on the set of *square matrices* $\mathbf{M}_n$. If A, B are two $n \times n$ square matrices, then $A\,B$ is always defined, and is actually also an $n \times n$ matrix. The set $\mathbf{M}_n$ endowed with matrix addition, scalar multiplication, and matrix multiplication is often referred to, as *the algebra of square matrices*.

In general, Equation (3.11) does not hold. Actually *most of the times* it doesn't hold. When it does hold, it's special and we give it a name.

DEFINITION 21 (**Commuting matrices**). If Equation (3.11) holds for $A, B \in \mathbf{M}_n$ we say that A and B commute.

Of course, A always commutes with itself, and the identity matrix I as well as the zero matrix O commute with all matrices.

EXAMPLE 45. The matrices

$$A = \begin{pmatrix} 1 & 2 & 3 \\ -2 & 0 & 1 \\ -1 & -3 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} -6 & -7 & 11 \\ -3 & -7 & -4 \\ 3 & -8 & -2 \end{pmatrix},$$

commute. Indeed, by direct calculations[7] we see that

$$A\,B = \begin{pmatrix} -3 & -45 & -3 \\ 15 & 6 & -24 \\ 21 & 12 & -3 \end{pmatrix} = B\,A.$$

EXAMPLE 46 (**Finding the set of matrices that commutes with a given matrix**). We often want to know the set of matrices that commute with a given matrix or even all matrices in a given set. If $S \subseteq \mathbf{M}_n$ then the set of matrices that commute with all elements of S is called the *centralizer* of S. Here is an example on how to find the centralizer of a single matrix.

Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. We want to find all matrices $M = \begin{pmatrix} x & y \\ z & t \end{pmatrix}$ that commute with A. In other words we want

$$A\,M = M\,A.$$

Now

$$A\,M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x & y \\ z & t \end{pmatrix} = \begin{pmatrix} x+z & y+t \\ z & t \end{pmatrix},$$

while

$$M\,A = \begin{pmatrix} x & y \\ z & t \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} x & x+y \\ z & z+t \end{pmatrix}.$$

[7] Do them!

Therefore we need

$$\begin{pmatrix} x & x+y \\ z & z+t \end{pmatrix} = \begin{pmatrix} x+z & y+t \\ z & t \end{pmatrix}.$$

This is equivalent to the linear system:

$$\begin{cases} x & = x \\ x+y & = y+t \\ z & = z \\ t \quad + z & = t \end{cases}.$$

Solving this is rather straightforward. From the second equation we have $x = t$ and from the last $z = 0$. So we conclude that in order to commute with A, M has to have the form

$$M = \begin{pmatrix} x & y \\ 0 & x \end{pmatrix}, \quad x, y \in \mathbb{R}.$$

DEFINITION 22 (**Algebra of Matrices**). We say that a *nonempty* subset $\mathbf{A} \subseteq \mathbf{M}_n$ is a *subalgebra*, or that it is an *algebra of matrices*, if $\mathbf{A}$ is closed under the operations of matrix addition, scalar multiplication, and matrix multiplication. This means that if $A, B \in \mathbf{A}$ and $\lambda \in \mathbb{R}$ then

(a) $\lambda A \in \mathbf{A}$.
(b) $A + B \in \mathbf{A}$.
(c) $A B \in \mathbf{A}$.

If in addition any two elements of $\mathbf{A}$ commute, that is, if in addition

(d) $A B = B A$,

then we say that $\mathbf{A}$ is a *commutative algebra of matrices*.

THEOREM 3.4.2. *If $\mathbf{A}$ is an algebra of matrices then:*

(a) $O \in \mathbf{A}$.
(b) $A \in \mathbf{A} \implies -A \in \mathbf{A}$.
(c) If $A \in \mathbf{A}$ and A is invertible then $A^{-1} \in \mathbf{A}$.
(d) If $\mathbf{A}$ contains invertible elements then $I \in \mathbf{A}$.

The first two properties follow from the fact that $\mathbf{A}$ is closed under scalar multiplication. Just take $\lambda = 0$ for the first and $\lambda = -1$ for the second. The fourth property follows from the third and the fact that $\mathbf{A}$ is closed under matrix multiplication.

The proof of the third property requires more ammunition than we have currently available. I will give a proof towards the end of this section but the proof will not be complete because it depends on a celebrated theorem, the Cayley-Hamilton Theorem that we will see later in the course.

EXAMPLE 47 (**Trivialities**). The subset $\{O\}$ consisting only of the zero matrix is clearly a subalgebra called *the zero subalgebra*. There are no invertible elements in this algebra.

EXAMPLE 48 (**The algebra of scalar matrices**). A slightly non trivial example is the algebra of *scalar matrices*. Let

$$\mathbf{R}_n = \{\lambda I_n : \lambda \in \mathbb{R}\}.$$

The elements of $\mathbf{R}_n$ are called scalar matrices because they behave like scalars. For example for $\mathbf{x} \in \mathbb{R}^n$ we have

$$(\lambda I)\mathbf{x} = \lambda \mathbf{x}.$$

Thus multiplying with a scalar matrix λI gives the same result as multiplying with the scalar λ. Similarly, adding two scalar matrices, results in the scalar matrix obtained by adding the corresponding scalars:

$$\lambda\, I + \mu\, I = (\lambda + \mu)\; I.$$

So $\mathbf{R}_n$ is a commutative algebra of matrices. The invertible elements are the scalar matrices $\lambda\, I$ with $\lambda \neq 0$, and of course

$$(\lambda\, I)^{-1} = \lambda^{-1}\, I.$$

EXAMPLE 49 (**The algebra of diagonal matrices**). A *diagonal* $n \times n$ matrix is a matrix D with $d_{ij} = 0$ for $i \neq j$, i.e. non-zero entries can occur only along the main diagonal. If $\lambda_1, \ldots, \lambda_n$ are n scalars then we define $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ to be the diagonal matrix with $\lambda_1, \ldots, \lambda_n$ in the main diagonal. For example

$$\operatorname{diag}(1, -7, 0, 42) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -7 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 42 \end{pmatrix}.$$

The scalar matrix $\lambda\, I$ is thus $\operatorname{diag}(\lambda, \ldots, \lambda)$.

Notice that the i-th row (as well as the i-th column) of $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is $\lambda_i\, \mathbf{e}_i$. This means that when we multiply a diagonal matrix with another matrix only one of the products in the sum that gives the i, j entry of the product matrix is (possibly) non-zero.

Let A be any matrix with n rows and let, as usual, $\mathbf{a}_1^*, \ldots, \mathbf{a}_n^*$ (respectively $\mathbf{a}_1, \ldots, \mathbf{a}_m$) be its row (respectively column) vectors. Then,

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n)\, A = \begin{pmatrix} \lambda_1\, \mathbf{a}_1^* \\ \vdots \\ \lambda_n\, \mathbf{a}_n^* \end{pmatrix}, \quad A\, \operatorname{diag}(\lambda_1, \ldots, \lambda_m) = \begin{pmatrix} \lambda_1\, \mathbf{a}_1 & \ldots & \lambda_m\, \mathbf{a}_m \end{pmatrix}.$$

So multiplying from a left by a diagonal matrix has the effect of multiplying the rows of A with the scalars along the diagonal, while multiplying from the right has the effect of multiplying the columns of A.

It follows that

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n)\; \operatorname{diag}(\mu_1, \ldots, \mu_n) = \operatorname{diag}(\lambda_1\, \mu_1, \ldots, \lambda_n\, \mu_n).$$

So the product of two diagonal matrices is also diagonal, and furthermore any two diagonal matrices commute. In particular, since scalar matrices are special cases of diagonal matrices we also see that the set of diagonal matrices is also closed under scalar multiplication.

We also have that

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n) + \operatorname{diag}(\mu_1, \ldots, \mu_n) = \operatorname{diag}(\lambda_1 + \mu_1, \ldots, \lambda_n + \mu_n),$$

and we established that the set of diagonal matrices is a commutative algebra.

It is rather straightforward[8] to see that a diagonal matrix is invertible if and only if all diagonal entries are non=zero. In that case,

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n)^{-1} = \operatorname{diag}(\lambda_1^{-1}, \ldots, \lambda_n^{-1}).$$

EXAMPLE 50 (**The algebra of upper triangular matrices**). A square matrix T is called *(upper) triangular*if all the entries below the main diagonal are 0, in other words $T = (t_{ij})$ is triangular if

$$i > j \implies a_{ij} = 0.$$

[8] Is it?

For example here is an upper triangular 4×4 matrix:

$$\begin{pmatrix} 1 & 11 & 0 & 0 \\ 0 & -7 & 41 & 42 \\ 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 42 \end{pmatrix}.$$

The set of $n \times n$ triangular matrices is denoted by $\mathbf{\Delta}_n$. It is easy[9] to see that $\mathbf{\Delta}_n$ is closed under addition and scalar multiplication. To see that it is also closed under multiplication notice that the i-th row of a triangular matrix has zero entries up to the $(i-1)$-th column, while its j-th column has all zero entries after the j-th row. So if A and B are triangular matrices and $i > j$ then the dot product $\mathbf{a}_i^* \cdot \mathbf{b}_j = 0$, and therefore the i, j entry of AB is 0.

So we established that $\mathbf{\Delta}_n$ is an algebra of matrices. For future use we observe that the diagonal entries of the product of two triangular matrices are just the products of the corresponding diagonal entries.

The invertible elements of $\mathbf{\Delta}_n$ are exactly the triangular matrices with all diagonal entries non-zero. For, if this the case then we have a matrix in echelon form with non-zero diagonals. If on the other hand there a 0 in the diagonal then the corresponding column is a free column, and therefore the matrix has non-zero nullity.

EXAMPLE 51 (**Centralizers**). Recall from Example 46 that if $S \subseteq \mathbf{M}_n$ then the centralizer of S is the set of all matrices that commute with all elements of S. Denoting the centralizer of S by $\mathcal{C}$ we thus have

$$A \in \mathcal{C} \iff \forall X \in S, \quad AX = XA.$$

I claim that $\mathcal{C}$ is an algebra. The claim follows from the following three facts:

- If A and X commute, so do λA and X.

  PROOF. Assume A and X commute. Then we have

  $$(\lambda A)\ X = \lambda\ (AX) = \lambda\ (XA) = X\ (\lambda A).$$

  □

- If A and X, and B and X commute, the $A + B$ and X and commute.

  PROOF. We have

  $$(A + B)\,X = AX + BX = XA + XB = X\,(A + B).$$

  □

- If A and X, and B and x commute, the AB and X and commute.

  PROOF. We have

  $$(AB)\,X = A(BX) = A\,(XB) = (AX)\,B = (XA)\,B = X\,(AB).$$

  □

EXAMPLE 52 (**A commutative algebra of matrices**). Let

$$\mathbf{A} = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} : a, b \in \mathbb{R} \right\}.$$

Then $\mathbf{A}$ is a commutative algebra of matrices. Indeed for $\lambda \in \mathbb{R}$ and $A \in \mathbb{A}$ we have

$$\lambda A = \begin{pmatrix} \lambda a & \lambda b \\ \lambda b & \lambda a \end{pmatrix}.$$

We see then that $\lambda A \in \mathbf{A}$. Thus $\mathbf{A}$ is closed under scalar multiplication.

[9] It is easy, right?

Now let A be as above and let $B = \begin{pmatrix} x & y \\ y & x \end{pmatrix}$ be a second element of **A**. Then

$$A + B = \begin{pmatrix} a+x & b+y \\ b+y & a+x \end{pmatrix}.$$

Thus $A + B \in \mathbb{A}$ and we established closure under matrix addition.

For multiplication we have

$$A\,B = \begin{pmatrix} a\,x + b\,y & a\,y + b\,x \\ b\,x + a\,y & b\,y + a\,x \end{pmatrix} = \begin{pmatrix} a\,x + b\,y & a\,y + b\,x \\ a\,y + b\,x & a\,x + b\,y \end{pmatrix}.$$

Hence, $A\,B \in \mathbf{A}$ and **A** is closed under multiplication as well.

We have established then that **A** is an algebra of matrices. To prove that it is commutative we compute $B\,A$ to verify that it is equal to $A\,B$.

$$B\,A = \begin{pmatrix} x\,a + y\,b & x\,b + y\,a \\ y\,a + x\,b & y\,b + x\,a \end{pmatrix} = \begin{pmatrix} a\,x + b\,y & a\,y + b\,x \\ a\,y + b\,x & a\,x + b\,y \end{pmatrix} = A\,B.$$

So **A** is a commutative algebra of matrices.

Now let's find the invertible elements of **A**. From Example 43 we know that A is invertible when its determinant is non-zero. Thus an element $A \in \mathbf{A}$ is invertible if and only if

$$a^2 - b^2 \neq 0 \iff a \neq \pm b.$$

In that case

$$A^{-1} = \frac{1}{a^2 - b^2}\begin{pmatrix} a & -b \\ -b & a \end{pmatrix}.$$

Definition 23 (**Powers of a matrix**). If $A \in \mathbf{M}_n$ and $k \in \mathbb{N}$ the power A^k is defined recursively as follows:

$$\begin{cases} A^0 & = I \\ A^{n+1} & = A^n\,A \end{cases}.$$

So,

$$\begin{aligned} A^1 &= A^0\,A \\ &= A, \end{aligned}$$

and

$$\begin{aligned} A^2 &= A^1\,A \\ &= A\,A, \end{aligned}$$

and continuing,

$$\begin{aligned} A^3 &= A^2\,A \\ &= (A\,A)\,A, \end{aligned}$$

and so on. In general, A^n is a product of n copies of A.

REMARK 13. Because of the associative property of multiplication (the first property in Theorem 3.4.1), we also have

$$A^{n+1} = A\,A^n.$$

This can be proven by induction. We just show it for the third power:

$$\begin{aligned} A\,A^2 &= A\,(A\,A) \\ &= (A\,A)\,A \\ &= A^2\,A \\ &= A^3. \end{aligned}$$

Powers of matrices enjoy some of the properties of powers that we are familiar with.

PROPOSITION 4. *If $A \in \mathbf{M}_n$ and $k, \ell \in \mathbb{N}$ we have*

(a) $A^k\,A^\ell = A^{k+\ell}$.
(b) $\left(A^k\right)^l = A^{k\,l}$.
(c) $(\lambda A)^k = \lambda^k\,A^k$.
(d) $I^k = I$.
(e) $O^k = O$.

However **it's not true**, that

$$(A\,B)^k = A^k\,B^k,$$

unless A and B commute. For example, by definition

$$(A\,B)^2 = A\,B\,A\,B.$$

But we can't swap the second and third factor, unless A and B commute.

In general, we need to be careful when we are doing algebraic manipulations with matrices. For example if A, B are $n \times n$ square matrices, then we have

$$(A + B)^2 = A^2 + A\,B + B\,A + B^2,$$

which, if A and B commute simplifies to the familiar

$$(A + B)^2 = A^2 + 2\,A\,B + B^2.$$

Similarly,

$$(A + B)(A - B) = A^2 - A\,B + B\,A - B^2,$$

which, if A and B commute, simplifies to the familiar

$$(A + B)(A - B) = A^2 - B^2.$$

Since I commutes with all matrices, we have that

$$A^2 - I = (A - I)(A + I)$$

and

$$(A \pm I)^2 = A^2 \pm 2\,A + I.$$

The following Theorem follows from the more general Theorem 3.4.5.

THEOREM 3.4.3. *If A is invertible, then A^k is also invertible for all natural numbers k and*

$$\left(A^k\right)^{-1} = \left(A^{-1}\right)^k.$$

Since later in this section we will prove a more general Theorem about the interaction of matrix multiplication and inverses let us just see why the theorem is true with an example. Say $k = 3$. Then (eschewing parenthesis as associativity allows us to)

$$A^3 = A\,A\,A, \quad \left(A^{-1}\right)^3 = A^{-1}\,A^{-1}\,A^{-1}.$$

Therefore:

$$\begin{aligned} A^3\left(A^{-1}\right)^3 &= A\,A\,A\,A^{-1}\,A^{-1}\,A^{-1} \\ &= A\,A\,I\,A^{-1}\,A^{-1} \\ &= A\,A\,A^{-1}\,A^{-1} \\ &= A\,I\,A^{-1} \\ &= A\,A^{-1} \\ &= I. \end{aligned}$$

Entirely similarly,

$$\left(A^{-1}\right)^3 A^3 = I.$$

Thus indeed,

$$\left(A^3\right)^{-1} = \left(A^{-1}\right)^3.$$

So we can now define negative powers, at least for invertible matrices.

DEFINITION 24. If A is an invertible matrix, then for any negative integer k we define

$$A^k = \left(A^{-1}\right)^{-k}.$$

Let's collect a few basic facts about powers of matrices. The proofs are either straightforward, already embedded in the discussion we've had so far, or are special cases of theorems we'll prove later in this section. Make sure that you can provide the proofs, if you can't at first reading come back after you finish this section.

PROPOSITION 5. *The following hold. The powers could be positive or negative integers; in the later case we assume that the involved matrices are invertible.*

(a) *Properties (1) through (3) in Proposition 4 hold for all integers, provided A is invertible. Property (4) also holds for all integers. Property (5) of course doesn't make sense for negative k[10].*
(b) *All powers of the same matrix commute.*
(c) *If $\mathbf{A}$ is an algebra of matrices, $A \in \mathbf{A}$ and $k \in \mathbb{Z}$ then $A^k \in \mathbf{A}$ if defined.*
(d) *If A is invertible then for all matrices B we have:*

$$\left(A^{-1}\,B\,A\right)^k = A^{-1}\,B^k\,A,$$

provided that B^k is defined.

EXAMPLE 53 (**Powers of diagonal matrices**). Refer to Example 49 for the notation used. Let $A = \operatorname{diag}(a_1, \ldots, a_n)$ then for all $k \geq 0$ we have

$$A^k = \operatorname{diag}\left(a_1^k, \ldots, a_n^k\right).$$

If all diagonal entries are non-zero this is true for negative k as well.

[10] Why?

Let's prove this by induction. It clearly it is true for $k = 0$. Now,

$$\begin{aligned} A^{k+1} &= A^k\, A \\ &= \operatorname{diag}\left(a_1^k, \dots, a_n^k\right)\, \operatorname{diag}\left(a_1, \dots, a_n\right) \\ &= \operatorname{diag}\left(a_1^k\, a_1, \dots, a_n^k\, a_n\right) \\ &= \operatorname{diag}\left(a_1^{k+1}, \dots, a_n^{k+1}\right). \end{aligned}$$

Now, if all diagonal entries are non-zero then (see Example 49) A is invertible and if $k < 0$ then $-k > 0$ and

$$A^k = \left(A^{-1}\right)^{-k} = \operatorname{diag}\left(a_1^{-1}, \dots, a_n^{-1}\right)^{-k} = \operatorname{diag}\left(a_1^k, \dots, a_n^k\right).$$

EXAMPLE 54. Consider the matrix

$$A = \frac{1}{2}\begin{pmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix}.$$

We have

$$A^2 = \frac{1}{4}\begin{pmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix}\begin{pmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1-3 & -\sqrt{3}-\sqrt{3} \\ \sqrt{3}+\sqrt{3} & -3+1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}.$$

Then

$$A^3 = A^2\, A = \frac{1}{4}\begin{pmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} -4 & 0 \\ 0 & -4 \end{pmatrix} = -I.$$

Then,

$$A^4 = A^3\, A = -A, \quad A^5 = A^4\, A = -A^2, \quad A^6 = I.$$

From now on the powers will repeat in cycles of length 6. The next cycle is

$$A^7 = A^6\, A = I\, A = A,$$

and then

$$\begin{aligned} A^8 &= A^2 \\ A^9 &= A^3 \\ A^{10} &= A^4 \\ A^{11} &= A^5 \\ A^{12} &= I. \end{aligned}$$

We can express this periodic pattern using *modular arithmetic*. Any integer m can be *uniquely* written as $m = 6\,k + i$ where $k \in \mathbb{Z}$ and $i \in \{0, 1, \dots, 5\}$, where k is the *quotient* and i the *remainder* of the division $m \div 6$. Then

$$A^m = A^{6\,k+i} = A^{6\,k}\, A^i = \left(A^6\right)^k\, A^i = I^k\, A^i = A^i.$$

For example, since 12435 leaves remainder 3 when divided by 6 we have

$$A^{12435} = A^3 = -I.$$

Or, since 134 leaves remainder 2 we have

$$A^{134} = A^2 = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}.$$

EXAMPLE 55. Consider the matrix

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

We have

$$A^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0+0 & 0+0 \\ 0+0 & 0+0 \end{pmatrix} = O.$$

Then,

$$A^3 = A^2\, O = O, \quad A^4 = A^3\, O = O, \quad \dots$$

Thus all power of A after the first are the zero matrix.

EXAMPLE 56. Let's find the powers of

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

Of course,

$$A^0 = I, \quad A^1 = A.$$

Now,

$$A^2 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1+0+1 & 0+0+0 & 1+0+1 \\ 0+0+0 & 0+0+0 & 0+0+0 \\ 1+0+1 & 0+0+0 & 1+0+1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & 2 \end{pmatrix}.$$

Now notice that

$$A^2 = 2\, A.$$

So,

$$A^3 = A^2\, A = (2\, A)\, A = 2\, A^2 = 2\,(2\, A) = 4\, A.$$

and

$$A^4 = A^3\, A = (4\, A)\, A = 4\, A^2 = 4\,(2\, A) = 8\, A.$$

And this pattern will continue, to get the fifth power we multiply A^4 with A, and we'll get $8\, A^2 = 16\, A$. Thus we have,

$$A^n = 2^{n-1}\, A = \begin{pmatrix} 2^{n-1} & 0 & 2^{n-1} \\ 0 & 0 & 0 \\ 2^{n-1} & 0 & 2^{n-1} \end{pmatrix}.$$

We can formalize the above argument to an inductive proof. So we will prove, using induction, that for all $n \geq 1$[11]

$$A^n = 2^{n-1}\, A.$$

For $n = 1$ the formula clearly holds since both sides are equal to A. Assuming it holds for n we get

$$A^{n+1} = A^n\, A = \left(2^{n-1}\, A\right) A = 2^{n-1}\, A^2 = 2^{n-1}\, 2\, A = 2^n\, A = 2^{n+1-1}\, A.$$

[11] Why the formula doesn't work for $n = 0$?

EXAMPLE 57. Let

$$A = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

A direct calculation shows that

$$A^2 = I.$$

Then,

$$A^3 = A^2\,A = I\,A = A, \qquad A^4 = A^3\,A = A\,A = I.$$

And therefore[12]

$$A^n = \begin{cases} I & n \text{ even} \\ A & n \text{ odd} \end{cases}.$$

Now that we have powers, scalar multiplication, and addition we can plug a matrix in any polynomial with real coefficients.

DEFINITION 25 (**Evaluating polynomials at matrices**). Let

$$p(x) = \sum_{j=0}^{d} a_j\,x^j = a_0\,x^0 + a_1\,x^1 + \cdots + a_{d-1}\,x^{d-1} + a_d\,x^d,$$

be a polynomial of degree d, where $a_i \in \mathbb{R}$, and let $A \in \mathbf{M}_n$. Then *the evaluation of* $p(x)$ *at* A is defined via

$$p\,(A) = \sum_{j=0}^{d} a_j\,A^j = a_0\,A^0 + a_1\,A^1 + \cdots + a_{d-1}\,A^{d-1} + a_d\,A^d.$$

If $p\,(A) = O$ then we say that A is a *root* or *zero* of $p(x)$.

REMARK 14. Since $A^0 = I$ we often write

$$p\,(A) = a_0\,I + a_1\,A + \cdots + a_{d-1}\,A^{d-1} + a_d\,A^d.$$

EXAMPLE 58. Let $A = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}$, and let $p(x) = x^3 - 2\,x^2 - 2\,x + 6$, and $q(x) = x^2 - 4\,x + 13$. We calculate:

$$A^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A^1 = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}, \quad A^2 = \begin{pmatrix} -5 & -12 \\ 12 & -5 \end{pmatrix}, \quad A^3 = \begin{pmatrix} -46 & -9 \\ 9 & -46 \end{pmatrix}.$$

Then

[12] Give an inductive proof of this.

$$\begin{aligned} p\,(A) &= A^3 - 2\,A^2 = 2\,A + 6 \\ &= \begin{pmatrix} -46 & -9 \\ 9 & -46 \end{pmatrix} - 2\begin{pmatrix} -5 & -12 \\ 12 & -5 \end{pmatrix} - 2\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} + 6\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -46 & -9 \\ 9 & -46 \end{pmatrix} + \begin{pmatrix} 10 & 24 \\ -24 & 10 \end{pmatrix} + \begin{pmatrix} -4 & 6 \\ -6 & -4 \end{pmatrix} + \begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix} \\ &= \begin{pmatrix} -34 & 21 \\ -21 & -34 \end{pmatrix}. \end{aligned}$$

And

$$\begin{aligned} q\,(A) &= A^2 - 4\,A + 13\,I \\ &= \begin{pmatrix} -5 & -12 \\ 12 & -5 \end{pmatrix} - 4\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} + 13\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -5 & -12 \\ 12 & -5 \end{pmatrix} + \begin{pmatrix} -8 & 12 \\ -12 & -8 \end{pmatrix} + \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

So A is a root of $q(x)$.

The following theorem is immediate[13].

THEOREM 3.4.4. *If* $\mathbf{A}$ *is an algebra of matrices,* $A \in \mathbf{A}$ *and* $p(x)$ *is any polynomial, then* $p(A) \in \mathbf{A}$.

3.4.2. Invertible Matrices. We now focus on invertible matrices. We already know quite a few characterizations of invertible matrices, and we will see a few more down the road. From an algebraic point of view perhaps the following definition is the most convenient.

DEFINITION 26 (**General Linear Group**). We say that a square matrix $A \in \mathbf{M}_n$ is invertible if there exists a matrix B in $\mathbf{M}_n$ such that

$$A\,B = I = B\,A. \tag{3.12}$$

In that case we call B the *inverse of* A and write $A^{-1} = B$.

The set of $n \times n$ invertible matrices is called *the General Linear Group* and is denoted by $\mathrm{GL}(n)$.

THEOREM 3.4.5 (**Invertible matrices form a group**). *We have*

(a) $A \in \mathrm{GL}(n) \implies A^{-1} \in \mathrm{GL}(n)$*, and actually*

$$\left(A^{-1}\right)^{-1} = A.$$

(b) $A, B \in \mathrm{GL}(n) \implies A\,B \in \mathrm{GL}(n)$*, and actually*[14]

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

[13] Provide the proof

[14] Notice the reverse of the order!

PROOF. The first is obvious since by definition we have

$$A\,A^{-1} = I = A^{-1}\,A,$$

and therefor A is the inverse of A^{-1}.

For the second we have:

$$(A\,B)\left(B^{-1}\,A^{-1}\right) = A\left(B\,B^{-1}\right)A^{-1} = A\,I\,A^{-1} = A\,A^{-1} = I.$$

Similarly,

$$\left(B^{-1}\,A^{-1}\right)(A\,B) = B^{-1}\left(A^{-1}A\right)B = B^{-1}\,I\,B = B^{-1}\,B = I.$$

□

It turns out that we don't need to check that both products in Equation (3.12) give the identity matrix. As the following Lemma shows, if one of the products is the identity the other will be as well.

LEMMA 3. *If $A\,B = I$ then we also have $B\,A = I$ and therefore $B = A^{-1}$. Similarly, if $A\,B = I$ then $B = A^{-1}$.*

PROOF. If $A\,B = I$ then for all $\mathbf{x} \in \mathbb{R}^n$ we have

$$A\,(B\,\mathbf{x}) = \mathbf{x}.$$

So, every $\mathbf{x} \in \mathbb{R}^n$ is in the range of A and therefore A is surjective. By Theorem 3.3.2 it follows that A is invertible. We have then

$$\begin{aligned} A\,B = I &\implies A^{-1}(A\,B) = A^{-1} \\ &\implies \left(A^{-1}\,A\right)B = A^{-1} \\ &\implies I\,B = A^{-1} \\ &\implies B = A^{-1}. \end{aligned}$$

□

The properties listed in Theorem 3.4.5 have many important consequences, so we abstract them by introducing the concept of a *group*. Groups play a fundamental role not only in modern mathematics, but in physics and other sciences as well.

DEFINITION 27 (**Group of functions**). Let G be a set of functions with domain and codomain the same set X. We say that G is *a group* if the following hold:

(a) The identity function of X is in G.
(b) G is closed under composition of functions.
(c) All elements of G are invertible, and their inverses are also in G. That is,

$$g \in \mathrm{G} \implies g^{-1} \in \mathrm{G}.$$

Thus Theorem 3.4.5 says that $\mathrm{GL}(n)$ is a group. Usually the operation of composition is written as multiplication and we can define powers g^n where $g \in \mathrm{G}$ and $n \in \mathbb{Z}$, that satisfy the algebraic properties (1), (2), and (4) of Proposition 4. We will not pursue this further at this point. We'll come back to these ideas later though.

In Examples 54 and 57 we have matrices where one of their powers is the identity matrix. Lemma 3 implies that such matrices are invertible because if $A^k = I$ then $A^{k-1}\,A = I$.

EXAMPLE 59. Let

$$A = \frac{1}{9}\begin{pmatrix} 4 & 7 & -4 \\ 1 & 4 & 8 \\ 8 & -4 & 1 \end{pmatrix}.$$

We calculate[15] that

$$A^2 = \frac{1}{9}\begin{pmatrix} -1 & 8 & 4 \\ 8 & -1 & 4 \\ 4 & 4 & -7 \end{pmatrix}, \quad A^3 = \frac{1}{9}\begin{pmatrix} 4 & 1 & 8 \\ 7 & 4 & -4 \\ -4 & 8 & 1 \end{pmatrix}, \quad A^4 = I.$$

We conclude that A is invertible and

$$A^{-1} = A^3 = \frac{1}{9}\begin{pmatrix} 4 & 1 & 8 \\ 7 & 4 & -4 \\ -4 & 8 & 1 \end{pmatrix}.$$

When $A^k = I$ the matrix A is a root of the polynomial $x^k - 1$. More generally we have the following proposition.

PROPOSITION 6. *If X is a root of a polynomial $p(x) = a_k x^k + \cdots + a_1 x + a_0$ with constant term $a_0 \neq 0$, then X is invertible.*

PROOF. We have

$$a_k X^k + \cdots + a_1 X + a_0 I = O \iff X\left(a_k X^{k-1} + \cdots + a_1 I\right) = -a_0 I.$$

If $a_0 \neq 0$ then we can divide both sides by a_0 to get

$$X\left(-\frac{a_k}{a_0} X^{k-1} - \cdots - \frac{a_1}{a_0} I\right) = I.$$

So X is invertible by Lemma 3, and furthermore

$$X^{-1} = -\frac{a_k}{a_0} X^{k-1} - \cdots - \frac{a_1}{a_0} I,$$

that is the inverse of X can be expressed as a polynomial in X. □

EXAMPLE 60. In Example 58 we saw that $A = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}$, is a root $q(x) = x^2 - 4x + 13$. So, A is invertible and

$$A^{-1} = -\frac{1}{13}(A - 4I) = -\frac{1}{13}\begin{pmatrix} -2 & -3 \\ 3 & -2 \end{pmatrix}.$$

EXAMPLE 61. Let

$$X = \begin{pmatrix} -4 & 1 & 1 & 1 \\ -16 & 3 & 4 & 4 \\ -7 & 2 & 2 & 1 \\ -11 & 1 & 3 & 4 \end{pmatrix}.$$

Then one can verify[16] that X is a root of the following polynomial:

$$p(x) = x^4 - 5x^3 + 9x^2 - 7x + 2.$$

It follows that X is invertible, and its inverse is

$$X^{-1} = -\frac{1}{2}\left(X^3 - 5X^2 + 9X - 7I\right).$$

[15] Do the calculations!

[16] Do the calculations!

Since, as you have already calculated,

$$X^2 = \begin{pmatrix} -18 & 2 & 5 & 5 \\ -56 & 5 & 16 & 16 \\ -29 & 4 & 8 & 7 \\ -37 & 2 & 11 & 12 \end{pmatrix}, \text{ and } X^3 = \begin{pmatrix} -50 & 3 & 15 & 15 \\ -144 & 7 & 44 & 44 \\ -81 & 6 & 24 & 23 \\ -93 & 3 & 29 & 30 \end{pmatrix},$$

we find that[17]

$$X^{-1} = \frac{1}{2}\begin{pmatrix} 3 & -2 & 1 & 1 \\ 8 & -2 & 0 & 0 \\ -1 & -4 & 5 & 3 \\ 7 & -2 & -1 & 1 \end{pmatrix}.$$

Down the road, as a consequence of the Cayley-Hamilton Theorem we will see that the converse of Proposition 6 is also true. We state the proposition postponing the proof.

PROPOSITION 7. *If X is invertible then it is a root of a polynomial with non-zero constant term.*

Combining this with 3.4.4 we get the proof of Theorem 3.4.2.

PROOF OF THEOREM 3.4.2. If A is invertible then by Propositions 7 and 6 we have that A^{-1} is a polynomial of A. Theorem 3.4.4 then implies that $A^{-1} \in \mathbf{A}$. □

[17] Do the calculations!

3.5. The transpose of a matrix and the adjoint of a linear operator

We have identified matrices with linear operations by letting matrices *act from the left* that is the image of $\mathbf{x}$ is obtained by multiplying $\mathbf{x}$ from the left, in other words

$$\mathbf{x} \longmapsto A\mathbf{x}.$$

In order for that to make sense we represent $\mathbf{x}$ as a column vector.

Now, an $m \times n$ matrix can also act on m-vectors, but it has to act from the right

$$\mathbf{x} \longmapsto \mathbf{x}A.$$

In order for this to make sense we need $\mathbf{x}$ to be a row vector. We have

$$\begin{pmatrix} x_1 & \cdots & x_m \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} x_1\, a_{11} + x_2\, a_{21} + \cdots + x_m\, a_{m1} & \cdots & x_1\, a_{1n} + x_2\, a_{2n} + \cdots + x_m\, a_{mn} \end{pmatrix}.$$

Thus the same matrix defines two linear functions,

$$\mathbb{R}^n \longrightarrow \mathbb{R}^m, \quad \mathbf{x} \longmapsto A\mathbf{x}$$

and

$$\mathbb{R}^m \longrightarrow \mathbb{R}^n, \quad \mathbf{x} \longmapsto \mathbf{x}A.$$

These two linear maps are called *adjoint maps*. If one of them is denoted by A the other is denoted by A^*.

DEFINITION 28 (**Adjoint operators, Transpose matrices**). Let $A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear operator induced by multiplication from the left by a matrix A. Then the *adjoint* of A, is the operator

$$A^*\colon \mathbb{R}^m \longrightarrow \mathbb{R}^n, \quad \mathbf{x} \longmapsto \mathbf{x}A.$$

The *transpose* of an $m \times n$ matrix A, denoted by A^* is the $n \times m$ matrix with row vectors equal to the column vectors of A, or equivalently, column vectors equal to the row vectors of A. In other words, if a_{ij} and a^*_{ij} are the elements in the i-th row and j-th column of A, and A^* respectively, then

$$a^*_{ij} = a_{ji}.$$

EXAMPLE 62. Consider the 4×3 matrix

$$A = \begin{pmatrix} -1 & 42 & 11 \\ 3 & 5 & -10 \\ 0 & 6 & 4 \\ 7 & -2 & 0 \end{pmatrix}.$$

The transpose of A is the 3×4 matrix A^*, with $a^*_{11} = a_{11}$, $a^*_{12} = a_{21}$, $a^*_{13} = a_{31}$, and $a^*_{14} = a_{41}$, and so on. Thus,

$$A^* = \begin{pmatrix} -1 & 3 & 0 & 7 \\ 42 & 5 & 6 & -2 \\ 11 & -10 & 4 & 0 \end{pmatrix}.$$

Of course, if we transpose the transpose, we'll get a matrix with rows the columns of A^*, that is the rows of A. Two matrices with the same rows are of course equal so

$$(A^*)^* = A.$$

For our example, we see that indeed,

$$(A^*)^* = \begin{pmatrix} -1 & 42 & 11 \\ 3 & 5 & -10 \\ 0 & 6 & 4 \\ 7 & -2 & 0 \end{pmatrix}.$$

Now because, the number of columns of A is equal to the number of rows of A^* the multiplication $A\,A^*$ is defined and the product is a 3×3 matrix. But also the number of columns of A^* equals to the number of rows of A, the multiplication $A^*\,A$ is also defined with product a 4×4 matrix.

Notice that both of these matrices are square matrices, but of different dimension. We have

$$A\,A^* = \begin{pmatrix} 59 & -41 & -41 \\ -41 & 1829 & 436 \\ -41 & 436 & 237 \end{pmatrix}, \quad A^*\,A = \begin{pmatrix} 1886 & 97 & 296 & -91 \\ 97 & 134 & -10 & 11 \\ 296 & -10 & 52 & -12 \\ -91 & 11 & -12 & 53 \end{pmatrix}.$$

Notice that both of these square matrices are *symmetric*, their rows are identical with their columns.

Now let's look at A as a linear operator, $A\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^4$. What is A^* the *adjoint* linear operator?

$$A^*\colon \mathbb{R}^4 \longrightarrow \mathbb{R}^3, \mathbf{x} \longmapsto \mathbf{x}\,A.$$

What's the matrix of A^*. It's columns are the images of the basic vectors $\mathbf{e}_i$ for $i = 1, 2, 3, 4$. We have,

$$\begin{aligned} \mathbf{e}_1 \longmapsto \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} -1 & 42 & 11 \\ 3 & 5 & -10 \\ 0 & 6 & 4 \\ 7 & -2 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -1\cdot 1 + 3\cdot 3 + 0\cdot 0 + 7\cdot 0 & 42\cdot 1 + 5\cdot 0 + 6\cdot 0 - 2\cdot 0 & 11\cdot 1 - 10\cdot 0 + 4\cdot 0 + 0\cdot 0 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 42 & 11 \end{pmatrix} \end{aligned}$$

Entirely similarly,

$$\mathbf{e}_2 \longmapsto (3, 5, 10), \quad \mathbf{e}_3 \longmapsto (0, 6, 4), \quad \mathbf{e}_4 \longmapsto (7, -2, 0).$$

Thus the columns of the matrix of the adjoint operator, has columns equal to the rows of A. Thus the matrix of the adjoint operator is the transpose of the matrix of A.

Let's also look at the reduced echelon forms of A and A^*.

$$A = \begin{pmatrix} -1 & 42 & 11 \\ 3 & 5 & -10 \\ 0 & 6 & 4 \\ 7 & -2 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

and,

$$A^* = \begin{pmatrix} -1 & 3 & 0 & 7 \\ 42 & 5 & 6 & -2 \\ 11 & -10 & 4 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & -\frac{292}{113} \\ 0 & 1 & 0 & \frac{361}{113} \\ 0 & 0 & 1 & \frac{3411}{226} \end{pmatrix}.$$

We notice that A, and A^* have the same rank.

EXAMPLE 63 (**Column and Row vectors as matrices**). So far we have treated $m \times 1$ and $1 \times n$ matrices as column and row *vectors*, respectively. Let's now look at them as matrices, and what operations they induce.

Let $\mathbf{a}$ be an $n \times 1$ matrix, then it induces the linear operation that sends 1-vectors to n-vectors.

$$\mathbf{a}\colon \mathbb{R}^1 \longrightarrow \mathbb{R}^n, \qquad \mathbf{a}\,\mathbf{x} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \begin{pmatrix} x \end{pmatrix} = \begin{pmatrix} a_1\,x \\ \vdots \\ a_n\,x \end{pmatrix} = x\,\mathbf{a}.$$

Thus, if we identify $\mathbb{R}^1$ with $\mathbb{R}$, then we see that the operation induced by $\mathbf{a}$ as a matrix, sends a real number x to x times the vector $\mathbf{a}$.

We can think of this as introducing coordinates in the line determined by $\mathbf{a}$, where 1 corresponds to $\mathbf{a}$.

The adjoint of $\mathbf{a}$ on the other hand, is induced by acting by $\mathbf{a}$ from the right, so

$$\mathbf{a}^*\colon \mathbb{R}^n \longrightarrow \mathbb{R}^1, \qquad \mathbf{x} \longmapsto \mathbf{a}\,\mathbf{x} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} x_1\,a_1 + \cdots + x_n\,a_m \end{pmatrix} = \mathbf{x} \cdot \mathbf{a}.$$

Thus $\mathbf{a}^*$, as an operator, sends a vector to its dot product with $\mathbf{a}$. Now since the standard basis is orthonormal, we have

$$\mathbf{e}_i \cdot \mathbf{a} = a_i,$$

and we see that the matrix of $\mathbf{a}^*$ is a row vector, with the same coordinates as $\mathbf{a}$, and of course, the transpose of $\mathbf{a}$ as a matrix.

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \Longrightarrow \mathbf{a}^* = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}$$

$$\mathbf{a} = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \Longrightarrow \mathbf{a}^* = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Adjoint Operators, Transpose matrices

The transpose of an $m \times n$ matrix, is an $n \times m$ matrix, such that for any $m \times 1$ matrix $\mathbf{x}$ we have:

$$A^*\,\mathbf{x} = \left(\mathbf{x}^*\,A\right)^*. \tag{3.13}$$

Using Equation (3.13) we can prove the following property of the transpose.

THEOREM 3.5.1 (**Transpose is an anti-homomorphism**). *The following hold.*

(a) Transpose respects scalar multiplication. That is, for any scalar λ

$$\left(\lambda\,A\right)^* = \lambda\,A^*.$$

(b) *Transpose respects matrix addition. That is,*

$$(A+B)^* = A^* + B^*.$$

(c) *If $A\,B$ is defined then $B^*\,A^*$ is also defined and*

$$(A\,B)^* = B^*\,A^*. \tag{3.14}$$

PROOF. The proof of (1) and (2) are straightforward and left as an exercise.

For the third, we will use Equation (3.13). To prove that two functions are equal, we have to prove that they take the same values on all elements of their domain. So, we have

$$\begin{aligned}(A\,B)^*\,\mathbf{x} &= \mathbf{x}^*\,(A\,B)\\ &= (\mathbf{x}^*\,A)\,B\\ &= B^*\,(\mathbf{x}^*\,A)^*\\ &= B^*\,(A^*\,\mathbf{x})\\ &= (B^*\,A^*)\,\mathbf{x}.\end{aligned}$$

□

Equation (3.14) has the same structure as property (2) in Theorem 3.4.5. Transposing, just like inverting, doesn't preserve multiplication but it doesn't totally destroy it either, it just reverses the order of the factors.

If $A \in \mathbf{M}_n$ (i.e it is an $n \times n$ square matrix) then $A^* \in \mathbf{M}_n$ as well. In that case A^k is defined for $k \in \mathbb{N}$ and the following holds.

THEOREM 3.5.2. *We have for $k \in \mathbb{N}$.*

(a) *If $A \in \mathbf{M}_n$ then*

$$\left(A^k\right)^* = (A^*)^k.$$

(b) *If $p(x)$ is any polynomial then*

$$p(A^*) = (p(A))^*.$$

PROOF. The first follows from Equation (3.14), and the fact that all powers of the same matrix commute, using induction. For $k = 0$ both sides are equal to the identity matrix so the statement is true. Now assume that we have proved it for k. Then we have

$$\left(A^{k+1}\right)^* = \left(A^k\,A\right)^* = A^*\,\left(A^k\right)^* = A^*\,(A^*)^k = (A^*)^{k+1}.$$

Evaluating a polynomial at a matrix involves scalar multiplication, matrix addition, and powers of matrices. We have seen that transposing respects all of these operations and the result follows. More formally, let

$$p(x) = a_d\,x^d + \cdots + a_1\,x + a_0\,x^0.$$

Then

$$\begin{aligned}p\left(A^*\right) &= a_d\,(A^*)^d + \cdots + a_1\,(A^*)^1 + a_0\,(A^*)^0\\ &= a_d\,\left(A^d\right)^* + \cdots + a_1\,A^* + a_0\,I\\ &= \left(a_d\,A^d\right)^* + \cdots + (a_1\,A)^* + (a_0\,I)^*\\ &= \left(a_d\,A^d + \cdots + a_1\,A + a_0\,I\right)^*\\ &= (p(A))^*.\end{aligned}$$

□

Furthermore, as the following Theorem shows, if A is invertible so is its transpose.

THEOREM 3.5.3 (**Transposing and Inverting commute**). *If A is invertible then so is A^*. Furthermore*

$$\left(A^*\right)^{-1} = \left(A^{-1}\right)^* .$$

PROOF. We have:

$$\begin{aligned} A\,B = I &\implies (A\,B)^* = I^* \\ &\implies B^*\,A^* = I. \end{aligned}$$

And the result follows from Lemma 3. □

DEFINITION 29 (**Symmetric and orthogonal matrices**). A *square* $n \times n$ matrix A is called *symmetric* if

$$A = A^*. \tag{3.15}$$

Equivalently, for all $i, j \in \{1, \ldots, n\}$ we have

$$a_{ij} = a_{ji}.$$

A linear operator $A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ that satisfies Condition (3.15), is said to be *self-adjoint*. The set of symmetric $n \times n$ matrices is denoted by $\mathbf{S}_n$.

A square $n \times n$ matrix is said to be *orthogonal* if

$$A\,A^* = I. \tag{3.16}$$

A linear operator $A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ that satisfies Condition (3.16), is said to be a *(linear) isometry*, or an *orthogonal* transformation.

The set of orthogonal $n \times n$ matrices is denoted by $\mathrm{O}(n)$ and called the *orthogonal group* of n-dimensional space.

The reason for the terminology *symmetric* should be clear. The entries a_{ij} and a_{ji} are in symmetric position with respect to the main diagonal, so when they are equal there is a symmetry in the matrix. Consider the symmetric matrix $A\,A^*$ of Example 62, we can see the symmetry by coloring symmetric entries with the same color:

$$\begin{pmatrix} 59 & -41 & -41 \\ -41 & 1829 & 436 \\ -41 & 436 & 237 \end{pmatrix},$$

The reason for the terminology *orthogonal* is that the columns of an orthogonal matrix A form *an orthonormal basis* of $\mathbb{R}^n$. When we officially introduce the dot product we will explore this property further. For now let's us just state the following Proposition.

PROPOSITION 8. *A is orthogonal if and only if*

$$\mathbf{a}_i \cdot \mathbf{a}_j = \delta_{ij}.$$

PROOF. The element in the i, j column of $A^*\,A$ is the inner product of the i-th row of A^* and the j-th row of A. But the i-th row of A^* is the i-th column of A. □

PROPOSITION 9. *If A is an $m \times n$ matrix then $A\,A^*$ and $A^*\,A$ are symmetric matrices.*

PROOF. We have

$$\left(A^*\,A\right)^* = A^*\,\left(A^*\right)^* = A^*\,A.$$

Similarly, we have

$$\left(A\,A^*\right)^* = A\,A^*.$$

□

We will prove that the set of symmetric matrices is closed under scalar multiplication and matrix addition. But first we note that in general $\mathbf{S}_n$ is not closed under matrix multiplication. That is, if A and B are symmetric their product is not necessarily symmetric. For example, let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & -1 & 0 \\ 3 & 0 & -4 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & -1 & -5 \\ -1 & 0 & 4 \\ -5 & 4 & -4 \end{pmatrix}.$$

two symmetric matrices.

Now we calculate,

$$A\,B = \begin{pmatrix} -14 & 11 & -9 \\ 7 & -2 & -14 \\ 29 & -19 & 1 \end{pmatrix}$$

and we see that $A\,B$ is not symmetric.

$B\,A$ is not symmetric either:

$$B\,A = \begin{pmatrix} -14 & 7 & 29 \\ 11 & -2 & -19 \\ -9 & -14 & 1 \end{pmatrix}.$$

However, the sum $A\,B + B\,A$ is symmetric:

$$A\,B + B\,A = \begin{pmatrix} -28 & 18 & 20 \\ 18 & -4 & -33 \\ 20 & -33 & 2 \end{pmatrix}.$$

THEOREM 3.5.4. *If $\lambda \in \mathbb{R}$ and $A, B \in \mathbf{M}_n$ then:*

(a) $A \in \mathbf{S}_n \implies \lambda\,A \in \mathbf{S}_n$.
(b) $A, B \in \mathbf{S}_n \implies A + B \in \mathbf{S}_n$.
(c) $A, B \in \mathbf{S}_n \implies (A\,B)^* = B\,A$.
(d) $A, B \in \mathbf{S}_n \implies A\,B + B\,A \in \mathbf{S}_n$.

PROOF. If $A^* = A$ and $B^* = B$ we have:

(a) $(\lambda\,A)^* = \lambda\,A^* = \lambda\,A$.
(b) $(A + B)^* = A^* + B^* = A + B$.
(c) $(A\,B)^* = B^*\,A^* = B\,A$.
(d) $(A\,B + B\,A)^* = (A\,B)^* + (B\,A)^* = B^*\,A^* + A^*\,B^* = B\,A + A\,B = A\,B + B\,A$.

□

The set of orthogonal matrices on the other hand is closed under taking inverses and under matrix multiplication. In, other words, $\mathrm{O}(n)$ is a *subgroup* of $\mathrm{GL}(n)$.

THEOREM 3.5.5. *If $\lambda \in \mathbb{R}$ and $A, B \in \mathbf{M}_n$ then:*

(a) $A \in \mathrm{O}(n) \implies A^{-1} \in \mathrm{O}_n$.
(b) $A, B \in \mathrm{O}(n) \implies A\,B \in \mathrm{O}(n)$.

PROOF. To prove that a matrix is orthogonal we have to prove that its inverse is equal to its transpose.

(a) We have

$$\begin{aligned} A^{-1} = A^* &\implies \left(A^{-1}\right)^* = (A^*)^* \\ &\implies \left(A^{-1}\right)^* = \left(A^{-1}\right)^{-1}. \end{aligned}$$

Therefore A^{-1} is orthogonal.

(b) Let A and B be two orthogonal matrices. Then we have

$$\begin{aligned}(A\,B)^{-1} &= B^{-1}\,A^{-1}\\ &= B^*\,A^*\\ &= (A\,B)^*\,.\end{aligned}$$

□

3.5.1. The rank of the transpose. In Section 3.1 we saw that a matrix $A \in \mathbf{M}_{m\times n}$ defines a linear map $A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ given by

$$\mathbf{x} \longmapsto A\,\mathbf{x},$$

where $\mathbf{x} \in \mathbb{R}^n$ is considered a column vector. We concluded (among other things) that the range of this linear is spanned by the columns of A, namely

$$\mathbf{y} = A\,\mathbf{x} \iff \mathbf{y} = \sum_{i=1}^{n} x_i\,\mathbf{a}_i.$$

Entirely similar arguments show that the range of the adjoint linear map A^* is spanned by the rows of A, namely

$$\mathbf{x} = \mathbf{y}\,A \iff \mathbf{x} = \sum_{i=1}^{m} y_i\,\mathbf{a}_i^*.$$

Another way to see this is to recall that the columns of the transpose matrix A^* are the rows of A. Either way we have the following Theorem.

THEOREM 3.5.6. *The range of A^* is spanned by the rows of A. Therefore the rank of A^* is the dimension of the linear span of the rows of A.*

Now recall that a basis for the range of A consists of the basic columns of A, that is the columns that contain a leading 1 in the reduced echelon form of A. Now if $\mathbf{a}_i$ is a basic column, then the row that contains the leading one has all zero entries to the left of the leading 1. Therefore all these rows are linearly independent.

Therefore there are *at least* $\operatorname{rank} A$ linearly independent rows. Therefore we conclude that the dimension of the linear span of the rows of A is greater of equal to the rank of A. So

$$\operatorname{rank} A \le \operatorname{rank} A^*.$$

Applying this to A^*, whose transpose is A, we conclude that

$$\operatorname{rank} A^* \le \operatorname{rank} A,$$

as well.

Therefore we have proved the following theorem.

THEOREM 3.5.7 (**Transpose matrices have the same rank**). *We have*

$$\operatorname{rank} A = \operatorname{rank} A^*.$$

When we restrict attention to square matrices we obtain the following corollary.

COROLLARY 3. *Let $A \in \mathbf{M}_n$ be a square matrix. Then A is invertible if and only if A^* is invertible.*

We already knew that of course, see Theorem 3.5.3.

3.6. Elementary matrices and row (or column) operations

We have seen two ways of solving systems of linear equations. In Section 1.1 we developed the method of using elementary row operation to get the (augmented) matrix of the system to a (reduced) echelon form. On the other hand, Theorem 3.3.3 suggests another way, assuming that the matrix of the system is invertible: just multiply the vector of constants with the inverse of the matrix. In other words, the solution of

$$A\mathbf{x} = \mathbf{c},$$

is

$$\mathbf{x} = A^{-1}\mathbf{c},$$

The second method, in practice, is not really that different, since our method of finding the inverse of a matrix involves row operations anyway (see Example 42). In this section we will see that row operations can be thought of as multiplication with some special matrices: when we use row operations we still multiply with the inverse of the matrix, but we do it in several steps.

Recall that there are three types of elementary operations:

(a) **Type I:** Interchange two rows.
(b) **Type II:** Multiply a row by a non-zero scalar.
(c) **Type III:** Add a row to an other row.

DEFINITION 30 (**Elementary Matrices**). An $n \times n$ matrix resulting from the application of a row operation to the identity matrix I_n is called an *elementary matrix* of the same type as the row operation.

EXAMPLE 64. The following are 4×4 elementary matrices of type I, II, and III respectively:

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Indeed in the first we have interchanged the second and third row, in the the second we multiplied the third row by -2, and the for the last we added the fourth row to the first.

THEOREM 3.6.1. *Let E be an elementary $n \times n$ matrix and A an $n \times m$ matrix. Then EA is obtained by performing to A the same elementary row operation that was performed to I to get E.*

REMARK 15. We have already seen this for Type II elementary matrices. Indeed those are diagonal matrices with all diagonal entries except one equal to 1 and one diagonal entry equal to a number λ, and the effect of multiplying with a diagonal matrix was discussed in Example 49.

PROOF. Recall that the i-th row of the product AB consists of the dot products of the i-th row of A with the columns of B. This means that the i-th row of the product AB depends only on the i-th row of A and no other rows.

Now, let E an elementary matrix of Type I where the rows k and ℓ of I have been interchanged. Then if $i \neq k, \ell$ the i-th row of E is the same as the i-th row of I and therefore the i-th row of the product EA is the same as the i-th row of the product IA, that is the i-th row of A. On the other hand the k-th row of E is the ℓ-th row of I, hence the k-th row of EA equals the ℓ-th row of $IA = A$. Similarly, the ℓ-th row of EA equals the k-th row of A.

If E is obtained by I by multiplying the row k by λ then all rows of $E\,A$ except the k-th are the same as the rows of A. On the other hand, the j-th entry of the k-th row of $E\,A$ is the dot product

$$(\lambda\,\mathbf{e}_k)\cdot\mathbf{a}_j = \lambda\,a_{kj}.$$

If E is an elementary matrix of type III, obtained, say, by adding the k-th row to the ℓ-th row, then all the rows of E, except the ℓ-th, are the same as the rows of I thus all the rows of $E\,A$, except the ℓ-th, are the same as the rows of A. On the other hand the ℓ-th row of E is $\mathbf{e}_k + \mathbf{e}_\ell$ and so the j-the entry of the ℓ-row is

$$(\mathbf{e}_k + \mathbf{e}_\ell)\cdot\mathbf{a}_j = a_{kj} + a_{\ell j}.$$

□

Next we prove that all elementary matrices are invertible. But before that let's introduce some notation.

DEFINITION 31. The elementary matrix obtained by interchanging the k-th and ℓ-th rows of I will be denoted by $P_{k\ell}$, the one obtained by multiplying the k-th row by λ will be denoted by $M_{k;\lambda}$, and the one obtained by adding the k-th row to the ℓ-th row by $S_{k\ell}$.

THEOREM 3.6.2. *All elementary matrices are invertible. Namely,*

(a) $P_{k\ell}^{-1} = P_{k\ell}$.
(b) $M_{k;\lambda}^{-1} = M_{k;\lambda^{-1}}$.
(c) $S_{k\ell}^{-1} = M_{k;-1}\,S_{k\ell}M_{k;-1}$.

PROOF. By Theorem 3.6.2 when multiplying $P_{k\ell}$ with $P_{k\ell}$ interchanges the k-th and ℓ-th rows of $P_{k\ell}$ and so

$$P_{k\ell}^2 = I.$$

Similarly,

$$M_{k;\lambda^{-1}}\,M_{k;\lambda} = I.$$

For (3) notice that $M_{k;-1}\,S_{k\ell}M_{k;-1}\,A$ has the effect of subtracting the k-th row of A from its ℓ-th row. Indeed $M_{k;-1}$ multiplies the k-th row by -1, then $S_{k\ell}$ adds it to the ℓ-th row, and finally $M_{k;-1}$ multiplies the k-th row by -1 again reverting it to the original row of A. Therefore,

$$M_{k;-1}\,S_{k\ell}M_{k;-1}\,S_{k\ell} = I.$$

□

So applying an elementary row operation to the augmented matrix of a system is equivalent to multiplying, from the left, both sides of the corresponding vector equation by an elementary matrix. Let's reconsider the 3×3 system of Example 5[18]

$$\begin{cases} x + y + z = 3 \\ x - y + z = 1 \\ 4x + 2y + z = 10 \end{cases}.$$

The corresponding vector equation is

$$A\,\mathbf{x} = \mathbf{c} \tag{3.17}$$

where

[18] I changed the variables to x, y and z.

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 4 & 2 & 1 \end{pmatrix}, \text{ and } \mathbf{c} = \begin{pmatrix} 3 \\ 1 \\ 10 \end{pmatrix}.$$

We started by subtracting the first row from the second. This is really a combination of two elementary row operations: we first multiplied the second row with -1 and then we replaced the second row by the sum of the first and second row. In terms of elementary matrices this is equivalent to first multiply Equation (3.17) with $M_{2;-1}$ and then by S_{12}.

$$A\,\mathbf{x} = \mathbf{c} \iff M_{2;-1}\,(A\,\mathbf{x}) = M_{2;-1}\,\mathbf{c} \iff S_{12}\,M_{2;-1}\,(A\,\mathbf{x}) = S_{12}\,M_{2;-1}\,\mathbf{c}.$$

Then we subtracted 4 times the first equation from the third. This is equivalent to the composition of four elementary operations: multiply the third equation by -1, then the first by 4, then add the first equation to the third, and finally multiply the first equation by $1/4$. In terms of elementary matrices we multiplied both sides of the vector equation with $M_{1;1/4}\,S_{13}\,M_{1;4}\,M_{3;-1}$ to get

$$M_{1;1/4}\,S_{13}\,M_{1;4}\,M_{3;-1}\,S_{12}\,M_{2;-1}\,(A\,\mathbf{x}) = M_{1;1/4}\,S_{13}\,M_{1;4}\,M_{3;-1}\,S_{12}\,M_{2;-1}\,\mathbf{c}.$$

Continuing in this fashion we see that overall we multiplied the vector equation by the matrix

$$B := M_{2;-1}\,M_{3;-1}\,S_{21}\,M_{2;1/2}\,S_{31}\,M_{3;1/3}\,S_{2;3}\,M_{3;-1}\,M_{1;1/4}\,S_{13}\,M_{1;4}\,M_{3;-1}\,S_{12}\,M_{2;-1}.$$

In other words, the whole process of Gauss-Jordan elimination can be summarized as

$$A\,\mathbf{x} = \mathbf{c} \iff B\,A\,\mathbf{x} = B\,\mathbf{c}.$$

But since the echelon form of A turned out to be the identity matrix we have that $B\,A = I$, which means that $B = A^{-1}$.

The Gauss-Jordan elimination method is an efficient way of multiplying both sides of Equation (3.17) by A^{-1}.

The above discussion also gives an algebraic interpretation of the method of finding the inverse of a matrix exposed in Example 42. Namely, we see that the inverse of an invertible matrix is a product of elementary matrices. Since any invertible matrix is the inverse of its inverse we see that every invertible matrix is a product of elementary matrices.

Conversely, since elementary matrices are invertible, a matrix that is a product of elementary matrices is invertible. We thus have the following Theorem.

THEOREM 3.6.3 (**Elementary matrices generate** $\mathrm{GL}(n)$**).** *A square matrix is invertible if and only it can be written as a product of elementary matrices.*

EXAMPLE 65. The 3×3 matrix

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

is invertible. To express A as a product of elementary matrices we need to find a sequence of row operations that reduces it to the identity matrix.

We start by multiplying the adding -2 times the first row to the third row. This corresponds to the product $M_{1;-1/2}\,S_{13}\,M_{1;-2}$. Then we add 3 times the second row to the third. This corresponds to the product $M_{2;1/3}\,S_{23}\,M_{2;3}$. Then we divide the third row by -5. This is accomplished by $M_{3;-1/5}$. This turns A into an upper triangular matrix.

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 2 & -1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & -3 & -2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & -5 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

The last matrix is equal to the product

$$M_{3;-1/5}\,M_{2;1/3}\,S_{23}\,M_{2;3}\,M_{1;-1/2}\,S_{13}\,M_{1;-2}\,A.$$

Next we add the third row to the second. This is accomplished by S_{32} . We then multiply the first row by -1 (corresponding to $M_{1;-1}$), add the third and then the second row to the first ($S_{31}\,S_{21}$), and finally we multiply the first row by -1 ($M_{1;-1}$).

$$\sim \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} -1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} -1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

So we have

$$M_{1;-1}\,S_{21}\,S_{31}\,M_{1,-1}\,S_{32}\,M_{3;-1/5}\,M_{2;1/3}\,S_{23}\,M_{2;3}\,M_{1;-1/2}\,S_{13}\,M_{1;-2}\,A = I.n$$

Therefore

$$\begin{aligned} A &= \left(M_{1;-1}\,S_{21}\,S_{31}\,M_{1,-1}\,S_{32}\,M_{3;-1/5}\,M_{2;1/3}\,S_{23}\,M_{2;3}\,M_{1;-1/2}\,S_{13}\,M_{1;-2}\right)^{-1} \\ &= M_{1;-1/2}\,M_{1;-1}\,S_{13}\,M_{1;-1}\,M_{1,-2}\,M_{2;1/3}\,M_{2;-1}\,S_{23}\,M_{2;-1}\,M_{23}\,M_{3;-5}\,M_{3;-1}\,S_{32}\,M_{3;-1}\,M_{2;-1}\,S_{21}\,M_{2;-1}\,M_{1,-1}. \end{aligned}$$

Notice that even for a relatively small matrix we get a rather complicated expression. We could simplify the expression a bit by noticing that, for example

$$M_{1;-1/2}\,M_{1;-1} = M_{1;1/2}$$

because multiplying the same row two consecutive times can be done with one step. We get a simpler, but still complicated expression:

$$A = M_{1;1/2}\,S_{13}\,M_{1;2}\,M_{2;-1/3}\,S_{23}\,M_{2;3}\,M_{3;5}\,S_{32}M_{3;-1}\,M_{2;-1}\,S_{21}\,M_{2;-1}\,M_{1,-1}.$$

REMARK 16. As the previous example demonstrates using elementary matrices to compute inverses is not really that practical. Computing with row operations, as we have been doing so far is more efficient. This doesn't mean that elementary matrices are useless though, looking at the same topic from different points of view increases our understanding and leads to new insights that could be much harder to reach otherwise.

Recall (see Definition 3) that two matrices A and B are called *row equivalent* if there is a finite sequence of elementary row operations that turn A on to be. By our discussion so far we have the following theorem.

THEOREM 3.6.4. *Two $m \times n$ matrices A, B are row equivalent if and only if there an invertible $m \times m$ matrix C such that*

$$B = C\,A.$$

3.6.1. Column operations. We introduced row operations in Section 1.1 as operations on the equations of a linear system. When we later introduced the vector form of a system, equations corresponded to rows of the matrix and so these operations ended up to act on the rows of the matrix. We represented the system

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & c_2 \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & c_m \end{cases}$$

as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}.$$

But that was a choice. We could also have represent it as

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \cdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} c_1 & c_2 & \cdots & c_m \end{pmatrix}.$$

In other words, we represented a vector $\mathbf{x} \in \mathbb{R}^n$ as a column but we could have represented it as row instead. Had we done that, the matrix of the system would have been the transpose A^* instead of A. After all,

$$A\mathbf{x} = \mathbf{c} \iff \mathbf{x}^* A^* = \mathbf{c}^*.$$

Perhaps we made that choice in a parallel universe. In that universe, the equations of the system would correspond to the columns of its matrix and the variables would correspond to its rows, and we would talk about *elementary column operations* and *(reduced) column echelon form*.

Of course, such a choice doesn't really change the system, or its solution set, a vector doesn't care whether we write it as a column or as a row, it's the same vector either way. So in that hypothetical universe[19] there would be a theory of linear systems that would get the same results by using elementary column operations. The elementary matrices that would represent their column operations would be the same as our elementary matrices though, just applied on the right of a matrix not on the left.

Column operations and column equivalence are completely analogous to row operations and row equivalence. Rather than copying the definitions changing "row" to column we develop them from an algebraic point of view starting with the analog of Theorem 3.6.4.

DEFINITION 32 (**Column Equivalence**). We say that $A, B \in \mathbf{M}_{m\times n}$ are *column equivalent* if

$$B = AC$$

for some invertible $n \times n$ matrix C.

For the remaining of this section, let's use the notation

$$A \cong B$$

to mean that A is column equivalent to B.

[19] This is not pure science fiction. There are books where this choice is made.

THEOREM 3.6.5. *Column equivalence is an equivalence relation. In other words if X, Y, Z are $m \times n$ matrices, we have:*

(a) $X \cong X$.
(b) $X \cong Y \implies Y \cong Z$.
(c) $X \cong Y$ *and* $Y \cong Z \implies X \cong Z$.

PROOF. (1) holds because $X = X\,I$.
(2) follows from the implication $Y = X\,C \implies X = Y\,C^{-1}$.
To prove (3), notice that if $Y = X\,C$ and $Z = Y\,D$ then $Z = X\,(C\,D)$. Furthermore, if C, D are invertible then so is $C\,D$ (see Theorem 3.4.5). □

THEOREM 3.6.6. *Two matrices are column equivalent if and only if their transposes are row equivalent. In other words,*

$$A \cong B \iff A^* \sim B^*.$$

PROOF. We have (see Theorem 3.5.1)

$$B = A\,C \iff B^* = C^*\,A^*.$$

and C is invertible if and only if C^* is invertible (see Theorem 3.5.3). Thus, if $A \cong B$ then (by Theorem 3.6.4) $A^* \sim B^*$.

Conversely, if $A^* \sim B^*$ then, again by Theorem 3.6.4 we have that $B^* = C\,A^*$ for some invertible matrix C. But then $B = A\,C^*$ and therefore A, B are row equivalent. □

As a corollary we have the following Theorem.

THEOREM 3.6.7. *Let A and B be $m \times n$ matrices. Then $A \cong B$ if and only if B can be obtained from A by applying a finite sequence of* elementary column operations*:*

- *Interchanging two columns.*
- *Multiplying a column by a non-zero scalar.*
- *Adding a column to an other column.*

The elementary matrices P_{ij} and $M_{k;\lambda}$ are symmetric. This is obvious for $M_{k;\lambda}$ since it's a diagonal matrix. On the other hand, the ij and ji as well as the kk entries, for $k \neq i, j$, of P_{ij} are 1 and all other entries are 0, and so P_{ij} is symmetric.

For the third type of elementary matrices we have that the diagonal entries as well as the ℓk entry of $S_{k\ell}$ are 1 and all other entries are 0. Thus the transpose of $S_{k\ell}$ has the diagonal entries and the $k\ell$ entry 1, and all other entries 0. So the transpose of $S_{k\ell}$ is $S_{\ell k}$.

Thus the following theorem holds.

THEOREM 3.6.8. *We have*

(a) $P^*_{k\ell} = P_{k\ell}$.
(b) $M^*_{k;\lambda} = M_{k;\lambda}$.
(c) $S^*_{k\ell} = S_{\ell k}$.

THEOREM 3.6.9. *Let A be an $m \times n$ matrix. Then*

(a) $A\,P_{k\ell}$ *has the same columns as A with the k and ℓ columns interchanged.*
(b) $A\,M_{k;\lambda}$ *has the same columns as A except the k-th that is equal to λ times the k-th column of A.*
(c) $A\,S_{k\ell}$ *has the same columns as A except the k-th that is the sum of the k-th and ℓ-th columns of A.*

PROOF. We have

$$(A\,P_{k\ell})^* = P_{k\ell}\,A^*.$$

Thus the columns of $A\,P_{k\ell}$ are the rows of $P_{k\ell}\,A^*$, that is the rows of A^* with the k-th and ℓ-th rows interchanged. In other words the columns of A with the k-th and ℓ-th columns interchanged. This proves (1).

Similarly,

$$(A\,M_{k;\lambda})^* = M_{k;\lambda}\,A^*.$$

Thus the columns of $A\,M_{k;\lambda}$ are the rows of $M_{k;\lambda}\,A^*$, that is the rows of A^* with the k-th row multiplied by λ. In other words the columns of A with the k-th column multiplied by λ. This proves (2).

Finally,

$$(A\,S_{k\ell})^* = S_{\ell k}\,A^*.$$

Thus the columns of $A\,S_{k\ell}$ are the rows of $S_{\ell k}\,A^*$ that is the rows of A^* with the ℓ-th row added to the k-th. In other words the columns of A with the ℓ-th column added to the k-th. □

DEFINITION 33. We say that two $m \times n$ matrices A, B are *equivalent*, and write $A \approx B$, if there is an invertible $m \times m$ matrix C and an invertible $n \times n$ matrix D such that

$$B = C\,A\,D.$$

Equivalently if B can be obtained by applying a finite sequence of elementary row or column operations.

Exercise 8. Prove that $\approx$ is an equivalence relation.

THEOREM 3.6.10. *Any matrix A is equivalent to a block matrix of the form*

$$\left(\begin{array}{cccc|ccc} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right).$$

The number of non-zero rows (and columns) is $\operatorname{rank} A$.

PROOF. We use row operations to bring the matrix to its reduced row echelon form. Then we use column operations to put all the free columns at the end. Finally we use the leading 1 of each row to kill all non-zero entries on that row.

The number of non-zero columns is the number of basic columns in the reduced row echelon form of A and therefore equal to $\operatorname{rank} A$. □

EXAMPLE 66. Consider the matrix

$$A = \begin{pmatrix} 1 & -2 & -1 & 4 & 0 & -1 & -2 & 0 \\ 2 & -4 & -5 & 11 & 0 & -4 & -16 & -1 \\ -2 & 4 & 5 & -11 & 1 & 4 & 16 & 1 \\ 4 & -8 & -9 & 21 & -2 & -7 & -27 & -2 \\ -1 & 2 & 5 & -8 & 1 & 3 & 16 & 1 \\ 1 & -2 & -2 & 5 & -1 & 0 & -1 & -1 \end{pmatrix}.$$

Its reduced row echelon form is

$$A \sim \begin{pmatrix} 1 & -2 & 0 & 3 & 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We use column interchanges to move the free columns to the end:

$$A \approx \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & -2 & 3 & 3 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 2 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Finally, we add 2 (respectively -3, -3) times the first column to the fifth (respectively sixth, seventh), add the second column to the sixth, add -2 times the second column to the seventh, and finally -3 times the fourth column to the seventh to get

$$A \approx \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We conclude that $\operatorname{rank} A = 5$.

Exercise. Prove Theorem 3.5.7 using Theorem 3.6.10.

CHAPTER 4

Abstract Vector Spaces

In this chapter we abstract the algebraic properties of the standard real vector spaces $\mathbb{R}^n$. We start by abstracting the properties of addition and multiplication of the real numbers and get the concept of a *field*. Roughly speaking a field is a set whose elements we can add or multiply and all the algebraic manipulations that are valid for addition and multiplication of real numbers are still valid. In particular almost all the theory we developed in the previous three chapters works in all fields. There are standard n-dimensional vector spaces defined over any field, they have vector subspaces, matrices define linear maps and so on.

Then we go to the next level of abstraction by isolating the properties of scalar multiplication and vector addition of the standard vector spaces that make the theory we developed in the first three chapters work. This leads to the definition of an abstract vector space. A vector space is thus, roughly, a set whose elements ("vectors") can be added and multiplied by scalars, and these operations satisfy the vector space axioms.

These are very powerful abstractions with vast scope of applications. An abstract "vector" can be a geometric vector, a matrix, a function, the state of a quantum system, and so on.

4.1. Fields

DEFINITION 34 (**Fields)**. A set K endowed with two binary operations $+$ and $\cdot$ called *addition* and multiplication respectively is a field if the following properties, called the *field axioms* hold:

(a) Addition is commutative. That is, for all $a, b \in K$

$$a + b = b + a.$$

(b) Addition is associative. That is, for all $a, b, c \in K$

$$a + (b + c) = (a + b) + c.$$

(c) Addition has a *neutral element*. That is there exists an element $0 \in K$ such that for all $a \in K$

$$a + 0 = a.$$

(d) Every element has an *opposite* (or *negative*) element. This means that for every $a \in K$ there exists $-a \in K$ such that

$$a + (-a) = 0.$$

(e) Multiplication is commutative. That is, for all $a, b \in K$

$$a\, b = b\, a.$$

(f) Multiplication is associative. That is, for all $a, b, c \in K$

$$a\,(b\,c) = (a\,b)\,c.$$

(g) Multiplication has a *neutral element*. That is, there exists an element $1 \in K$ such that for all $a \in K$

$$1 \cdot a = a.$$

(h) Every non-zero element has an *inverse* (or *reciprocal*) element. This means that for every $a \in K$ there exists $a^{-1} \in K$ such that

$$a \cdot a^{-1} = 1.$$

(i) Multiplication distributes over addition. That is, for all $a, b, c \in K$ we have

$$a\,(b + c) = a\,b + a\,c.$$

It turns out that these nine axioms imply that we can do algebra in any field, pretty much the same way we do algebra with real numbers. For the rest of this section elaborates this vague remark.

There is only one 0 and only one 1 in any field, i.e. the elements that satisfy Properties (3) and (7) are unique. For, assume that for $0' \in K$ we have that for all $a \in K$

$$a + 0' = a.$$

Then for $a = 0$ we have

$$0 + 0' = 0.$$

On the other hand, by Property (3) for $a = 0'$ we have

$$0' + 0 = 0'.$$

But addition is commutative, and therefore $0 + 0' = 0' + 0$. It follows then that

$$0 = 0'.$$

The proof of the uniqueness of 1 is entirely similar.

The opposite and inverse of an element are also unique. Let's prove it for the inverse. Let $a \neq 0$ be an element of K and let a^{-1} be an element that satisfies Property (9) and let $b \in K$ be another element that satisfies

$$a\,b = 1.$$

We will prove, using the field axioms, that $b = a^{-1}$. Indeed,

$$\begin{aligned}
a\,b = 1 &\implies a^{-1}\,(a\,b) = a^{-1} \cdot 1 && \text{Multiply both sides by } a^{-1}\\
&\implies a^{-1}\,(a\,b) = 1 \cdot a^{-1} && \text{Using (5)}\\
&\implies a^{-1}\,(a\,b) = a^{-1} && \text{Using (7)}\\
&\implies \left(a^{-1}\,a\right)\,b = a^{-1} && \text{Using (6)}\\
&\implies \left(a \cdot a^{-1}\right)\,b = a^{-1} && \text{Using (5)}\\
&\implies 1 \cdot b = a^{-1} && \text{Using (8)}\\
&\implies b = a^{-1} && \text{Using (7).}
\end{aligned}$$

The following theorem summarizes some algebraic properties that follow from the field axioms and that we will be using freely from now on. We provide proofs of some of them and invite the reader to provide proofs of the rest and in genera to feel the details.

THEOREM 4.1.1. *Let K be a field. Then the following hold.*

(a) Let $a \in K$. If for some $b \in K$ we have $a + b = b$ then $a = 0$.

(b) Let $a \in K$. If for some $b \in K$ we have $a\,b = b$ and $b \neq 0$, then $a = 1$.

(c) For all $a \in K$ we have

$$-(-a) = a.$$

(d) For all $a \in K$,

$$0 \cdot a = 0.$$

(e) *For all $a \in K$,*

$$(-1)\ a = -a.$$

(f) *For all $a, b \in K$,*

$$-(a+b) = (-a) + (-b).$$

(g) *For all $a, b \in K$,*

$$a\,b = 0 \iff a = 0 \text{ or } b = 0.$$

(h) *For all $a \in K$ if $a \neq 0$ then $a^{-1} \neq 0$ and*

$$\left(a^{-1}\right)^{-1} = a.$$

(i) *If $a, b \in K$ with $a \neq 0$ and $b \neq 0$ then $a\,b \neq 0$ and*

$$(a\,b)^{-1} = a^{-1}\,b^{-1}.$$

PROOF. (a) Starting with $a + b = b$ we add $-b$ to both sides and get

$$(a+b) + (-b) = 0.$$

Using associative property we get

$$a + (b + (-b)) = 0,$$

and since $b + (-b) = 0$,

$$a + 0 = 0,$$

and finally

$$a = 0.$$

(b) Note that since $b \neq 0$ the inverse b^{-1} exists. The proof then proceeds as in (1): we multiply both sides by b^{-1}, and then use the properties of multiplication.

(c) We have $a + (-a) = 0$ and by commutativity of addition $(-a) + a = 0$. By the discussion about the uniqueness of the opposite it follows that the opposite of $-a$ is a.

(d) Since $0 + 0 = 0$ we have, by the distributive property

$$0\,a = (0+0)\,a = 0\,a + 0\,a,$$

that is, $0\,a + 0\,a = 0\,a$. The result then follows from (1).

(e) We have,

$$0 = 0\,a\,(1 + (-1))\ a = 1 \cdot a + (-1)\,a = a + (-1)\,a.$$

Adding $-a$ to both sides then yields the result.

(f) By (5) this is equivalent to

$$(-1)\,(a+b) = (-1)\,a + (-1)\,b$$

which holds by the distributivity of multiplication.

(g) If $a = 0$ or $b = 0$ then by (4) we have $a\,b = 0$.

To prove the converse, we'll assume $a\,b = 0$ and prove that if $b \neq 0$ we must have $a = 0$, thus establishing the result. Now, if $b \neq 0$ the inverse b^{-1} exists and we can multiply both sides of $a\,b = 0$ by b^{-1}, to get

$$(a\,b)\,b^{-1} = 0 \cdot b^{-1}$$

which using distributivity of multiplication and (4) gives

$$a\,(b\,b^{-1}) = 0.$$

And this using the field axioms (8) and (7) gives

$$a = 0.$$

(h) We will show that $(a\,b)\,(a^{-1}\,b^{-1}) = 1$. Here is the calculations in full detail:

$$\begin{aligned}
(a\,b)\,\left(a^{-1}\,b^{-1}\right) &= \left((a\,b)\,a^{-1}\right)\,b^{-1} && \text{(Using distibutivity)}\\
&= \left((b\,a)\,a^{-1}\right)\,b^{-1} && \text{(Using commutativity)}\\
&= \left(b\,\left(a\,a^{-1}\right)\right)\,b^{-1} && \text{(Using distibutivity)}\\
&= (b\,1)\,b^{-1} && \text{(Using Axiom (8))}\\
&= (1\,b)\,b^{-1} && \text{(Using commutativity)}\\
&= b\,b^{-1} && \text{(Using Axiom (7))}\\
&= 1 && \text{(Using Axiom (8)).}
\end{aligned}$$

□

The associative property of addition means that if $a, b, c \in K$ the triple sum

$$a + b + c$$

is well defined and we don't need to write parentheses. More generally, we can prove that for any $a_1, \ldots, a_n$ the sum

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n$$

is well defined and gives the same answer no matter how we insert the parentheses. For example:

$$a_1 + ((a_2 + a_3) + (a_4 + a_5)) = (a_1 + a_2) + (a_3 + (a_4 + a_5))\,.$$

Similarly the product of $a_1, \ldots, a_n$

$$\prod_{i=1}^{n} a_i = a_1\,a_2 \cdots a_n$$

is well defined, no matter how we parenthesize we'll get the same answer. From now on we will not write parentheses for multiple sums and products.

DEFINITION 35 (**Subtraction and Division in a field)**. If K is a field and $a, b \in K$ then the *difference of a and b* is defined via

$$a - b = a + (-b).$$

If $b \neq 0$ we define the *quotient of a and b* via

$$\frac{a}{b} = a\,b^{-1}.$$

As an example of how our familiar algebraic manipulations work in any field we prove the following proposition.

PROPOSITION 10. *If K is a field and $a, b, c, d \in K$ with $b \neq 0$ and $d \neq 0$ then*

$$\frac{a}{b} + \frac{c}{d} = \frac{a\,d + b\,c}{b\,d}.$$

PROOF. We have

$$\begin{aligned}
\left(a\,b^{-1} + c\,d^{-1}\right)\,b\,d &= a\,b^{-1}\,b\,d + c\,d^{-1}\,b\,d\\
&= a\,1\,d + c\,d^{-1}\,d\,b\\
&= a\,d + c\,1\,b\\
&= a\,d + c\,b.
\end{aligned}$$

Thus we have

$$\left(a\,b^{-1}+c\,d^{-1}\right)\,b\,d = a\,d + c\,b.$$

Multiplying both sides with $(b\,d)^{-1}$ gives

$$a\,b^{-1}+c\,d^{-1} = (a\,d+c\,b)\ (b\,d)^{-1}.$$

The last equation is equivalent to

$$\frac{a}{b}+\frac{c}{d}=\frac{a\,d+b\,c}{b\,d}.$$

□

Finally in any field we can define the powers a^n with $n \in \mathbf{Z}$.

DEFINITION 36 (**Integer powers in a field**). Let $a \in K$. For $n \geq 0$ we define recursively

$$\begin{cases} a^0 & = 1 \\ a^{n+1} & = a^n\,a \end{cases}.$$

If $a \neq 0$ and $n \geq 0$ we also define

$$a^{-n} = \left(\frac{1}{a}\right)^n.$$

The familiar properties of powers hold in any field. We list some of them in the following theorem. The proofs are straightforward and are left as an exercise.

THEOREM 4.1.2. *Let K be a field, $a, b \in K$ and $m, n \in \mathbb{Z}$. Assuming that when the exponents are negative the base is non-zero, the following hold:*

(a) $a^m\,a^n = a^{m+n}$.
(b) $(a^m)^n = a^{m\,n}$.
(c) $(a\,b)^n = a^n\,b^n$.
(d) $\dfrac{a^n}{a^m} = a^{n-m}$.
(e) $\left(\dfrac{a}{b}\right)^n = \left(\dfrac{b}{a}\right)^{-n}$.

EXAMPLE 67 (**The reals**). The set of real numbers $\mathbb{R}$ endowed with the usual addition and multiplication is a field. Indeed all the field axioms hold.

EXAMPLE 68 (**The rationals**). The set of rational numbers $\mathbb{Q}$ endowed with the usual addition and multiplication is a field. Recall that a real number q is said to be rational if there are two integers m, n such that $q = m/n$ (n is non-zero of course).

Now notice that

(a) $0 \in \mathbb{Q}$.

Indeed we can write

$$0 = \frac{0}{1}$$

and since $0, 1 \in \mathbb{Z}$ it follows that $0 \in \mathbb{Q}$.

(b) $1 \in \mathbb{Q}$. Indeed, $1 = 1/1$.

(c) $\mathbb{Q}$ is closed under addition. That is,

$$q_1, q_2 \in \mathbb{Q} \implies q_1 + q_2 \in \mathbb{Q}.$$

Indeed, if $q_i = m_i/n_i$ with $m_i, n_i \in \mathbb{Z}$, and $n_i \neq 0$ for $i = 1, 2$ then

$$q_1 + q_2 = \frac{m_1\,n_2 + m_2\,n_1}{n_1\,n_2}.$$

Now the sums and products of integers are integers and therefore $m_1\,n_2 + m_2\,n_1 \in \mathbb{Z}$ and $n_1\,n_2 \in \mathbb{Z}$. Thus $q_1 + q_2 \in \mathbb{Q}$.

(d) $\mathbb{Q}$ is closed under opposites. That is,

$$q \in \mathbb{Q} \implies -q \in \mathbb{Q}.$$

Indeed if $q = m/n$ with $m, n \in \mathbb{Z}$ then

$$-q = \frac{-m}{n}$$

and since $-m$ is an integer we conclude that $-q \in \mathbb{Q}$.

(e) $\mathbb{Q}$ is closed under multiplication. That is,

$$q_1, q_2 \in \mathbb{Q} \implies q_1\, q_2 \in \mathbb{Q}.$$

Indeed,

$$q_1\, q_2 = \frac{m_1\, m_2}{n_1\, n_2}$$

and $\mathbb{Z}$ is closed under multiplication.

(f) $\mathbb{Q}$ is closed under inverses. That is,

$$q \in \mathbb{Q}, \quad q \neq 0 \implies q^{-1} \in \mathbb{Q}.$$

Indeed if $q = m/n$ with $m, n \in \mathbb{Z}$ then if $q \neq 0$ we have that $m \neq 0$. Then

$$q^{-1} = \frac{n}{m} \in \mathbb{Q}.$$

Since $\mathbb{Q} \subseteq \mathbb{R}$ and $\mathbb{R}$ is a field we conclude that all the field axioms hold for Q.

EXAMPLE 69 (**$\mathbb{N}$, $\mathbb{Z}$ are not fields**). The set of natural numbers with the usual operations of addition and multiplication is not a field. One can check that all field axioms except (4) and (8) hold, i.e. opposites and inverses do not exist in $\mathbb{N}$.

The set of integers $\mathbb{Z}$ has opposites but it fails axiom (8). For example 2^{-1} doesn't exist in $\mathbb{Z}$.

DEFINITION 37 (**Subfield**). Let K be a field and $F \subseteq K$. We say that F is a *subfield* if the following hold.

(a) $0 \in F$.

(b) $1 \in F$.

(c) F is closed under addition. That is,

$$a, b \in F \implies a + b \in F.$$

(d) F is closed under opposites. That is,

$$a \in F \implies -a \in F.$$

(e) F is closed under multiplication. That is,

$$a, b \in F \implies a\, b \in F.$$

(f) F is closed under inverses. That is,

$$a \neq 0 \text{ and } a \in F \implies a^{-1} \in F.$$

THEOREM 4.1.3 (**Subfields are fields**). *If F is a subfield of a field K, then F is also a field with addition, multiplication, 0, 1, opposites, and inverses the same as K.*

PROOF. Exercise. □

EXAMPLE 70 (**The complexes**). $\mathbb{C}$, the set of complex numbers, is a field. There various ways to define $\mathbb{C}$. We use the one in our see Exercise E.6 of Homework 4. For a more comprehensive treatment of complex numbers and their properties, consult Appendix A.

Let's then define the set of complex numbers to be the following set of matrices

$$\mathbb{C} = \left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} : a, b \in \mathbb{R} \right\}.$$

Addition and multiplication of complex numbers is the usual addition and multiplication of matrices. Notice that for all $a \in \mathbb{R}$ the scalar matrix $a\, I_2 \in \mathbb{C}$ and if we identify the real numbers with 2×2 scalar matrices we have

$$\mathbb{R} \subseteq \mathbb{C}.$$

Identifying $a \in \mathbb{R}$ with $a\, I_2 \in \mathbb{C}$ is justified since as we saw in Example 48, addition and multiplication of scalar matrices mimics addition and multiplication of real numbers. In particular we have identified I_2, the 2×2 identity matrix, with 1.

Let us now set

$$i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathbb{C}.$$

We have

$$i^2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathbb{C}. = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -1\, I_2 = -1.$$

Now,

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + b \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = a + b\, i,$$

since we have identified the scalar matrix $a\, I_2$ with a. Thus every element of $\mathbb{C}$ can be written in the form $a + b$ß for $a, b \in \mathbb{R}$. Thus we have

$$\mathbb{C} = \{a + b\, i : a, b \in \mathbb{R}\}.$$

Now notice that

$$a + b\, i = c + d\, i \iff a = c \text{ and } b = d.$$

So we have that a complex number $z \in \mathbb{C}$ and be written *uniquely* as

$$z = a + b\, i, \quad a, b \in \mathbb{R}.$$

We call a the *real part*, and b the *imaginary part*, of z.

CLAIM 2. *$\mathbb{C}$ endowed with matrix addition and multiplication is a field, where the unit 1 is the identity matrix, and the zero element 0 is the zero matrix*[1].

PROOF. Since addition in $\mathbb{C}$ is just matrix addition we have that addition in $\mathbb{C}$ is commutative and associative. Similarly we have that multiplication is associative and distributes over addition. Since 0 is the zero matrix we also have that for all $z \in \mathbb{C}$

$$z + 0 = 0$$

and thus axiom (3) holds.

Now by the properties of matrix addition we have

$$a\, I_2 + b\, i + (-a\, I_2 - b\, i) = (a - a)\, I_2 + (b - b)\, i = O,$$

that is,

$$a + b\, i + (-a - b\, i) = 0.$$

[1] Notice that these matrices are identified with the real numbers 1 and 0 via our identification of real numbers and scalar matrices.

So, axiom (4) also holds where

$$-(a+bi) = -a-bi.$$

Now,

$$\begin{aligned}(a+bi)(c+di) &= ac+adi+bci+bdi^2\\ &= ac-bd+(ad+bc)i.\end{aligned}$$

And so,

$$\begin{aligned}(c+di)(a+bi) &= ca-db+(da+cb)i\\ &= ac-bd+(ad+bc)i.\end{aligned}$$

Therefore we proved that multiplication in $\mathbb{C}$ is commutative.

The only axiom left to prove is axiom (8) the existence of inverses for non-zero complex numbers. Let $z = a+bi \in \mathbb{C}$ with $z \neq 0$. Then multiplying with its conjugate $\overline{z} = a-bi$ we get

$$z\,\overline{z} = a^2+b^2,$$

and since $z \neq 0$ we have $a^2+b^0 \neq 0$. Thus the last equation can be written

$$z\,\frac{\overline{z}}{a^2+b^2} = 1.$$

Thus axiom (8) also holds, and

$$z^{-1} = \frac{\overline{z}}{a^2+b^2}.$$

□

Example 71 ($\mathbb{Z}/2$**: The smallest possible field)**. The definition requires that a field has at least two elements the zero element 0 and the unity 1. It turns out that there is a field, the smallest possible, that has only those two elements. Let

$$\mathbb{Z}/2 = \{0,1\}$$

and define addition a multiplication as follows:

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

The multiplication table is determined by the axioms, if $\mathbb{Z}/2$ is to be a field then $0\cdot 0 = 0$, etc. The addition table is also defined by the axioms. The only non-obvious entry is probably

$$1+1=0.$$

This follows because 1 needs an opposite and this opposite can't be 0 because $1+0=1$. Thus we must have $-1=1$, or in other words, $1+1=0$.

The verification of the axioms is straight forward and involves checking finitely many identities.

We can see for example that every element has an opposite, namely itself, and that the only non-zero element 1 has an inverse, again itself.

To verify that addition is commutative for example, we need to verify that for all $a,b \in \mathbb{Z}/2$ we have

$$a+b=b+a.$$

Since this is obviously true when $a=b$ we need to verify it only for $a=0, b=1$ and $a=1, b=0$, and by symmetry we only need to check that

$$0+1=1+0,$$

which is true since both sides of this equation equal to 1.

To verify that multiplication distributes over addition we need to verify

$$a\,(b+c)=a\,b+a\,c$$

for all eight choices of a, b, c. By commutativity of addition these reduce to six, we chose a and the pair $\{b,c\}$. So we have to verify that

$$\begin{array}{ll} 0\,(0+0)=0\cdot 0+0\cdot 0 0\,(0+1)=0\cdot 0+0\cdot 1 & 0\,(1+1)=0\cdot 1+0\cdot 1 \\ 1\,(0+0)=0\cdot 0+0\cdot 0 1\,(0+1)=0\cdot 0+0\cdot 1 & 1\,(1+1)=0\cdot 1+0\cdot 1. \end{array}$$

The verification of all these is straightforward. Similarly we can check that addition and multiplication is associative.

Doing algebra on $\mathbb{Z}/2$ is easy because $x+x=0$ and $x^2=x$ for all $x\in\mathbb{Z}/2$. So for example the *freshman's dream identity* holds:

$$(x+y)^2=x^2+y^2.$$

There are (at least) two interesting interpretations of the field $\mathbb{Z}/2$. The first one is via logic. If we consider 0 to stand for *False* and 1 to stand for *True*, then multiplication is the logical **and** operation while addition is the logical **exclusive or**. Indeed, $a\,b$ is 1 only when both a and b are 1, just as p **and** q is true only when both p and q are true.

Similarly $a+b=1$ when exactly one of the a, b is 1, just as the exclusive disjunction of two propositions is true when exactly one of them is true.

The other interpretation of $\mathbb{Z}/2$ is as *integers modulo* 2. In this case 0 stands for *even* and 1 stands for *odd*. Then $1+1=0$ means that if we add two odd integers the result is even, $0\cdot 1=0$ means that if we multiply an even and an odd integer the result is even, and so on.

In general $\mathbb{Z}/m$ is defined for all integers $m\geq 2$. We can define for example

$$\mathbb{Z}/m=\{0,1,\dots,m-1\}$$

and think of its elements as the possible remainders of the division by m. It turns out that if a and c leave the same remainder when divided by m, and so do b and d, then $a+b$ leaves the same remainder as $c+d$, and $a\,b$ leaves the same remainder as $c\,d$. Then for $x, y\in\mathbb{Z}_m$ we define $x+y$ to be the remainder of the sum of x and y as integers, Similarly $x\,y$ is the remainder of the product of x and y as integers.

For example in $\mathbb{Z}/8$ we have $7+5=4$ because when we add 7 and 5 as integers we get 12, and 12 leaves remainder 4 when divided by 12. On the other hand $5\cdot 7=3$ because the product of the integers 5 and 7 is 35, and it leaves remainder 3.

Addition and multiplication of integers satisfy all the field axioms except the existence of inverses. It follows that so do modular addition and multiplication. For example the modular products $x\,(y\,z)$ and $(x\,y)\,z$ are equal because the products $x\,(y\,z)$ and $(x\,y)\,z$ are equal in $\mathbb{Z}$, and thus leave the same reminder.

Sometimes of course we get a filed, for example $\mathbb{Z}/2$ is a field as we saw. It turns out that for prime modulus axiom (8) is satisfied as well.

THEOREM 4.1.4. *For any $m\geq 2$ modular addition and multiplication satisfy all the field axioms with the possible exception of axiom (8), in other words inverses don't always exist. In fact, $\mathbb{Z}/p$ is a field if and only if p is a prime.*

PROOF. See Theorem B.2.4 in Appendix B. □

EXAMPLE 72 (**$\mathbb{Z}/3$ is a field)**. We have the following tables for addition and multiplication in $\mathbb{Z}/3$:

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

.

From the multiplication table we see that $1^{-1} = 1$ and $2^{-1} = 2$ and thus every non-zero element has an inverse. Thus axiom (8) is satisfied. Since the remaining axioms are satisfied for all $\mathbb{Z}/m$ we conclude that indeed $\mathbb{Z}/3$ is a field.

EXAMPLE 73 (**$\mathbb{Z}/4$ is not a field)**. Indeed in $\mathbb{Z}/4$ we have $2 \cdot 2 = 0$ and $2 \neq 0$. This contradicts Property (7) in Theorem 4.1.1.

EXAMPLE 74 (**$\mathbb{Z}/5$ is a field)**. For $\mathbb{Z}/5$ we have

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\cdot$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

.

We can think of the elements of $\mathbb{Z}/5$ arranged in a circle at the vertices of a regular pentagon, see Figure 1.

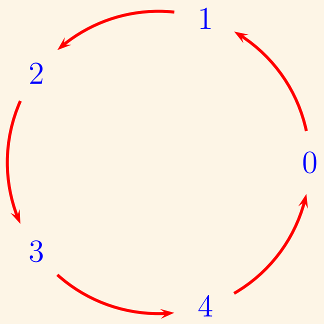


FIGURE 1. Modular arithmetic for $m = 5$.

Addition and multiplication are then defined as follows:

To find $a + b \in \mathbb{Z}/5$, start at a and walk along the circle for b steps.

Thus to find $2 + 4$ I start at 2 and walk 4 steps, thus going through $3, 4, 0$ and ending in 1. Thus $2 + 4 = 1$.

Multiplication can be defined as repeated addition.

To find $a\,b \in \mathbb{Z}/5$, start at 0 and and walk b steps a times.

Thus to find $4{\cdot}3$, we start at 0 and walk 3 steps, for 4 times, going through $3, 1, 4$ and ending at 2 thus $4 \cdot 3 = 2$.

As an example of how our familiar algebra works in any field let us prove the following claim.

CLAIM 3. *For all $x \in \mathbb{Z}/5$ we have:*

$$x^5 = x.$$

PROOF. Let $p(x) = x^5 - x$. We will prove that for all $a \in \mathbb{Z}/5$ we have $p(a) = 0$. We factor:

$$\begin{aligned} x^5 - x &= x\,(x^4 - 1) \\ &= x\,(x^2 - 1)\,(x^2 + 1) \\ &= x\,(x - 1)\,(x + 1)\,(x^2 - 4) \\ &= x\,(x - 1)\,(x + 1)\,(x - 2)\,(x + 2) \\ &= x\,(x - 1)\,(x - 4)\,(x - 2)\,(x - 3) \\ &= (x - 0)\,(x - 1)\,(x - 2)\,(x - 3)\,(x - 4). \end{aligned}$$

In the above calculations we used that $1 = -4$ and $2 = -3$ in $\mathbb{Z}/5$.

Thus all elements of $\mathbb{Z}/5$ are roots of $p(x)$. □

REMARK 17. There are similar interpretations for the arithmetic operations in $\mathbb{Z}/m$ for all m. We also remark that an analogous result to Claim 3 holds in all fields $\mathbb{Z}/p$. Indeed we have

$$\forall x \in \mathbb{Z}/p, \quad x^p = x.$$

EXAMPLE 75. The set

$$\mathbb{Q}[\sqrt{2}] = \left\{a + b\sqrt{2} : a, b \in \mathbb{Q}\right\}$$

is a subfield of $\mathbb{R}$. This means that it contains 0 and 1, and it is closed under addition, multiplication, opposites, and inverses.

(a) $0 \in \mathbb{Q}[\sqrt{2}]$. Indeed $0 = 0 + 0\sqrt{2}$.
(b) $1 \in \mathbb{Q}[\sqrt{2}]$. Indeed $0 = 1 + 0\sqrt{2}$.
(c) $\mathbb{Q}[\sqrt{2}]$ is closed under addition. Indeed,

$$\left(a + b\sqrt{2}\right) + \left(c + d\sqrt{2}\right) = (a + c) + (b + d)\sqrt{2}.$$

(d) $\mathbb{Q}[\sqrt{2}]$ is closed under multiplication. Indeed,

$$\left(a + b\sqrt{2}\right)\left(c + d\sqrt{2}\right) = (a\,c + 2\,b\,d) + (a\,d + b\,c)\sqrt{2}.$$

(e) $\mathbb{Q}[\sqrt{2}]$ is closed under opposites. Indeed,

$$-(a + b\sqrt{2}) = -a + (-b)\sqrt{2}.$$

(f) $\mathbb{Q}[\sqrt{2}]$ is closed under inverses. This is a bit more challenging. We have to show that for $a, b \in \mathbb{Q}$ with $a + b\sqrt{2} \neq 0$ we have

$$\frac{1}{a + b\sqrt{2}} \in \mathbb{Q}[\sqrt{2}].$$

We have

$$\left(a+b\sqrt{2}\right)\left(a-b\sqrt{2}\right)=a^2-2b^2,$$

and so if $a^2-2b^2\neq 0$ we have

$$\frac{1}{a+b\sqrt{2}}=\frac{a}{a^2-2b^2}-\frac{b}{a^2-2b^2}\sqrt{2}\in Q[\sqrt{2}].$$

Thus we have to show that if $a,b\in\mathbf{Q}$ not both zero, then $a^2-2b^2\neq 0$. To see this notice that if $b=0$ then $a\neq 0$ and so $a^2-2b^2=a^2\neq 0$.

On the other hand, if $b\neq 0$ then

$$a^2-2b^2=0\iff\sqrt{2}=\frac{a}{b}\implies\sqrt{2}\in\mathbb{Q}.$$

Since $\sqrt{2}$ is irrational we conclude that $a^2-2b^2\neq 0$.

4.1.1. Standard vector spaces over arbitrary fields. All the material we developed in the previous three chapters can be extended over any field K. We can solve linear systems, using Gauss, or Gauss-Jordan elimination. Any matrix is row equivalent to one in echelon form and has a unique reduced echelon form. The solutions of an $m\times n$ systems are n-dimensional vectors i.e. elements of K^n, the standard n-dimensional vector space over k. In K^n we have vector addition, and scalar multiplication by scalars $\lambda\in K$. We have vector subspaces of K^n, linear combinations, bases, dimension and so on.

This is so because all the operations we used make sense in K as well and have the same algebraic properties. We give a few examples of applying the theory we developed to fields different than $\mathbb{R}$.

EXAMPLE 76 (**Solving a system in $\mathbb{Z}/3$**). Consider the following 3×3 system of linear equations over the field with three elements $\mathbb{Z}_3$

$$\begin{cases} x+2y-z=2\\ 2x+\ y\qquad=1\\ \qquad y+z=0\end{cases}.$$

The augmented matrix of the system is.

$$\left(\begin{array}{ccc|c}1&2&-1&2\\2&1&0&1\\0&1&1&0\end{array}\right).$$

First notice that in $\mathbb{Z}/3$ we $-2=1$. We add the first row to the second, then interchange the second and third rows.

$$\left(\begin{array}{ccc|c}1&2&-1&2\\0&0&-1&0\\0&1&1&0\end{array}\right)\sim\left(\begin{array}{ccc|c}1&2&-1&2\\0&1&1&0\\0&0&-1&0\end{array}\right).$$

Now we add the third row to the second, subtract it from the first and multiply it by -1. Finally we add the second row to the first.

$$\left(\begin{array}{ccc|c}1&2&0&2\\0&1&0&0\\0&0&1&0\end{array}\right)\sim\left(\begin{array}{ccc|c}1&0&0&2\\0&1&&0\\0&0&1&0\end{array}\right).$$

Thus the system has a unique solution $(x,y,z)=(2,0,0)$.

EXAMPLE 77. Find a basis for the subspace of $\mathbb{C}^5$ spanned by the vectors

$$\mathbf{v}_1 = (1, i, 1+i, 1+3i, -2i) \qquad \mathbf{v}_2 = (1, i, 1+i, 1+3i, -2i)$$
$$\mathbf{v}_3 = (1, i, 1+i, 1+3i, 1+i, -1-2i) \qquad \mathbf{v}_4 = (0, 0, 0, 1, -1)$$
$$\mathbf{v}_5 = (0, 0, 0, 0, 1)$$

Let A be the matrix with columns $\mathbf{v}_i$, $i = 1, \ldots, 5$

$$A = \begin{pmatrix} 1 & i & 0 & 1 & 0 \\ i & -1 & 0 & i & 0 \\ 1+i & -1+i & 0 & 1+i & 0 \\ 1+3i & -3+i & 0 & 1+3i & 1 \\ -2i & 2 & 1 & -1-2i & -1 \end{pmatrix}.$$

We'll bring A to its reduced row echelon form. Let's start by using the second row to get rid of of the imaginary parts of below it. So we subtract the second row from the third, add -3 times the second row to the fourth, and 2 times the second row to the fifth.

$$A \sim \begin{pmatrix} 1 & i & 0 & 1 & 0 \\ i & -1 & 0 & i & 0 \\ 1 & i & 0 & 1 & 0 \\ 1 & i & 0 & 1 & 1 \\ 0 & 0 & 1 & -1-4i & -1 \end{pmatrix}.$$

Now we add $-i$ times the first row to the second, and subtract the first row from the third and fourth:

$$A \sim \begin{pmatrix} 1 & i & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 & -1 \end{pmatrix}.$$

Next we move the zero row to the bottom and the last row to the second place

$$A \sim \begin{pmatrix} 1 & i & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Finally, we add the third row to the second

$$A \sim \begin{pmatrix} 1 & i & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus a basis for the span of $\mathbf{v}_i$, $i = 1, \ldots, 5$ is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and the span has dimension 3.

EXAMPLE 78. Find A^{-1} where A is the following matrix with entries in $\mathbb{Z}/5$.

$$A = \begin{pmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \\ 0 & 2 & 3 \end{pmatrix}$$

In $\mathbb{Z}/5$ we have $2+3=1+4=0$ and $2\cdot 3=4\cdot 4=1$.

We have

$$\left(\begin{array}{ccc|ccc}1&3&2&1&0&0\\4&0&1&0&1&0\\0&2&3&0&0&1\end{array}\right)\sim\left(\begin{array}{ccc|ccc}1&3&2&1&0&0\\0&3&3&1&1&0\\0&2&3&0&0&1\end{array}\right)\sim\left(\begin{array}{ccc|ccc}1&3&2&1&0&0\\0&3&3&1&1&0\\0&0&1&1&1&1\end{array}\right),$$

where we first added the first row the second and then the second to the third.

Now we add 2 (respectively 3) times the third row to the second (respectively first), and then subtract the second row from the first. Finally we divide the second row by 3 (i.e. we multiply with 2)

$$\sim\left(\begin{array}{ccc|ccc}1&3&0&4&3&3\\0&3&0&3&3&2\\0&0&1&1&1&1\end{array}\right)\sim\left(\begin{array}{ccc|ccc}1&0&0&1&0&1\\0&3&0&3&3&2\\0&0&1&1&1&1\end{array}\right)\sim\left(\begin{array}{ccc|ccc}1&0&0&1&0&1\\0&1&0&1&1&4\\0&0&1&1&1&1\end{array}\right).$$

Thus,

$$A^{-1}=\begin{pmatrix}1&0&1\\1&1&4\\1&1&1\end{pmatrix}.$$

EXAMPLE 79. Show that the map

$$T\colon \mathbb{C}^2\to\mathbb{C}^3,\quad T(z_1,z_2)=(2\,z_1-i\,z_2,(1-3\,i)\,z_1,(3+i)\,z_1-2\,i\,z_2).$$

is linear and find its matrix.

We will first find the matrix of T assuming it is linear and then we will show that T is given by multiplying column vectors from the left with that matrix thus establishing the linearity of T. We calculate

$$T(1,0)=(2,1-3\,i,3+i),\quad T(0,1)=(-i,0,-2\,i)$$

and so if T is linear its matrix will be

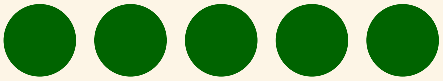
$$\begin{pmatrix}2&-i\\1-3\,i&0\\3+i&-2\,i\end{pmatrix}.$$

Now, we have

$$\begin{pmatrix}2&-i\\1-3\,i&0\\3+i&-2\,i\end{pmatrix}\begin{pmatrix}z_1\\z_2\end{pmatrix}=\begin{pmatrix}2\,z_1-i\,z_2\\(1-3\,i)\,z_1\\(3+i)\,z_1-2\,i\,z_2\end{pmatrix}=T(z_1,z_2).$$

Thus T is linear.

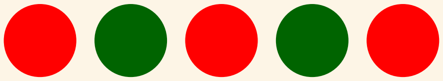
EXAMPLE 80 (**A puzzle**). We have five coins each with one side green and and the other red. We place them in a row with the green sides up.



We are allowed to flip any coin and its immediate neighbors. Thus we are allowed to

(a) Flip the first two coins, or
(b) flip the first three coins, or
(c) flip the middle three coins, or
(d) flip the last three coins, or
(e) flip the last two coins.

To solve the puzzle, we have, using only these allowed operations, to get the first, third, and fifth coin with the red side up, and the second and fourth with the green side up.



This is a typical example of a system described by finitely many *bits*. Each coin can be flipped or not. If we let 0 stand for “not flipped” and 1 for “flipped” the state of our system can be represented by a tuple of five bits, that is by a vector in $(\mathbb{Z}/2)^5$. Thus flipping the first and the fourth coin is represented by the vector $(1,0,0,1,0)$, and flipping only the third coin by $(0,0,1,0,0)$. Adding the vectors that correspond to two states corresponds to performing the corresponding flipping operations consecutively.

Now the allowed operations correspond to the vectors

$$\mathbf{v}_1 = (1,1,0,0,0),\ \mathbf{v}_2 = (1,1,1,0,0),\ \mathbf{v}_3 = (0,1,1,1,0),\ \mathbf{v}_4 = (0,0,1,1,1),\ \mathbf{v}_5 = (0,0,0,1,1),$$

and the final state we want to achieve corresponds to the vector

$$\mathbf{v} = (1,0,1,0,1),$$

Thus to solve the puzzle we have to express v as a linear combination of $\mathbf{v}_i$, $i = 1,\dots,5$. So we have to solve the system

$$\mathbf{v} = \sum_{i=1}^{5} x_i\,\mathbf{v}_i$$

where x_i is either 0 or 1.

Taking the augmented matrix we have

$$\left(\begin{array}{ccccc|c} 1&1&0&0&0&1\\ 1&1&1&0&0&0\\ 0&1&1&1&0&1\\ 0&0&1&1&1&0\\ 0&0&0&1&1&1 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1&1&0&0&0&1\\ 0&0&1&0&0&1\\ 0&1&1&1&0&1\\ 0&0&1&1&1&0\\ 0&0&0&1&1&1 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1&1&0&0&0&1\\ 0&1&1&0&0&1\\ 0&0&1&0&0&1\\ 0&0&1&1&1&0\\ 0&0&0&1&1&1 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1&1&0&0&0&1\\ 0&1&1&0&0&1\\ 0&0&1&0&0&1\\ 0&0&1&1&1&0\\ 0&0&0&0&0&1 \end{array}\right).$$

The last row implies that the system has no solutions and therefore the puzzle is impossible.

EXAMPLE 81 (**An other puzzle)**. Assume that we have the same puzzle as in Example 81 but now when the allowed move is to flip any coin and its coin to the left (if present).

This puzzle leads to the augmented matrix

$$\left(\begin{array}{ccccc|c} 1&0&0&0&0&1\\ 1&1&0&0&0&0\\ 0&1&1&0&0&1\\ 0&0&1&1&0&0\\ 0&0&0&1&1&1 \end{array}\right).$$

We proceed to get the row-echelon form:

$$\sim \left(\begin{array}{ccccc|c} 1&0&0&0&0&1\\ 0&1&0&0&0&1\\ 0&1&1&0&0&1\\ 0&0&1&1&0&0\\ 0&0&0&1&1&1 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1&0&0&0&0&1\\ 0&1&0&0&0&1\\ 0&0&1&0&0&0\\ 0&0&1&1&0&0\\ 0&0&0&1&1&1 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1&1&0&0&0&1\\ 0&1&0&0&0&1\\ 0&0&1&0&0&1\\ 0&0&0&1&0&0\\ 0&0&0&1&1&1 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1&1&0&0&0&1\\ 0&1&0&0&0&1\\ 0&0&1&0&0&0\\ 0&0&0&1&0&0\\ 0&0&0&0&1&1 \end{array}\right).$$

Thus

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_5,$$

and thus the puzzle can be solved by flipping the first two coins, then the second and the third, and then the fifth. We show the solution in Figure 2

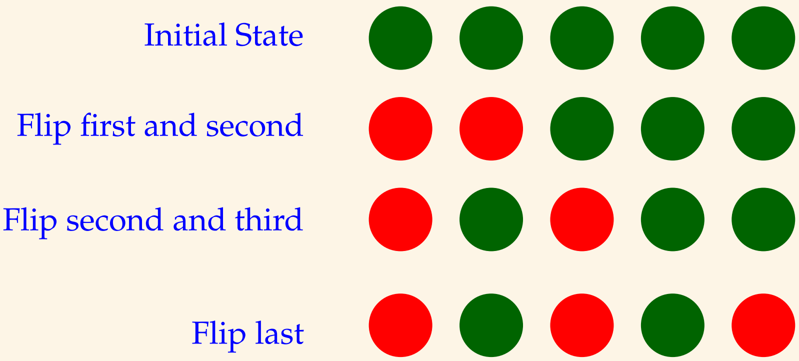


FIGURE 2. The solution of the puzzle in Example 81.

We remark that we could perform the three operations in any order since vector addition is commutative.

4.2. Vector Spaces

DEFINITION 38 (**Vector space**). Let K be a field. A set V is said to be a *vector space over* K if there are is a binary operation, called *(vector) addition*

$$V \times V \to V, \quad (\mathbf{v}, \mathbf{w}) \mapsto \mathbf{v} + \mathbf{w}$$

and a binary operation, called *scalar multiplication*

$$K \times V \to V, \quad (\lambda, \mathbf{w}) \mapsto \lambda \mathbf{w}$$

that satisfy the following properties (called *vector space axioms*):

(a) Addition is commutative. That is for all $\mathbf{v}, \mathbf{w} \in V$ we have
$$\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}.$$

(b) Addition is associative. That is for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ we have
$$\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}.$$

(c) Addition has a neutral element $\mathbf{0}$. That is for all $\mathbf{v} \in V$
$$\mathbf{v} + \mathbf{0} = \mathbf{v}.$$

(d) Every $\mathbf{v} \in V$ has an *opposite* $-\mathbf{v}$. That is, there exists $-v \in V$ such that
$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0}.$$

(e) Scalar multiplication distributes over vector addition. That is for $\lambda \in K$ and $\mathbf{v}, \mathbf{w} \in V$ we have
$$\lambda (\mathbf{v} + \mathbf{w}) = \lambda \mathbf{v} + \lambda \mathbf{w}.$$

(f) Scalar multiplication distributes over field addition. That is for $\lambda, \mu \in K$ and $\mathbf{v} \in V$ we have
$$(\lambda + \mu)\, \mathbf{v} = \lambda \mathbf{v} + \mu \mathbf{v}.$$

(g) For all $\lambda, \mu \in K$ and $\mathbf{v} \in V$,
$$\lambda (\mu \mathbf{v}) = (\lambda \mu) \mathbf{v}.$$

(h) For all $\mathbf{v} \in V$ we have

$$1\,\mathbf{v} = \mathbf{v}.$$

As we remarked in Section 2.1 all the usual algebraic properties of scalar multiplication and vector addition follow from these axioms.

THEOREM 4.2.1 (Some consequences of the axioms). *We have:*

- *The zero vector is unique. That is there is only one element $\mathbf{0} \in V$ that satisfies Axiom (3).*
- *For $\mathbf{v} \in V$ the opposite $-\mathbf{v}$ is unique. That is there is only one element $-\mathbf{v} \in V$ that satisfies Axiom (4).*
- *For all vectors $\mathbf{a}$, $\mathbf{b}$ the equation*

  $$\mathbf{a} + \mathbf{x} = \mathbf{b}$$

  has a unique solution.
- *For any vector $\mathbf{a}$*

  $$-1\,\mathbf{a} = -\mathbf{a}$$
- *For any scalar λ we have*

  $$\lambda\,\mathbf{0} = \mathbf{0}.$$
- *For any vector $\mathbf{a}$*

  $$0\,\mathbf{a} = \mathbf{0}.$$
- *For scalar λ and vector $\mathbf{a}$*

  $$\lambda\,\mathbf{a} = \mathbf{0} \iff \lambda = 0 \text{ or } \mathbf{a} = \mathbf{0}.$$

PROOF. Exercise. For the first two the proof mimics the proof of the analogous properties of a field, see the proofs in Section 4.1. The proofs of the other properties are exactly the same as the proofs for the standard vector spaces, see Section 2.1. □

EXAMPLE 82. K^n is a vector space over K. This follows from Theorem 2.1.1.

EXAMPLE 83. If V is a vector subspace of K^n (see Definition 6) then V with the scalar multiplication and vector addition inherited from K^n is a vector space. First of all we note that these operations are well defined by definition. Also we have $\mathbf{0} \in V$ and if $\mathbf{v} \in V$ its opposite $-\mathbf{v} \in V$. Since the vector space axioms hold for K they also hold for V.

More generally we have the following definition of vector subspace.

DEFINITION 39. Let V be a vector space over a field K and $W \subseteq V$. We say that W is a *vector subspace* of V if the following hold.

(a) W contains the zero vector of V, that is $\mathbf{0} \in W$.
(b) W is closed under vector addition, that is

$$\mathbf{x}, \mathbf{y} \in W \implies \mathbf{x} + \mathbf{y} \in W.$$

(c) W is closed under scalar multiplication, that is

$$\lambda \in K, \mathbf{x} \in W \implies \lambda\,\mathbf{x} \in W.$$

And of course we have the following theorem.

THEOREM 4.2.2 (**Alternative definition of vector subspace**). *Let V be a vector space over a field K. A subset $W \subseteq V$ is a subspace if and only if the following two properties hold:*

- *$W \neq \varnothing$.*
- *For all $\lambda, \mu \in K$ and $\mathbf{a}, \mathbf{b} \in V$*

$$\mathbf{a}, \mathbf{b} \in W \implies \lambda\,\mathbf{a} + \mu\,\mathbf{b} \in W.$$

PROOF. Entirely analogous to the proof of Theorem 2.1.3. □

We can now generalize Example 83.

THEOREM 4.2.3. *If V is a vector space and W is a vector subspace of V then W with the operations inherited from V is also a vector space.*

PROOF. Exercise. Follow the proof that a vector subspace of K^n is a vector space given in Example 83. □

EXAMPLE 84. Let m, n be two positive integers. Then the set $\mathrm{M}_{m\times n}(K)$ of $m \times n$ matrices with entries in K is a vector space over K with the zero matrix playing the role of the zero vector $\mathbf{0}$. See the discussion at the beginning of Section 3.4.

EXAMPLE 85 (**Function Spaces)**. Let X be any set, and let V be a vector space over a field k. Denote by $V^X$the set of functions $X \to V$, that is

$$V^X = \{f \mid f\colon X \to V\}.$$

For $f, g \in V^X$ and $\lambda \in K$ we define the functions $f + g$ and λf as follows:

$$(f+g)(x) = f(x) + g(x), \quad (\lambda f)(x) = \lambda f(x).$$

Then V^X endowed with these operations is a vector space over K. The role of $\mathbf{0}$ is played by the zero function $O\colon X \to V$ defined via

$$O(x) = \mathbf{0},$$

and the opposite $-f$ is the function

$$(-f)(x) = -f(x).$$

To prove that V^X is indeed a vector space we have to prove that all the Vector Space Axioms listed in Definition 38 hold. To prove that two functions , say f, g, are equal we need to prove that $f(x) = g(x)$ for all $x \in X$. We have,

(a) Let $x \in X$ then

$$\begin{aligned} (f+g)(x) &= f(x) + g(x) && \text{(By Definition)}\\ &= g(x) + f(x) && \text{(Addition in } V \text{ is commutative)}\\ &= (g+f)(x) && \text{(By definition) .} \end{aligned}$$

Therefore $f + g = g + f$.

(b) Let $x \in X$ then

$$\begin{aligned} (f+(g+h))(x) &= f(x) + (g+h)(x) && \text{(By Definition)}\\ &= f(x) + (g(x) + h(x)) && \text{(By Definition)}\\ &= (f(x) + g(x)) + h(x) && \text{(Addition in } V \text{ is associative)}\\ &= ((f+g)+h)(x) && \text{(By Definition) .} \end{aligned}$$

Thus $f + (g + h) = (f + g) + h$.

(c) For $x \in X$ we have

$$\begin{aligned} (f+O)(x) &= f(x) + O(x)\\ &= f(x) + \mathbf{0}\\ &= f(x) \end{aligned}$$

Thus $f + O = f$.

(d) We have for $x \in X$

$$\begin{aligned}(f + (-f))\,(x) &= f(x) + (-f)(x)\\ &= f(x) + (-f(x))\\ &= \mathbf{0}\\ &= O(x).\end{aligned}$$

Thus $f + (-f) = O$.

(e) Let $x \in X$, then

$$\begin{aligned}(\lambda\,(f+g))\,(x) &= \lambda\,((f+g)(x))\\ &= \lambda\,(f(x)+g(x))\\ &= \lambda\,f(x) + \lambda\,g(x)\\ &= (\lambda\,f)\,(x) + (\lambda\,g)\,(x)\\ &= (\lambda\,f + \lambda\,g)\,(x).\end{aligned}$$

Thus, $\lambda\,(f+g) = \lambda\,f + \lambda\,g$.

(f) For $x \in X$ we have

$$\begin{aligned}((\lambda+\mu)\,f)\,(x) &= (\lambda+\mu)\,f(x)\\ &= \lambda\,f(x) + \mu\,f(x)\\ &= (\lambda\,f)\,(x) + (\mu\,f)\,(x)\\ &= (\lambda\,f + \mu\,f)\,(x).\end{aligned}$$

Thus, $(\lambda+\mu)\,f = \lambda\,f + \mu\,f$.

(g) For $x \in X$ we have

$$\begin{aligned}((\lambda\,\mu)\,f)\,(x) &= (\lambda\,\mu)\,f(x)\\ &= \lambda\,(\mu\,f(x))\\ &= \lambda\,((\mu\,f)(x))\\ &= (\lambda\,(\mu\,f))\,(x)\end{aligned}$$

Thus, $(\lambda\,\mu)\,f = \lambda\,(\mu\,f)$.

(h) For $x \in X$, we have

$$\begin{aligned}(1\,f)\,(x) &= 1\,f(x)\\ &= f(x).\end{aligned}$$

Thus $1\,f = f$.

From Theorem 4.2.3 and Example 85 we have the following examples of function spaces that are vector spaces.

EXAMPLE 86 (**The vector space of continuous functions**). The set $\mathcal{C}(\mathbb{R})$ of *continuous* functions $\mathbb{R} \to \mathbb{R}$ is a vector space. Indeed, as we know from Calculus, the sum of two continuous functions is continuous as is the product of a real number and a continuous function.

EXAMPLE 87 (**The vector space of differentiable functions**). The set $\mathcal{C}^1(\mathbb{R})$ of *differentiable* functions $\mathbb{R} \to \mathbb{R}$ is a vector space. Indeed, as we know from Calculus, the sum of two differentiable functions is differentiable as is the product of a real number and a differentiable function.

EXAMPLE 88 (**The set of functions that vanish on a given point**). Let $a \in \mathbb{R}$ be an arbitrary (but fixed) real number. Then

$$V = \{f \in \mathcal{C}(\mathbb{R}) : f(a) = 0\}$$

is a vector subspace of $\mathcal{C}(\mathbb{R})$. Indeed, the zero function 0 vanishes at a and so $0 \in V$. Furthermore, if $\lambda, \mu \in \mathbb{R}$ and $f, g \in V$, we have

$$(\lambda f + \mu g)(0) = \lambda f(0) + \mu g(0) = 0$$

and therefore $\lambda f + \mu g \in V$.

EXAMPLE 89 (**The vector space of everywhere convergent powerseries**). A function $f\colon \mathbb{R} \to \mathbb{R}$ that is defined via a powerseries

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

that converges for all $x \in \mathbb{R}$ is said to be *analytic*. The set $\mathcal{C}^{\omega}(\mathbb{R})$ of *analytic* functions $\mathbb{R} \to \mathbb{R}$ is a vector space.

Indeed, as we know from Calculus, the sum of two convergent series is convergent as is the product of a real number and a convergent series.

EXAMPLE 90. $\mathbb{R}$ is a vector space over the field of rational numbers $\mathbb{Q}$. Vector addition is the usual addition of real numbers, and scalar multiplication is the usual multiplication of real numbers. This makes sense because $\mathbb{Q} \subseteq \mathbb{R}$ so for $\lambda \in \mathbb{Q}$ and $x \in \mathbb{R}$ we have $\lambda x \in \mathbb{R}$. The zero vector is the real number $0 \in \mathbb{R}$, and the opposite of a is the usual opposite $-a$.

The first four of the Vector Space Axioms hold because vector addition is just field addition. Axioms (5) and (6) follow from the fact that in a field, and thus $\mathbb{R}$, multiplication distributes over addition. Axiom (7) holds because multiplication is associative, and (8) is true because 1 is neutral for multiplication.

In general we have the following theorem.

THEOREM 4.2.4. *If F is a subfield of K then K is a vector space over F.*

PROOF. Exercise. □

EXAMPLE 91 (**The vector space of polynomials**). See Appendix C for the basic definitions and properties of polynomials.

Let K be a field. A *polynomial* of one variable x, over K, is an expression of the form

$$p(x) = \sum_{k=0}^{n} a_k x^k = a_0 + a_1 x + \cdots + a_n x^n,$$

where $a_k \in K$, and x an indeterminate.

The set of all polynomials of one variable x with coefficients in K is denoted by $K[x]$. It is sometimes convenient to write polynomials as a sum of infinite many terms, with only finitely many of them non-zero. In other words we think of a polynomial as having infinitely many coefficients, one for each power x^n, but after a certain power of x all coefficients are 0. Thus we write

$$p(x) = \sum_{k \in \mathbb{N}} a_k x^k = a_0 + a_1 x + \cdots + a_n x^n + \cdots,$$

and we assume that $a_n = 0$ for all but finitely many $n \in \mathbb{N}$.

We define addition and scalar multiplication via

$$\left(\sum_{n\in\mathbb{N}} a_n x^n\right) + \left(\sum_{n\in\mathbb{N}} b_n x^n\right) = \sum_{n\in\mathbb{N}} (a_n + b_n) x^n, \quad \lambda \left(\sum_{n\in\mathbb{N}} a_n x^n\right) = \sum_{n\in\mathbb{N}} (\lambda a_n) x^n.$$

Note that these formulas really define polynomials because only finitely many of the sums $a_n + b_n$, or the products $\lambda\, a_n$, are non-zero.

With these definitions, $K[x]$ is a vector space over K, with the role of the zero vector played by the zero polynomial, that is the polynomial with all coefficients 0. The opposite of a polynomial $p(x)$ is the polynomial with coefficients the opposites of the coefficients of $p(x)$, i.e.

$$p(x) = \sum_{n\in\mathbb{N}} a_n\, x^n \implies -p(x) = \sum_{n\in\mathbb{N}} (-a_n)\, x^n.$$

The proof that $K[x]$ is indeed a vector space is straightforward. The verification of all the axioms follows from the field axioms. For example the proof that Axiom (8) holds goes as follows:

$$\begin{aligned} 1\left(\sum_{n\in\mathbb{N}} a_n\, x^n\right) &= \sum_{n\in\mathbb{N}} (1\, a_n)\, x^n \\ &= \sum_{n\in\mathbb{N}} a_n\, x^n. \end{aligned}$$

Exercise 9. Prove that $K[x]$ is indeed a vector space over K.

4.2.1. The vector space of linear forms. Let X be a set of variables and K a field. Then we define a *linear form* to be a *formal sum*

$$\omega = \sum_{x\in X} \lambda_x\, x$$

where $\lambda_x \in K$, and only finitely many of λ_x are non-zero. The scalars λ_x are called the *coefficients* of ω. The set of all linear forms with set of indeterminates X is denoted by $K\langle X\rangle$. Two linear forms on the same set of indeterminates are considered equal if they have the same coefficients, that is

$$\sum_{x\in X} \lambda_x\, x = \sum_{x\in X} \mu_x\, x \iff \forall x\in X,\ \lambda_x = \mu_x.$$

If X is a finite set then we often write the sums in *expanded form*. For example if $X = \{x_1, x_2, x_3, x_4\}$ then instead of

$$\omega = \sum_{x\in X} \lambda_x\, x$$

we write

$$\omega = \lambda_{x_1}\, x_1 + \lambda_{x_2}\, x_2 + \lambda_{x_3}\, x_3 + \lambda_{x_4}\, x_4.$$

When we use this expanded form we omit terms with 0 coefficient, thus writing

$$\omega = -3\, x_1 + 4\, x_3$$

instead of

$$\omega = -3\, x_1 + 0\, x_2 + 4\, x_3 + 0\, x_4.$$

If all the coefficients are 0 we simply write 0. In other words, the zero-form

$$\sum_{x\in X} 0\, x$$

is written simply 0. If $x \in X$ write x instead of $1\, x$ and $-x$ instead of $-1\, x$.

Thus instead of

$$1\, x_1 + 2\, x_2 + 0\, x_3 + (-1)\, x_4$$

we write simply

$$x_1 + 2\, x_2 - x_4.$$

If X is finite the requirement that only finitely many of the coefficients are non-zero is always satisfied. The condition is non-trivial only when we have infinitely many variables. For example if $X = \{x_1, \ldots, x_n, \ldots\}$ the expression

$$x_1 + x_2 + \cdots + x_n + \cdots = \sum_{i \in \mathbb{N}} 1\, x_i \notin K\langle X\rangle$$

since infinitely many coefficients are non-zero. Informally speaking, we allow the sum of infinitely many zero-terms but we only add finitely many non-zero terms.

We add two linear forms, with the same set of variables X, by adding their coefficients, and we multiply a form with a scalar by multiplying all the coefficients with that scalar.

DEFINITION 40. Let $\omega_1 = \sum_{x \in X} \lambda_x\, x$ and $\omega_2 = \sum_{x \in X} \mu_x\, x$ be two linear forms and $\lambda \in K$. Then we define

$$\omega_1 + \omega_2 = \sum_{x \in X} (\lambda_x + \mu_x)\, x$$

and

$$\lambda\, \omega_1 = \sum_{x \in X} (\lambda\, \lambda_x)\, x.$$

EXAMPLE 92. Consider linear forms over $\mathbb{R}$ with variables x, y, z. We have

$$(x + 3\,y - 5\,z) + (-2, x + y + 8\,z) = -x + 4\,y + 3\,z,$$

while

$$7\,(2\,x + y - 3\,z) = 14\,x + 7\,y - 21\,z.$$

REMARK 18. We have used subtraction as a an abbreviation. $2\,x + y - 3\,z$ stands for $2\,x + y + (-3)\,z$.

THEOREM 4.2.5. *Let X be a set of indeterminates and K a field. Then, $K\langle X\rangle$ endowed with addition and scalar multiplication is a vector space over K. The zero vector is the linear form 0, that is*

$$\sum_{x \in X} 0\, x.$$

The opposite of a form has the opposite coefficients, that is

$$-\sum_{x \in X} \lambda_x\, x = \sum_{x \in X} (-\lambda_x)\, x.$$

PROOF. Exercise. □

We consider $X \subset K\langle X\rangle$ by identifying the variable $y \in X$ with the form

$$\sum_{x \in X} \delta_{xy} x$$

that is the form where all variables have coefficient 0 except y that has coefficient 1. With this convention we see that every element of $K\langle X\rangle$ can be expressed as a *linear combination* of the variables in X in a *unique* way. In other words X is a *basis* of $K\langle X\rangle$.

4.3. Linear dependence, basis, dimension

The concepts studied in Section 2.2 (such as linear span, linear (in)dependence, basis, dimension) are defined in exactly the same way in all vector spaces and the results (and their proofs) proved there carry over almost verbatim.

There is an important caveat however, not all vector spaces are finite-dimensional, that is not all vector spaces have bases with finitely many elements. Some of the results of Chapter 2 do not hold for infinite-dimensional vector spaces, and some results that do hold, have different proofs.

That said, we will mostly concentrate on finite-dimensional vector spaces, infinite-dimensional ones will occasionally occur but mostly in examples, they are not studied per se.

NOTE (**Notation**). We will be using normal fonts for elements of an arbitrary vector space V. Thus we will use v, u, w for elements of V, and 0 for the zero vector. When there is a chance of confusion we may use special notation for elements of V, such as $\overrightarrow{0}$ for the zero vector. We reserve bold font for the elements of the standard vector space K^n.

DEFINITION 41. Let V be a K-vector space and $S \subseteq V$. A *linear combination* of elements of S is a sum

$$\sum_{v \in S} \lambda_v \, v$$

where only finitely many coefficients λ_v are non-zero.

By convention a linear combination of elements of the empty subset of V is a sum with zero terms and is equal to the zero vector of V.

The *linear span* of S, denoted $K\langle S \rangle$ or when K is understood simply $\langle S \rangle$, is the set of all linear combinations of S.

A subset S of V is called *spanning* if $\langle S \rangle = V$, that is if every element of V can be written as a linear combination of S.

A subset S of V is said to be *linearly independent* if every element of $\langle S \rangle$ can be written as a linear combination of S in a unique way. That is, if the following condition holds

$$\sum_{v \in S} \lambda_v \, v = \sum_{v \in S} \mu_v \, v \iff \forall v \in S, \ \lambda_v = \mu_v .$$

If S is not linearly independent we say that S is *linearly dependent*.

The *trivial linear combination* (or *zero linear combination*) of S is the linear combination with all coefficients zero.

A *linear dependency in S* is a non-trivial linear combination of S equal to the zero-vector.

If S is both spanning and linearly independent then S is said to be a *basis*. Thus S is a basis if every element of V can be written as a linear combination of S in a *unique way*.

REMARK 19. If S is an infinite set, then according to Definition 41 a linear combination of S is a sum with infinitely many terms. This doesn't cause a real problem because only finitely many terms are non-zero, and we make the convention that a sum of infinitely many zeros equals to zero.

Essentially, linear combinations are finite sums, potentially padded with (potentially infinitely many) zeros.

This convention makes all linear combinations of S to have the same numbers of terms, and this is convenient in many circumstances.

This approach introduces some annoying insolvencies as well. For, if $S' \subsetneq S$ then a linear combination of S' is not a linear combination of S: there are no terms corresponding to the elements of $S \setminus S$, i.e. the elements of S that are not in S'. For example, if $S' = \{v, u\}$ and $S = \{v, u, w\}$ then a linear combination of S' has two terms $\lambda v + \mu u$ while a linear combination of S has three terms $\lambda v + \mu u + \nu w$, and strictly speaking these are different expressions even if $\nu = 0$.

Of course, we can easily fix this, we extend any linear combination of S' to a linear combination of S by adding zero terms. So we identify the linear combinations

$$\sum_{v \in S'} \lambda_v \, v = \sum_{v \in S} \mu_v \, v$$

where

$$\mu_v = \begin{cases} \lambda_v & v \in S' \\ 0 & v \notin S' \end{cases}. \tag{4.1}$$

Also, unless we want to emphasize them, we will omit the zero terms in a linear combination. Thus instead of $2\,v + 0\,u - 3\,w$ we write $2\,v - 3\,w$. If needed we may consider $2\,v - 3\,w$ as a linear combination of $\{x, v, w\}$ as well. No harm is caused by this because the only linear combinations that extends to a trivial linear combination is the trivial one. Indeed using the notation of Equation (4.1), we have that if $\mu_v = 0$ for all $v \in S$, then $\lambda_v = 0$ for all $v \in S'$.

In what follows many of the proofs are only sketched because the are (nearly) identical to the proofs in Chapter 2.

THEOREM 4.3.1. *Let $S \subseteq V$ where V is a vector space over a field K. The following hold:*

(a) S is linearly dependent if and only if there is a linear dependency in S.
(b) S is linearly independent if and only if the only linear combination of S equal to 0 is the trivial linear combination.
(c) If $S' \subseteq S$ and S' is spanning then so is S.
(d) If $S' \subseteq S$ and S' is linearly dependent then so is S.
(e) If $S' \supseteq S$ and S' is linearly independent then so is S.
(f) If 0 in S then S is linearly dependent.
(g) If some $w \in S$ is a linear combination of $S \setminus \{w\}$ then S is linearly dependent.

PROOF. (a) If there is a linear dependency

$$\sum_{v \in S} \lambda_v \, v = 0$$

with some $\lambda_v \neq 0$, then 0 can be written as a linear combination of S in two different ways:

$$\sum_{v \in S} \lambda_v \, v = 0 = \sum_{v \in S} 0 \, v.$$

Thus S is not linearly independent.

Conversely, if S is linearly dependent then some $w \in V$ can be expressed as a linear combination of S in two different ways, say

$$w = \sum_{v \in S} \lambda_v \, v, \quad w = \sum_{v \in S} \mu_v \, v$$

with $\lambda_u \neq \mu_u$ for some $u \in S$. Then by subtracting the two equations we get

$$0 = \sum_{v \in S} \lambda_v \, v - \sum_{v \in S} \mu_v \, v = \sum_{v \in S} (\lambda_v - \mu_v) \, v.$$

Since $\lambda_u \neq \mu_u$ we have $\lambda_u - \mu_u \neq 0$ and thus

$$\sum_{v \in S} (\lambda_v - \mu_v) \, v = 0$$

is a non-trivial linear combination equal to zero, i.e. a linear dependency.

(b) Logically equivalent to Item (a).

(c) Every element of V is a linear combination of S'. Since, (see Remark 19) linear combinations of S' are also linear combinations of S it follows that every element of V is a linear combination of S. Thus S is spanning.
(d) A linear dependency on S' is also a linear dependency on S.
(e) Logically equivalent to Item (d).
(f) $\left\{\overrightarrow{0}\right\}$ is linearly dependent since $1 \cdot \overrightarrow{0} = \overrightarrow{0}$. and by Item (d) so is every superset of $\left\{\overrightarrow{0}\right\}$.
(g) If $w = \lambda_1 v_1 + \cdots + \lambda_n v_n$ with $\lambda_i \in K$ and $v_i \in S \setminus \{w\}$ we have

$$-1\, w + \lambda_1 v_1 + \cdots + \lambda_n v_n = 0,$$

a linear dependency in S since $-1 \neq 0$.

□

EXAMPLE 93 (**Examples of linearly dependent and independent sets**). Here are some examples of linearly dependent, and linearly independent sets for several vector spaces.

(a) Consider $\mathbb{R}$ as a vector space over $\mathbb{Q}$. Then the set

$$\left\{\sqrt{2}, \sqrt{3}\right\}$$

is linearly independent.

Indeed, assume to the contrary that there are two rational numbers $p, q \in \mathbb{Q}$ such that

$$p\sqrt{2} + q\sqrt{3} = 0. \tag{4.2}$$

Then squaring both sides we get

$$2p^2 + 2pq\sqrt{6} + 3q^2 = 0.$$

Now if $p = 0$ we get $3q^2 = 0$ and so $q = 0$ as well. Similarly, if $q = 0$ we have that $p = 0$ as well.

Assume then that $p \neq 0$ and $q \neq 0$. Then we get

$$\sqrt{6} = -\frac{2p^2 + 3q^2}{2pq}.$$

Since $p, q \in \mathbb{Q}$ the RHS of the last equation is a rational number and thus we get that $\sqrt{6} \in \mathbb{Q}$. But this is a contradiction because $\sqrt{6}$ is irrational (see Corollary 5 in Appendix B).

(b) The set $\left\{\sqrt{20}, \sqrt{45}\right\} \subseteq \mathbb{R}$ is linearly dependent over $\mathbb{Q}$.

Indeed $\sqrt{20} = 2\sqrt{5}$ and $\sqrt{45} = 3\sqrt{5}$. Therefore,

$$3\sqrt{20} - 2\sqrt{45} = 0.$$

(c) The set $\{\mathbf{v}, \mathbf{u}, \mathbf{w}\} \subseteq \mathbb{C}^3$ where

$$\mathbf{v} = (1+2i, 1-i, i), \quad \mathbf{u} = (2, -3i, 3-4i), \quad \mathbf{w} = (11, 1-12i, 11-11i),$$

is linearly dependent.

Considering the matrix A with columns $\mathbf{v}, \mathbf{u}, \mathbf{w}$ we have

$$A = \begin{pmatrix} \mathbf{v} & \mathbf{u} & \mathbf{w} \end{pmatrix} = \begin{pmatrix} 1+2i & 2 & 11 \\ 1-i & -3i & 1-12i \\ i & 3-4i & 11-11i \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -2i+1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}.$$

And thus $\{\mathbf{v}, \mathbf{u}, \mathbf{w}\}$ is linearly independent. We actually have

$$\mathbf{w} = (1 - 2\,i)\,\mathbf{v} + 3\,\mathbf{u}$$

because A is the augmented matrix of the vector equation

$$z_1\,\mathbf{v} + z_2\,\mathbf{u} = \mathbf{w}$$

considered as a system.

(d) Let S be the subset of $\mathbf{M}_{2\times 3}$ consisting of the matrices

$$A = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 2 & -2 \\ 1 & 0 & 3 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 3 & -4 \\ 1 & 0 & 0 \end{pmatrix}$$

is linearly independent.

Indeed the equation

$$x\,A + y\,B + z\ = O$$

is equivalent to

$$x \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix} + y \begin{pmatrix} 0 & 2 & -2 \\ 1 & 0 & 3 \end{pmatrix} + z \begin{pmatrix} 1 & 3 & -4 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

which in turn is equivalent to the system

$$\begin{cases} 2x \phantom{{}+2y} + 3z = 0 \\ -x + 2y + 3z = 0 \\ - 2y - 4z = 0 \\ y + z = 0 \\ x = 0 \\ x + 3y = 0 \end{cases}.$$

Solving this is straightforward, the fifth equation gives $x = 0$, and then the sixth gives $y = 0$, and then the fourth gives $z = 0$. Thus only the trivial linear combination of S equals to zero, and S is linear independent.

(e) Consider the subset S of $\mathbb{R}[x]$ consisting of the polynomials

$$p_0(x) = 1, \;\; p_1(x) = x, \;\; p_2(x) = x\,(x-1), \;\; p_3(x) = x\,(x-1)\,(x-2).$$

To check whether S is linearly independent we have to solve the equation

$$\sum_{i=0}^{3} \lambda_i\, p_i(x) = 0, \tag{4.3}$$

for $\lambda_i \in \mathbb{R}$.

We can proceed in two ways:

First way: Find the coefficients We expand

$$p_2(x) = x^2 - x$$

and then

$$p_3(x) = (x^2 - x)\,(x-2) = x^3 - 3\,x^2 + 2\,x.$$

Thus the LHS of Equation (4.3) is

$$\lambda_0 + \lambda_1\,x + \lambda_2\,(x^2 - x) + \lambda_3\,(x^3 - 3\,x^2 + 2\,x) = \lambda_0 + (\lambda_1 - \lambda_2 + 2\,\lambda_3)\,x + (\lambda_2 - 3\,\lambda_3)\,x^2 + \lambda_3\,x^3.$$

In order for this to be the zero polynomial all coefficients have to be 0. So we get the system

$$\begin{cases} \lambda_0 = 0 \\ \lambda_1 - \lambda_2 + 2\lambda_3 = 0 \\ \lambda_2 - 3\lambda_3 = 0 \\ \lambda_3 = 0 \end{cases}.$$

Clearly the system has only the trivial solution and therefore S is linearly independent.

Second way: Evaluate at select points We write Equation (4.3) as $p(x) = 0$ where,

$$p(x) = \lambda_0 + \lambda_1 x + \lambda_2 x(x-1) + \lambda_3 x(x-1)(x-2).$$

Since $p(x)$ is the zero polynomial, $p(a) = 0$ for all real numbers a.

Evaluating at 0 we have $p(0) = \lambda_0$ and thus $\lambda_0 = 0$. Thus

$$p(x) = \lambda_1 x + \lambda_2 x(x-1) + \lambda_3 x(x-1)(x-2).$$

Evaluating at 1 we have $p(1) = \lambda_1$ and so $\lambda_1 = 0$. Therefore

$$p(x) = \lambda_2 x(x-1) + \lambda_3 x(x-1)(x-2).$$

Evaluating at 2 we get $p(2) = 2\lambda_2$ and so $\lambda_2 = 0$. But then

$$p(x) = \lambda_3 x(x-1)(x-2).$$

and evaluating at any number other than $0, 1, 2$ we get that $\lambda_3 = 0$. Thus we conclude again that Equation (4.3) has only the trivial solution.

(f) The set of functions

$$S = \{f_1, f_2, f_3\} \subseteq \mathbb{R}^{\mathbb{R}}$$

where,

$$f_k \colon \mathbb{R} \to \mathbb{R}, \quad f_k(x) = \cos kx, \quad k = 1, 2, 3$$

is linearly independent.

Indeed consider a linear dependency

$$a\cos x + b\cos 2x + c\cos 3x = 0.$$

Evaluating at $x = \pi/2$ gives $-b = 0$ and so $b = 0$. Evaluating then at $x = \pi/6$ gives

$$a\cos\frac{\pi}{6} = 0 \implies a = 0$$

and finally evaluating at $x = 0$ gives $c = 0$ as well.

Therefore only the trivial linear combination of S equals the zero function establishing that S is linearly independent.

THEOREM 4.3.2 (**Characterizations of basis**). *Let V be a vector space and $B \subseteq V$. Then the following are equivalent.*

(a) B is a basis.

(b) B is spanning and linearly independent.

(c) Every $v \in V$ can be written uniquely as a linear combination of elements of B.

(d) B is a maximal *linearly independent subset of V. That is, B is linearly independent, and if $B \subsetneq B'$ then B' is not linearly independent.*

(e) B is a minimal *spanning subset of V. That is B is spanning and if $B' \subsetneq B$ then B' is not spanning.*

PROOF. We have that (a), (b), and (c) are equivalent by definition.
(a) $\implies$ (d). If B is a basis and $w \notin B$ then w is a linear combination of elements of B and therefore, by Item (g) of Theorem 4.3.1, $B \cup \{w\}$ is linearly dependent.
(d) $\implies$ (a). If B is a maximal linearly independent subset of V, then B is spanning. For, if there was a $w \in V \smallsetminus \langle B \rangle$ we would have that $B \cup \{w\}$ is linearly independent contradicting the maximality of B.
(a) $\implies$ (e). Let B be a basis of V and B' a proper subset of B. Then there is a $w \in B \smallsetminus B'$ and such a w is not a linear combination of elements of B', otherwise, by Item (g) of Theorem 4.3.1, B would be linearly dependent. Since $w \notin \langle B' \rangle$ we have that B' is not spanning.
(d) $\implies$ (a). If B is a minimal spanning subset of V then B is linearly independent. For, if there is a linear dependency in B

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = 0$$

with, say, $\lambda_1 \neq 0$ then

$$v_1 = \frac{\lambda_2}{\lambda_1} v_2 + \cdots + \frac{\lambda_n}{\lambda_1} v_n$$

and this means that $B \smallsetminus \{v_1\}$ is spanning. Indeed, in any linear combination of B, we can replace v_1 with the RHS of the above equation to get a linear combination that does not involve v_1. But $B \smallsetminus \{v_1\}$ is a proper subset of B and thus it is not spanning, and we've arrived at a contradiction. Thus B is linearly independent and thus a basis. □

It turns out that every vector space has a basis. The idea of the proof is essentially the same as in the case of subspaces of $\mathbb{R}^n$ (see Theorem 2.2.6) but there are some logical subtleties arising from the fact that a basis is not necessarily a finite set. So we just state the following theorem and provide a rough sketch of the proof in Appendix D.

THEOREM 4.3.3 (**Existence of basis**). *Every vector space has a basis. Furthermore, if B_1 and B_2 are bases of the same vector space V then there is a bijection $B_1 \to B_2$*[2].

If V has a finite basis all the proofs of Section 2.2 go through. We collect the main results in the following theorem.

THEOREM 4.3.4. *et V be a vector space over a field K and let $B = \{v_1, \ldots, v_d\}$ be a basis of V. Then the following hold.*

(a) Every other basis of V has d elements.
(b) If $B' \subseteq V$ is linearly independent and $|B'| = d$ then B' is a basis.
(c) If $B' \subseteq V$ is spanning and $|B'| = d$ then B' is a basis.
(d) If $B' \subseteq V$ and $|B'| > d$ then B' is linearly dependent.
(e) If $B' \subseteq V$ and $|B'| < d$ then B' is not spanning.

PROOF. Exercise. Go through Section 2.2, find the corresponding statements and verify that the proofs go through. □

DEFINITION 42 (**Dimension**). The cardinality of a basis of V is called the *dimension* of V and is denoted by $\dim V$. If $\dim V$ is finite we say that V is *finite dimensional*. If V is not finite dimensional we say that V is *infinite dimensional*.

Occasionally we write $\dim V < \infty$ (respectively $\dim V = \infty$) to mean that V is finite dimensional (respectively infinite dimensional).

REMARK 20 (**Basis and dimension of the zero vector space**). We make the convention that the empty set $\varnothing$ is a basis of the zero vector space $\{0\}$ and therefore $\dim\{0\} = 0$.

[2] This means, by definition, that B_1 and B_2 have the same *cardinality*, i.e. the same number of elements.

THEOREM 4.3.5. *Let V be a finite dimensional vector space over a field K and $W \subseteq V$ a vector subspace. Then*

$$\dim W \leq \dim V,$$

with equality holding if and only if $W = V$.

Furthermore any basis of W can be extended to a basis of V.

EXAMPLE 94 (**The standard basis of $\mathbf{M}_{m\times n}$**). The set of basic matrices $B = \{E_{ij} : i = 1, \dots, m,\ j = 1, \dots, n\}$ where E_{ij} has all entries 0, except the (i, j)-th entry that is 1 (see Definition 20) form a basis of $\mathbf{M}_{m\times n}(K)$.

The proof is very similar to the proof of the real case. See Proposition 2.

It follows that $\dim \mathbf{M}_{m\times n} = m\,n$.

As an application we can prove the following proposition.

PROPOSITION 11. *Let $A \in \mathbf{M}_n(K)$. Then there exists a non-zero polynomial $p(x) \in K[x]$ such that $p(A) = O$.*

PROOF. Since $\dim \mathbf{M}_n = n^2$ the set $\{A^k : k = 0, \dots, n^2\}$ is linearly dependent. Therefore there is a non-trivial linear combination

$$c_0\, I + c_1\, A + c_2\, A^2 + \cdots + c_{n^2}\, A^{n^2} = O$$

with at least one coefficient $c_i \neq 0$. That means that A is a root of the non-zero polynomial

$$p(x) = c_0 + c_1\, x + \cdots + c_n\, x^n.$$

□

EXAMPLE 95 (**Basis for upper triangular matrices**). Recall (see Example 50 in Chapter 3) that Δ_n stand for the set of $n \times n$ triangular matrices, and that Δ_n is closed under addition and multiplication, in particular, Δ_n is a vector subspace of $\mathbf{M}_n$.

The set of basic matrices

$$B = \{E_{ij} : 1 \leq j \leq n\}$$

is a basis of Δ_n. For example, for $n = 3$ we have the basis

$$\{E_{11}, E_{12}, E_{13}, E_{22}, E_{23}, E_{33}\}\,.$$

It follows that

$$\dim \Delta_n = n + (n-1) + \cdots + 2 + 1 = \frac{n\,(n+1)}{2}.$$

EXAMPLE 96 (**The standard basis of $K[x]$**). The set

$$B = \{x^n : n \in \mathbb{N}\} = \left\{1, x, x^2, \dots, x^n, \dots\right\}$$

is a basis of $K[x]$. Indeed B is spanning because every polynomial is, by definition, a linear combination of elements of B. To see that B is linearly independent observe that if

$$\sum_{n\in\mathbf{N}} a_n\, x^n = 0$$

then the polynomial $\sum a_n\, x^n$ is the zero polynomial and therefore $a_n = 0$ for all $n \in \mathbb{N}$.

We call B the *standard basis of* $K[x]$. Since B is infinite we conclude that $K[x]$ is infinite dimensional.

EXAMPLE 97 (**Polynomials of degree up to n**). Let $\mathbf{P}_n(K)$ be the set of polynomials with coefficients in K and degree at most n. That is

$$\mathbf{P}_n(K) = \left\{\sum_{k=0}^{n} a_k\, x^k : a_i \in K,\ \text{for } i = 0, \dots, n\right\}.$$

Clearly,

$$\mathbf{P}_n = K\langle 1, x, \dots, x^n\rangle$$

and therefore $\mathbf{P}_n(K)$ is a vector subspace of $K[x]$. By Example 96 we have that

$$B = \{x^n : n = 0, \dots, n\}$$

is linearly independent.

Therefore $\dim \mathbf{P}_n(K) = n + 1$.

EXAMPLE 98 (**An other basis of $\mathbb{R}[x]$**). The set B consisting of polynomials

$$\begin{aligned} p_0(x) &= 1 \\ p_1(x) &= x \\ p_2(x) &= x\,(x-1) \\ p_3(x) &= x\,(x-1)\,(x-2) \\ &\dots\dots \\ p_n(x) &= x\,(x-1)\,(x-2)\cdots(x-n+1) \\ &\dots\dots \end{aligned}$$

is a basis of $\mathbb{R}[x]$.

Formally, we define $p_n(x)$ recursively as follows:

$$p_0(x) = 1, \quad p_{n+1}(x) = p_n(x)\,(x-n).$$

We can see[3] that B is linearly independent as in Item (e) of Example 93.

To prove that B is spanning we will prove that every element of the standard basis of $\mathbb{R}[x]$ (see Example 96) is a linear combination of elements of B.

Thus, we will prove using mathematical induction that for all $n \in \mathbb{N}$, $x^n \in \langle B\rangle$.

For $n = 0$ we have $x^0 = 1 \in B$. Now assume that for some $n_1, n_2, \dots, n_k \in \mathbb{N}$ we have

$$x^n = \lambda_1\, p_{n_1} + \lambda_2\, p_{n_2} + \cdots + \lambda_k\, p_{n_k}.$$

Then, since

$$x\, p_n(x) = p_{n+1}(x) + n\, p_n(x),$$

we have

$$\begin{aligned} x^{n+1} &= x^n\, x \\ &= (\lambda_1\, p_{n_1} + \lambda_2\, p_{n_2} + \cdots + \lambda_k\, p_{n_k})\; x \\ &= \lambda_1\; (p_{n_1+1}(x) + n_1\, p_{n_1}(x)) + \cdots + \lambda_k\; (p_{n_k+1}(x) + n_k\, p_{n_k}(x)) \\ &= \lambda_1\, p_{n_1+1}(x) + n_1\, \lambda_1\, p_n(x) + \cdots + \lambda_k\, p_{n_k+1}(x) + n_k\, \lambda_k\, p_{n_k}(x). \end{aligned}$$

So x^{n+1} is also a linear combination of elements of B. Thus we established that B is spanning.

4.4. Linear maps

The definition of linear map (see Definition 14) carries over almost verbatim.

DEFINITION 43 (**Linear function**). Let V and W be vector spaces over a field K. A function

$$f\colon V \to W$$

is said to be *linear* if it enjoys the following two properties.

[3] Do this yourself.

(a) *It respects vector addition*. This means that for any two vectors $v, w \in V$ we have
$$f(v+w) = f(v) + f(w).$$
(b) *It respects scalar multiplication*. This means that for all $\lambda \in K$ and $v \in V$ we have
$$f(\lambda v) = \lambda f(v).$$

And of course Theorems 3.1.1 and 3.1.2 also hold, as does Corollary 2.

EXAMPLE 99. The function
$$f\colon K[x] \to K[x], \quad f(p(x)) = 3p(x) - 2$$
is not linear. Indeed, $f(0) = -2 \neq 0$.

EXAMPLE 100 (**The zero function and the identity function are linear)**. Let V, W be vector space over a field K. Then we can define the zero map that sends all vectors to the zero vector of W:
$$O\colon V \to W, \quad O(v) = 0.$$
Clearly O is linear[4]

Also clearly[5], the identity function $I\colon V \to V$ that sends every vector to itself (i.e. $I(v) = v$) is linear.

EXAMPLE 101 (**The derivative is a linear operator)**. Let $\mathcal{C}(\mathbb{R})$ be the vector space of continuous functions $\mathbb{R} \to \mathbb{R}$, and let $\mathcal{C}^1(\mathbb{R})$ be the vector space of continuously differentiable functions $\mathbb{R} \to \mathbb{R}$, that is functions that have continuous derivatives. Then the function
$$D\colon \mathcal{C}^1(\mathbb{R}) \to \mathcal{C}(\mathbb{R}), \quad D(f) = f'$$
is linear.

Indeed, let f, g be continuous differentiable functions and $a, b \in \mathbb{R}$. Then from Calculus we know that
$$D(af + bg) = (af + bg)' = af' + bg' = a\,D(f) + b\,D(g).$$

EXAMPLE 102 (**The definite integral is a linear operator)**. Let $\mathcal{C}([\nvdash,\nVdash])$ be the vector space of continuous functions $[0,1] \to \mathbb{R}$. Then the function
$$S\colon \mathcal{C}([0,1]) \to \mathbb{R}, \quad S(f) = \int_0^1 f(x)\,dx$$
is linear.

Indeed, let f, g be continuous real functions on the unit interval $a, b \in \mathbb{R}$. Then from Calculus we know that
$$S(af + bg) = \int_0^1 (af + bg)(x)\,dx = \int_0^1 af(x)\,dx + b\int_0^1 g(x)\,dx = a\,S(f) + b\,S(g).$$

To simplify notation let's write $\mathbb{F}_p$ to stand for the field $\mathbb{Z}/p$.

EXAMPLE 103. Let
$$L\colon \mathbb{F}_2[x] \to \mathbb{F}_2[x], \quad L(p(x)) = p(x)^2.$$
Then L is a linear map.

We first establish the following result.

CLAIM 4. *Let $p(x) \in \mathbb{F}_2[x]$. Then*
$$p(x) + p(x) = 0.$$

[4] Is it clear? Prove it.

[5] Ditto.

PROOF. Let $p(x) = \sum a_n\, x^n$, with $a_n \in \mathbb{F}_2$. Then

$$p(x) + p(x) = \sum a_n\, x^n + \sum a_n\, x^n = \sum (a_n + a_n)\, x^n = \sum 0\, x^n = 0$$

because for $a \in \mathbb{F}_2$ we have $a + a = 0$. □

Let then $p(x), q(x) \in \mathbb{F}_2[x]$. We have

$$L\left(p(x) + q(x)\right) = \left(p(x) + q(x)\right)^2 = p(x)^2 + p(x)\, q(x) + p(x)\, q(x) + q(x)^2 = p(x)^2 + q(x)^2 = L(p(x)) + L(q(x)).$$

Thus Poperty (a) of Definition 43 is satisfied. Poperty (b) is also satisfied since

$$L(0\, p(x)) = (0\, p(x))^2 = 0^2 = 0,$$

and

$$L(1\, p(x)) = (1\, p(x))^2 = p(x)^2 = 1\, p(x)^2 = 1\, L(p(x)).$$

EXAMPLE 104 (**Evaluation is linear**). Let K be a field and let $a \in K$. Let $E_a \colon K[x] \to K$ be the map that assigns to a polynomial its value at a. That is,

$$E_a(p(x)) = p(a).$$

Then E_a is linear.

I leave the proof as an exercise. Start by writing $p(x) = \sum a_n\, x^n$, $q(x) = \sum b_n\, x^n$, and compute $\lambda\, p(x) + \mu\, q(x)$ and substitute a for x.

EXAMPLE 105 (**Taking transpose is linear**). The function

$$T \colon \mathbf{M}_{m\times n} \to \mathbf{M}_{n\times m}, \quad T(A) = A^*$$

is linear.

We already proved this, see Theorem 3.5.1.

EXAMPLE 106 (**Extracting the diagonal is linear**). The function that extracts the diagonal of a square matrix is linear. That is, (refer to Example 49 for the notation), the map

$$D \colon \mathbf{M}_n \to \mathbf{M}_n \quad T(A) = \operatorname{diag}(a_{11}, a_{22}, \dots, a_{nn}).$$

is linear.

Indeed

$$\begin{aligned} D(\lambda\, A + \mu\, B) &= \operatorname{diag}(\lambda\, a_{11} + \mu, b_{11}, \dots, \lambda\, a_{nn} + \mu, b_{nn}) \\ &= \operatorname{diag}(\lambda\, a_{11}, \dots, \lambda\, a_{nn}) + \operatorname{diag}(\mu, b_{11}, \dots, \mu, b_{nn}) \\ &= \lambda\ \operatorname{diag}(a_{11}, \dots, a_{nn}) + \mu\ \operatorname{diag}(b_{11}, \dots, b_{nn}) \\ &= \lambda\, D(A) + \mu\, D(B). \end{aligned}$$

EXAMPLE 107. Let $\mathbf{S}_n$ be the vector spaces of symmetric matrices. The the function

$$f \colon \mathbf{M}_n \to \mathbf{S}_n, \quad f\,(A) = A + A^*$$

where A^* is the transpose of A, is linear.

We first note that f is *well defined*, that is $f(A)$ is indeed a symmetric matrix since

$$\left(A + A^*\right)^* = A^* + \left(A^*\right)^* = A^* + A = A + A^*.$$

The linearity of f follows from the properties of the transpose. Indeed, if $\lambda, \mu \in \mathbb{R}$ and $A, B \in \mathbf{M}_n$ we have

$$f\,(\lambda\, A + \mu\, B) = \left(\lambda\, A + \mu\, B\right)^* = \lambda\, A^* + \mu\, B^* = \lambda\, f\,(A) + \mu\, f\,(B)\,.$$

EXAMPLE 108. Let $X \in \mathbf{M}_m$ be an $m \times m$ matrix. Then for any positive integer n, the function

$$f\colon \mathbf{M}_{m\times n} \to \mathbf{M}_{m\times n}, \quad f(A) = X\,A$$

is linear.

Indeed, Poperty (a) of Definition 43 holds because matrix multiplication distributes over matrix addition (Property (b) of Theorem 3.4.1):

$$f(A+B) = X\,A + X\,B = f(A) + f(B)$$

Poperty (b) of Definition 43 holds because of Property (c) of Theorem 3.4.1:

$$f(A) = X\,(\lambda\,A) + \lambda\,(X\,A) = \lambda, f(A).$$

The *kernel* and the *range* of a linear map are also defined the same way as in the case of linear maps between standard vector spaces. That is, if $f\colon V \to W$ is a linear map then

$$\ker f = \{v \in V : f(v) = 0\}$$

is a subspace of V, and

$$\mathcal{R}(f) = \{f(v) : v \in V\}$$

is a subspace of W.

THEOREM 4.4.1. *Let V, W be vector spaces over a field K and let let $f\colon V \to W$ be a linear map. Then f is injective if and only if*

$$\ker f = \{0\}\,.$$

PROOF. Since f is linear we have $f(0) = 0$, and thus if f is injective we have

$$v \in kerf \implies f(v) = 0 \implies f(v) = f(0) \implies v = 0.$$

Conversely, the linearity of f gives

$$f(v) = f(w) \implies f(v) - f(w) = 0 \implies f(v-w) = 0 \implies v - w \in \ker f.$$

Thus if $\ker f = \{0\}$, we have

$$v - w \in \ker f \implies v - w = 0 \implies v = w,$$

that is, f is injective. □

As in the case of the standard vector spaces, a linear map is completely determined by the images of a basis. We express this idea in the following two theorems, the first, Theorem 4.4.2 says that if two linear maps agree on a basis then they are equal. The second, Theorem 4.4.3 says that to define a linear function we only need to define it on a basis.

THEOREM 4.4.2. *Let $f, g\colon V \to W$ be two linear maps and let B be a basis of V. If for all$v \in B$ we have*

$$f(v) = g(v)$$

then $f = g$.

PROOF. We need to prove that for all $v \in V$ we have $f(v) = g(v)$. So, let $v \in V$. Since B is a basis of B we have

$$v = \sum_{u\in B} \lambda_u\,u,$$

for some (unique) $\lambda_u \in K$. But then using the linearity of f and g we have

$$f(v) = f\left(\sum_{u\in B} \lambda_u\,u\right) = \sum_{u\in B} \lambda_u\,f(u) = \sum_{u\in B} \lambda_u\,g(u) = g\left(\sum_{u\in B} \lambda_u\,u\right) = g(v).$$

□

THEOREM 4.4.3. *Let V, W be vector spaces over a field K, and let B be a basis of V. Then any function*

$$f\colon B \to W$$

can be uniquely extended to a linear function

$$L\colon V \to W.$$

That is there one, and only one, linear function

$$L\colon V \to W$$

such that for all $v \in B$ we have

$$L(v) = f(v).$$

Furthermore, the range of L is the linear span

$$\mathcal{R}(L) = K\left\langle f(u) : u \in B \right\rangle.$$

PROOF. By Theorem 4.4.2, there can be only one such linear function. Now for $v \in V$, there are unique coefficients λ_u such that

$$v = \sum_{u \in B} \lambda_u\, u$$

. We can then define,

$$L(v) = \sum_{u \in B} \lambda_u\, f(u).$$

We remark, that the uniqueness of the coefficients λ_u guarantees that this indeed defines a unique vector $L(v)$.

Let now, $\lambda, \mu \in K$ and $v, w \in V$. We need to prove that

$$L(\lambda\, v + \mu\, w) = \lambda\, L(v) + \mu\, L(w).$$

Let us express v and w as linear combinations of B:

$$v = \sum_{u \in B} \lambda_u\, u, \quad w = \sum_{u \in B} \mu_u\, u$$

and observe that the expression of $\lambda\, v + \mu\, w$ is

$$\lambda\, v + \mu\, w = \sum_{u \in B} (\lambda\, \lambda_u + \mu\, \mu_u)\, u.$$

It follows that

$$\begin{aligned} L(\lambda\, v + \mu\, w) &= \sum_{u \in B} (\lambda\, \lambda_u + \mu\, \mu_u)\, f(u) \\ &= \lambda \sum_{u \in B} \lambda_u\, f(u) + \mu \sum_{u \in B} \mu_u\, f(u) \\ &= \lambda\, L(v) + \mu\, L(w). \end{aligned}$$

Now, by definition, $L(v)$ is a linear combination of $\{f(u) : u \in B\}$, and any linear combination is the image of a vector $v \in V$. Indeed

$$\sum_{u \in B} \lambda_u\, f(u) = L\left(\sum_{u \in B} \lambda_u\, u \right).$$

Thus,

$$\mathcal{R}(L) = K\left\langle f(u) : u \in B \right\rangle.$$

□

In practice when we use Theorem 4.4.3 we use the same symbol for f and L, and say something "like let $L\colon V \to W$ be the linear function defined on B via" and give the values $L(u)$ for $u \in B$. Here are a few examples.

EXAMPLE 109. Find a formula for the linear function $L\colon \mathbf{M}_n(K) \to K$ defined on the standard basis of $\mathbf{M}_n(K)$ by

$$L(E_{ij}) = \delta_{ij}.$$

In other words if $i = j$ then $L(E_{ij}) = 1$ otherwise $L(E_{ij}) = 0$.

Let $X = (x_{ij})$ be an $n \times n$ matrix with entries in K. Then

$$X = \sum_{i,j=1}^{n} x_{ij}\, E_{ij}$$

and therefore

$$L(X) = \sum_{i,j=1}^{n} x_{ij}\, \delta_{ij} = \sum_{i=1}^{n} x_{ii}.$$

Thus $L(X)$ is the sum of the diagonal entries of X.

The linear function of Example 109 is an important one, it will appear later in this course.

DEFINITION 44 (**The trace of a square matrix**). The function defined in Example 109 is called the *trace*, and is denoted by trace. Thus for $X = (x_{ij}) \in \mathbf{M}_n(K)$ we have

$$\operatorname{trace} X = \sum_{i=1}^{n} x_{ii}.$$

EXAMPLE 110. Let $\mathbb{C}\langle z_1, z_2, z_3\rangle$ be the vector space of linear forms with indeterminates z_1, z_2, z_3 (see Section 4.2.1), and let

$$f\colon \mathbb{C}\langle z_1, z_2, z_3\rangle \to \mathbb{C}^3$$

be defined on the basis $\{z_1, z_2, z_3\}$ via

$$f(z_1) = (i, 0, 0), \quad f(z_2) = (1, 1 - i, 0), \quad f(z_3) = (i, -i, 3\,i).$$

Find a formula for f, and determine its kernel and its range.

We have for $a_1, a_2, a_3 \in \mathbb{C}$

$$\begin{aligned} f(a_1\, z_1 + a_2\, z_2 + a_3\, z_3) &= a_1\, f(z_1) + a_2\, f(z_2) + a_3\, f(z_3) \\ &= a_1\,(i, 0, 0) + a_2\,(1, 1 - i, 0) + a_3\,(i, -i, 3\,i) \\ &= (a_2 + (a_1 + a_3)\, i, a_2 - (a_2 + a_3)\, i, 3\, a_3\, i)\,. \end{aligned}$$

The kernel of f consists of all linear forms $a_1\, z_1 + a_2\, z_2 + a_3\, z_3$ such that

$$f(a_1\, z_1 + a_2\, z_2 + a_3\, z_3) = \mathbf{0}.$$

This means

$$(a_2 + (a_1 + a_3)\, i, a_2 - (a_2 + a_3)\, i, 3\, a_3\, i) = \mathbf{0},$$

which is equivalent to the system

$$\left\{ \begin{array}{ll} a_2 + (a_1 + a_3)\, i & = 0 \\ a_2 - (a_2 + a_3)\, i & = 0 \\ 3\, a_3\, i & \end{array} \right. . \tag{4.4}$$

It's easy to see that The only solution is the trivial one and it follows that

$$\ker f = \{0\}\,.$$

The range of f is the linear span $\mathbb{C}\langle f(z_1), f(z_2), f(z_3)\rangle$. But since the system (4.4) has only the trivial solution the set $\{f(z_1), f(z_2), f(z_3)\}$ is linearly independent, and since $\dim \mathbb{C}^3 = 3$ it follows that it is a basis of $\mathbb{C}^3$. Thus

$$\mathcal{R} = \mathbb{C}^3.$$

EXAMPLE 111. Consider the linear map $L\colon \mathbf{M}_3(\mathbb{R}) \to \mathbb{R}[x]$ defined on the standard basis as follows:

$$L(E_{ij}) = x^{i+j}.$$

Then we have[6]

$$L(A) = a_{11}\, x^2 + (a_{12} + a_{21}\ )\, x^3 + (a_{13} + a_{22} + a_{31}\ )\, x^4 + (a_{23} + a_{32}\ )\, x^5 + a_{33}\, x^6.$$

The range of L is then

$$\mathcal{R}(L) = \left\langle x^2, x^3, x^4, x^5, x^6 \right\rangle$$

a 5-dimensional subspace of $\mathbb{R}[x]$.

By solving the equation

$$L(A) = 0$$

we find that

$$\ker L = \left\{ \begin{pmatrix} 0 & a & b \\ -a & c & d \\ -c-d & -b & 0 \end{pmatrix} : a, b, c, d \in \mathbb{R} \right\},$$

a 4-dimensional subspace of $\mathbf{M}_3$.

Notice that

$$\dim \mathbf{M}_3 = \dim \ker L + \dim \mathcal{R}(L).$$

This indicates that a version of the rank-nullity theorem, holds in general. We will indeed see that this is the case.

DEFINITION 45. The set of linear functions $V \to W$ is denoted by $\mathcal{L}(V, W)$. If $V = W$ we simply write $\mathcal{L}(V)$.

THEOREM 4.4.4 (**A linear combination of linear maps is linear**). *Let V, W be vector spaces over a field K. Then $\mathcal{L}(V, W)$ is a vector subspace of W^V (see Example 85 for the definition of W^V).*

PROOF. Exercise. See the proofs of Theorems 3.1.6 and 3.1.7. □

THEOREM 4.4.5 (**Composition of linear maps is linear**). *Let V, U, W be vector spaces over a field K and let $g\colon V \to U$ and $f\colon U \to W$ be two linear maps. Then the composition*

$$f \circ g\colon V \to W$$

is linear.

PROOF. Identical to the proof of Theorem 3.1.9. □

THEOREM 4.4.6 (**Inverse of a linear map is linear**). *Let $f\colon V \to W$ be an invertible map. If f is linear then so is f^{-1}.*

PROOF. Recal that a function is invertible if and only if it is a bijection, and in particular an invertible function is surjective. Let then $\lambda_1, \lambda_2 \in K$ and w_1, w_2 in W and ler $f\colon V \to W$ be

[6] Verify this.

an invertible linear map. Since f is surjective, for $i = 1, 2$ we have $w_i = f(v_i)$ (or equivalently $v_i = f^{-1}(w_i)$) for some (unique) $v_i \in V$. Then,

$$\begin{aligned} f^{-1}(\lambda_1 w_1 + \lambda_2 w_2) &= f^{-1}\left(\lambda_1 f(v_1) + \lambda_2 f(v_2)\right) \\ &= f^{-1}\left(f(\lambda_1 v_1 + \lambda_2 v_2)\right) \\ &= \lambda_1 v_1 + \lambda_2 v_2 \\ &= \lambda_1 f^{-1}(w_1) + \lambda_2 f^{-1}(w_2). \end{aligned}$$

Therefore f^{-1} is also linear. □

4.5. Isomorphisms

The concept of *isomorphism* is fundamental in modern mathematics. Roughly speaking, two mathematical objects are *isomorphic* with respect to some "structure" if they are copies of each other, as far as that structure is concerned. Even when we don't explicitly mention it we use *isomorphisms*, particular identifications that preserve the structure of interest, all the time. For example we say that $\mathbb{R} \subset \mathbb{C}$, but strictly speaking that's false. For example, if we take $\mathbb{C}$ to be a 2-dimensional real vector space, with a certain multiplication defined as we do in Appendix A, $\mathbb{R}$ and $\mathbb{C}$ have no elements in common. The real numbers that sit inside $\mathbb{C}$ are an *isomorphic copy* of $\mathbb{R}$. But we don't really care, because that copy has, exactly the same properties as $\mathbb{R}$ as far we are concerned. We don't really care about ontological questions, what we care about is that $\mathbb{R}$ is a field that has certain properties, any field with those properties will do. The same goes for $\mathbb{C}$, we could consider $\mathbb{C}$ to be a set of matrices as we did in Example 70, and nothing of importance would change, we would have exactly the same theorems, because we don't really use the nature of the elements of $\mathbb{C}$, once we establish the basic properties that we want, we forget all about the fact that complex numbers were defined as matrices, or as vectors, or whatever definition we used.

When we study a vector space, the important properties from our point of view, are those properties that are defined in terms of scalar multiplication and vector addition. Things like bases, dimension, etc. If two vector spaces have all those properties the same, we might as well consider them identical. An *isomorphism of vector spaces* is then a way to identify two vector spaces, i.e. a bijection, that respects scalar multiplication and vector addition. This means that if we have identified $v \in V$ with $w \in W$ then λv should be identified with λw for all scalars λ. Similarly, if $v_1, v_2 \in V$ are identified with $w_1, w_2 \in W$, respectively, then $v_1 + v_2$ should be identified with $w_1 + w_2$. In other words, the bijection we use to identify V and W should be a *linear map*.

DEFINITION 46. If V, W are vector spaces over K, and $f\colon V \to W$ is a linear bijection, we say that f is an *isomorphism of vector spaces*, or simply an *isomorphism*.

If there is an isomorphism $f\colon V \to W$ we say that V is *isomorphic*[7] to W and we write $V \cong W$.

THEOREM 4.5.1. *Let V, U, W be vector spaces over a field K, and let $g\colon V \to U$ and $f\colon U \to W$ be isomorphisms. Then the following hold:*

(a) The identity map

$$I_V\colon V \to V, \quad I_V(v) = v$$

is an isomorphism.

[7] The term "isomorphic" comes from the greek words "ΙΣΟΣ" (isos) meaning equal or the same and "ΜΟΡΦΗ" (morphe) meaning form or shape. Thus two isomorphic vector spaces have the same form, they look the same as vector spaces.

(b) The inverse

$$f^{-1}\colon W \to U$$

is an isomorphism.

(c) The composition

$$f \circ g\colon V \to W, \quad (f \circ g)(v) = f(g(v))$$

is an isomorphism. Furthermore, its inverse

$$(f \circ g)^{-1}\colon W \to V$$

is

$$(f \circ g)^{-1} = g^{-1} \circ f^{-1}.$$

PROOF. The first item is obvious and the second was proven in Theorem 4.4.6. For the third item we have to prove three things:

(a) $f \circ g$ is linear.

Let $\lambda_1, \lambda_2 \in K$ and $v_1, v_2 \in V$. Then

$$\begin{aligned}(f \circ g)(\lambda_1 v_1 + \lambda_2 v_2) &= f(g(\lambda_1 v_1 + \lambda_2 v_2))\\ &= f(\lambda_1 g(v_1) + \lambda_2 g(v_2))\\ &= \lambda_1 f(g(v_1)) + \lambda_2 f(g(v_2))\\ &= \lambda_1 \; (f \circ g))(v_1) + \lambda_2 \; (f \circ g))(v_2),\end{aligned}$$

establishing that $f \circ g$ is linear.

(b) $f \circ g$ is injective.

Since $f \circ g$ is linear it suffices to prove that

$$\ker f \circ g = \{0\},$$

or equivalently, that for $v \in V$

$$(f \circ g)(v) = 0 \implies v = 0.$$

Indeed,

$$\begin{aligned}(f \circ g)(v) = 0 &\implies f(g(v)) = 0 && \text{by definition}\\ &\implies g(v) = 0 && f \text{ is injective}\\ &\implies v = 0 && g \text{ is injective}\;.\end{aligned}$$

Thus $f \circ g$ is injective.

(c) $f \circ g$ is surjective.

Let $w \in W$ then since f is surjective, there is a (unique) $u \in U$ such that $f(u) = w$. Now, since g is surjective, there is a (unique) $v \in V$ with $g(v) = u$. So

$$w = f(u) = f(g(v)) = (f \circ g)(v).$$

Thus $f \circ g$ is surjective.

Thus $f \circ g$ is an isomorphism.

The identity

$$(f \circ g)^{-1} = g^{-1} \circ f^{-1}$$

holds for any composable bijections. We have already seen of it, but let's prove it again anyway.

For $w \in W$ we have

$$f\left(g\left(g^{-1}\left(f^{-1}(w)\right)\right)\right) = f\left(f^{1}(w)\right) = w,$$

and similarly for $v \in V$

$$g^{-1}\left(f^{-1}\left(f\left(g(v)\right)\right)\right) = g^{-1}\left(g(v)\right) = v.$$

□

We have then the following corollary.

COROLLARY 4. *The relation of isomorphism is an equivalence relation. That is*

(a) It is reflexive. That is for any vector space V we have

$$V \cong V.$$

(b) It is symmetric. That is for any vector spaces V, U we have

$$V \cong U \implies U \cong V.$$

(c) It is transitive. That is for any vector spaces V, U, W we have

$$V \cong U \text{ and } U \cong W \implies V \cong W.$$

PROOF. Exercise. Use Theorem 4.5.1 to find the required isomorphisms. □

We already have seen many examples of isomorphisms $\mathbb{R}^n \to \mathbb{R}^n$. Indeed the linear maps defined by invertible matrices are isomorphisms.

EXAMPLE 112. Consider the subspace[8]

$$V = \{(x, y, 0)\} \subseteq \mathbb{R}^3.$$

Then

$$L\colon V \to \mathbb{R}^2, \quad L(x, y, 0) = (x, y)$$

is an isomorphism.

Clearly L is linear. It is also, but perhaps not quite so clearly, a bijection. To see that let

$$M\colon \mathbb{R}^2 \to V, \quad M(x, y) = (x, y, 0).$$

Then

$$L\left(M(x, y)\right) = L(x, y, 0) = (x, y), \quad M\left(L(x, y, 0)\right) = M(x, y) = (x, y, 0).$$

Thus $M = L^{-1}$ and so L is a bijection.

EXAMPLE 113. The subspace

$$V = \{(2\,x - 3\,z, y + z, x + 2\,z, x) : x, y, z \in \mathbb{R}\} \subseteq \mathbb{R}^4$$

is isomorphic to the subspace

$$W = \{(x, x + y, x + z, x + y + z, x - z) : x, y, z \in \mathbb{R}\} \subseteq \mathbb{R}^4.$$

Indeed,

$$L\colon V \to W, \quad L(2\,x - 3\,z, y + z, x + 2\,z, x) = (x, x + y, x + z, x + y + z, x - z)$$

is an isomorphism.

Let's prove first that L is linear. For $i = 1, 2$, let $\lambda_i \in \mathbb{R}$ and $\mathbf{v}_i = (2\,x_i - 3\,z_i, y_i + z_i, x_i + 2\,z_i, x_i) \in V$.

[8] Can you see that this is really a subspace? Prove it.

Using column vectors for readability we have

$$\lambda_1\,\mathbf{v}_1+\lambda_2\,\mathbf{v}_2=\lambda_1\begin{pmatrix}2\,x_1-3\,z_1\\ y_1+z_1\\ x_1+2\,z_1\\ x_1\end{pmatrix}+\lambda_2\begin{pmatrix}2\,x_2-3\,z_2\\ y_2+z_2\\ x_2+2\,z_2\\ x_2\end{pmatrix}$$

$$=\begin{pmatrix}2\,\lambda_1\,x_1-3\,\lambda_1\,z_1+2\,\lambda_2\,x_2-3\,\lambda_2\,z_2\\ \lambda_1\,y_1+\lambda_1\,z_1+\lambda_2\,y_2+\lambda_2\,z_2\\ \lambda_1\,x_1+2\,\lambda_1\,z_1+\lambda_2\,x_2+2\,\lambda_2\,z_2\\ \lambda_1\,x_1+\lambda_2\,x_2\end{pmatrix}$$

$$=\begin{pmatrix}2\,(\lambda_1\,x_1+\lambda_2\,x_2)-3\,(\lambda_1\,z_1+\lambda_2\,z_2)\\ (\lambda_1\,y_1+\lambda_2\,y_2)+(\lambda_1\,z_1+\lambda_2\,z_2)\\ (\lambda_1\,x_1+\lambda_2\,x_2)+2\,(\lambda_1\,z_1+\lambda_2\,z_2)\\ \lambda_1\,x_1+\lambda_2\,x_2\end{pmatrix}.$$

Therefore,

$$L(\lambda_1\,\mathbf{v}_1+\lambda_2\,\mathbf{v}_2)=\begin{pmatrix}\lambda_1\,x_1+\lambda_2\,x_2\\ (\lambda_1\,x_1+\lambda_2\,x_2)+(\lambda_1\,y_1+\lambda_2\,y_2)\\ (\lambda_1\,x_1+\lambda_2\,x_2)+(\lambda_1\,y_1+\lambda_2\,y_2)\\ (\lambda_1\,x_1+\lambda_2\,x_2)+(\lambda_1\,z_1+\lambda_2\,z_2)\\ (\lambda_1\,x_1+\lambda_2\,x_2)+(\lambda_1\,y_1+\lambda_2\,y_2)+(\lambda_1\,y_1+\lambda_2\,y_2)\\ (\lambda_1\,x_1+\lambda_2\,x_2)-(\lambda_1\,z_1+\lambda_2\,z_2)\end{pmatrix}$$

Similar calculations show that

$$\lambda_1\,L(\mathbf{v}_1)+\lambda_2\,L(\mathbf{v}_2)=\begin{pmatrix}\lambda_1\,x_1+\lambda_2\,x_2\\ (\lambda_1\,x_1+\lambda_2\,x_2)+(\lambda_1\,y_1+\lambda_2\,y_2)\\ (\lambda_1\,x_1+\lambda_2\,x_2)+(\lambda_1\,y_1+\lambda_2\,y_2)\\ (\lambda_1\,x_1+\lambda_2\,x_2)+(\lambda_1\,z_1+\lambda_2\,z_2)\\ (\lambda_1\,x_1+\lambda_2\,x_2)+(\lambda_1\,y_1+\lambda_2\,y_2)+(\lambda_1\,y_1+\lambda_2\,y_2)\\ (\lambda_1\,x_1+\lambda_2\,x_2)-(\lambda_1\,z_1+\lambda_2\,z_2)\end{pmatrix},$$

establishing the linearity of L.

Now, L is invertible and its inverse is

$$L^{-1}\colon W\to V,\quad L(x,x+y,x+z,x+y+z,x-z)=(2\,x-3\,z,y+z,x+2\,z,x),$$

as can be easily seen by computing $L\circ L^{-1}$ and $L^{-1}\circ L$.

EXAMPLE 114. Let $\mathbf{\Delta}_4$ be the vector space of upper triangular 4×3 matrices and $\mathbf{S}_4$ the vector space of symmetric 4×4 matrices. Let $L\colon \mathbf{\Delta}_4\to\mathbf{S}_4$ be given by

$$\begin{pmatrix}a & x & y & z\\ 0 & b & p & q\\ 0 & 0 & c & s\\ 0 & 0 & 0 & d\end{pmatrix}\longmapsto\begin{pmatrix}a & x & y & z\\ x & b & p & q\\ y & p & c & s\\ z & q & s & d\end{pmatrix}.$$

Then L is an isomorphism.

It is clear that L is a bijection. One way to see this is to prove that it has an iverse, that is there is a function $L^{-1}\colon\mathbf{S}_4\to\mathbf{\Delta}_4$ such that for $X\in\mathbf{\Delta}_4$ and $Y\in\mathbf{S}_4$ we have

$$L^{-1}\left(L(X)\right)=X\text{ and }L\left(L^{-1}(Y)\right)=Y. \tag{4.5}$$

Indeed if we define $L^{-1}\colon \mathbf{S}_4 \to \mathbf{\Delta}_4$ by $L^{-1}\colon \mathbf{S}_4 \to \mathbf{\Delta}_4$ by

$$\begin{pmatrix} a & x & y & z \\ x & b & p & q \\ y & p & c & s \\ z & q & s & d \end{pmatrix} \longmapsto \begin{pmatrix} a & x & y & z \\ 0 & b & p & q \\ 0 & 0 & c & s \\ 0 & 0 & 0 & d \end{pmatrix},$$

then clearly Conditions (4.5) hold.

Linearity can be proved by straightforward calculations. A slicker way to prove it is to note that

$$L = I + T - D$$

where I is the identity map, T is the map defined in Example 105, and D is the map defined in Example 106. Therefore, by Theorem 4.4.4 L is linear.

EXAMPLE 115. Let $a, b \in K$ and consider the following subspaces of $K[x]$

$$V = \{p(x) : p(a) = 0\}, \quad W = \{p(x) : p(b) = 0\}.$$

To see that V and W are really subspaces notice that

$$V = \ker E_a, \quad W = \ker E_b$$

where E_a (respectively E_b) is the "evaluate at a" (respectively at b) as in Example 104.

I claim that $V \cong W$. Indeed, by Theorem C.1.5, we have for every $p(x) \in V$ we have the quotient

$$\frac{p(x)}{x-a} \in K[x].$$

We define then,

$$L\colon V \to W, \quad p(x) \longmapsto \frac{p(x)}{x-a}(x-b).$$

We have

$$L(\lambda p(x) + \mu q(x)) = \frac{\lambda p(x) + \mu q(x)}{x-a}(x-b) = \lambda \frac{p(x)}{x-a}(x-b) + \mu \frac{q(x)}{x-a}(x-b) = \lambda L(p(x)) + \mu L(q(x)),$$

and so L is linear.

L is also a bijection with inverse

$$L^{-1}\colon W \to V, \quad p(x) \longmapsto \frac{p(x)}{x-b}(x-a),$$

as we easily verify.

4.5.1. Go forth, do your business, come back.

In the first three chapters we developed a powerfull method, namely getting matrices to reduced row echelon form, that can answer all kinds of questions in $\mathbb{R}^n$. We can solve linear systems, find inverses of linear transformations, finding bases of subspaces and so on. As we remarked in Section 4.1.1, and as we have seen in examples, these methods work for the vector spaces K^n, as well, where K is an arbitrary field. Since all vector spaces of dimension n are isomorphic to K^n, it turns out that we can use these methods for any finite dimensional vector space.

The basic idea is indicated schematically in Figure 3. Let's say we have a question in an arbitrary n-dimensional vector space V, then we use an isomorphism $L\colon V \to K^n$ to *transfer* the question to K^n. For example if we want to express a vector $v \in V$ as a linear combination of $S \subseteq V$, we go to K^n and express $L(v)$ as a linear combination of $L(S)$. Once we get our answer, we use the inverse isomorphism L^{-1} to transfer it back to V.

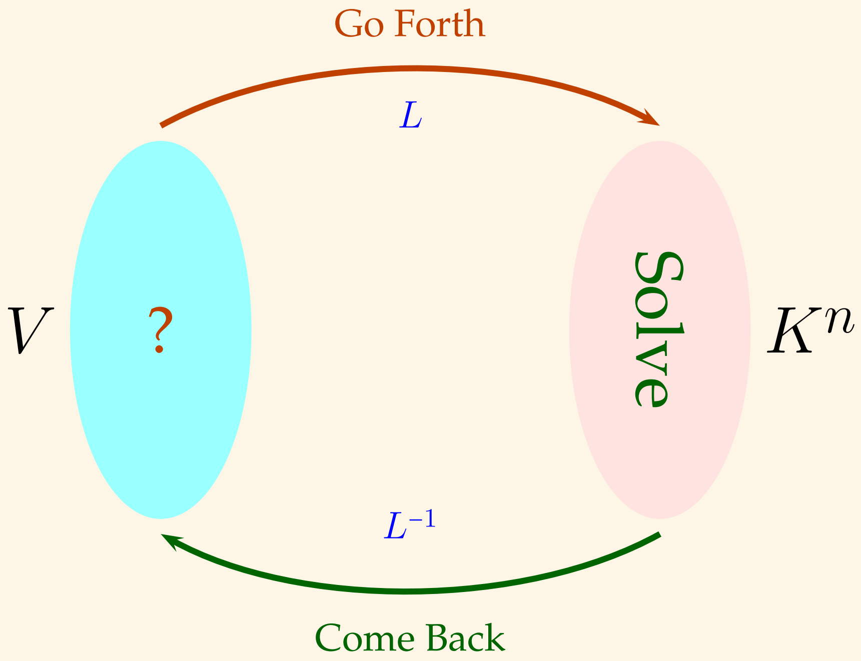


FIGURE 3. Go forward, do stuff, come back

EXAMPLE 116. Consider the polynomials

$$B = \left\{1, x+1, (x+1)^2, (x+1)^3, (x+1)^4\right\} \subseteq \mathbf{P}_4.$$

Let's prove that B is a basis of $\mathbf{P}_4$ and then express the polynomial $p(x) = 3\,x^4 - 5\,x^3 - x + 4$ as a linear combination of B.

Using the binomial theorem we express the elements of B in terms of the standard basis:

$$\begin{aligned} 1 &= 1 \\ x+1 &= x+1 \\ (x+1)^2 &= x^2 + 2\,x + 1 \\ (x+1)^3 &= x^3 + 3\,x^2 + 3\,x + 1 \\ (x+1)^4 &= x^4 + 4\,x^3 + 6\,x^2 + 4\,x + 1 \end{aligned}$$

.

Using the isomorphism $L\colon \mathbf{P}^4 \to \mathbb{R}^5$ with

$$L(1) = \mathbf{e}_1, \ldots, L(x^4) = \mathbf{e}_5$$

we have the image $L(B)$ consists of the collumns of the matrix

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 1 & 3 & 6 \\ 0 & 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Thus $L(B)$ is linearly independent, and therefore a basis of $\mathbb{R}^5$. It follows that B is a basis of $\mathbf{P}_4$.

Now, $L(p(x)) = (3, -5, 0, -1, 4)$. Taking the reduced echelon form of the augmented matrix we get

$$\left(\begin{array}{ccccc|c} 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 & 4 & -5 \\ 0 & 0 & 1 & 3 & 6 & 0 \\ 0 & 1 & 2 & 3 & 4 & -1 \\ 1 & 1 & 1 & 1 & 1 & 4 \end{array}\right) \sim \left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 9 \\ 0 & 1 & 0 & 0 & 0 & -16 \\ 0 & 0 & 1 & 0 & 0 & 21 \\ 0 & 0 & 0 & 1 & 0 & -13 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{array}\right).$$

Therefore

$$L(p(x)) = 9\,L(1) - 16\,L(1+x) + 21\,L\left((1+x)^2\right) - 13\,L\left((1+x)^3\right) + 3\,L\left((1+x)^4\right).$$

Going back to $\mathbf{P}_4$ via L^{-1} we have that

$$p(x) = 9 - 16\,(1+x) + 21\,(1+x)^2 - 13\,(1+x)^3 + 3\,(1+x)^4.$$

EXAMPLE 117. Lets find the dimension of the linear span of the following 2×2 real matrices:

$$\begin{pmatrix} 1 & -5 \\ -4 & 2 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ -1 & 5 \end{pmatrix} \quad \begin{pmatrix} 2 & -4 \\ -5 & 7 \end{pmatrix} \quad \begin{pmatrix} 1 & -7 \\ -5 & 1 \end{pmatrix}.$$

Let $L\colon \mathbf{M}_4 \to \mathbb{R}^4$ be the isomorphism that sends E_{11} to $\mathbf{e}_1$, E_{12} to $\mathbf{e}_2$, E_{21} to $\mathbf{e}_3$, and E_{22} to $\mathbf{e}_4$. Then the matrices transform to the vectors

$$(1,-5,-4,2) \quad (1,1,-1,5) \quad (2,-4,-5,7) \quad (1,-7,-5,1).$$

We have

$$\begin{pmatrix} 1 & 1 & 2 & 1 \\ -5 & 1 & -4 & -7 \\ -4 & -1 & -5 & -5 \\ 2 & 5 & 7 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 & 4/3 \\ 0 & 1 & 1 & -1/3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus the dimension of the linear span is 2.

EXAMPLE 118. Consider $\mathbf{P}_4$ and let

$$B = \{1, x, x\,(x+1), x\,(x+1)\,(x+2), x\,(x+1)\,(x+2)\,(x+3)\}\,.$$

(a) Verify that B is a basis.
(b) Let $T\colon \mathbf{P}_4 \to \mathbf{P}_4$ be the isomorphism defined by

$$T(x^n) = \prod_{k=0}^{n-1}(x+1), n = 0,1,2,3,4.$$

For $n = 0$ we have the empty product that by convention is 1. Find a formula for T.
(c) Find a formula for T^{-1}.

We will use the isomorphism defined by

$$L(x^k) = \mathbf{e}_{k+1}, i = 0,1,2,3,4.$$

We have

$$\begin{aligned} 1 &= 1 \\ x &= x \\ x\,(x+1) &= x^2 + x \\ x\,(x+1)\,(x+2) &= x^3 + 3\,x^2 + 2\,x \\ x\,(x+1)\,(x+2)\,(x+3) &= x^4 + 6\,x^3 + 11\,x^2 + 6\,x. \end{aligned}$$

Now using L we consider the linear transformation $A = L \circ T \circ L^{-1}$ as shown in the diagram below.

$$\begin{tikzcd}
\mathbf{P}_4 \arrow[r, "L"] \arrow[d, "T"'] & \mathbb{R}^5 \arrow[d, "A"] \\
\mathbf{P}_4 \arrow[r, "L"'] & \mathbb{R}^5
\end{tikzcd}$$

We have

$$A\,\mathbf{e}_1 = L\left(T\left(L^{-1}(\mathbf{e}_1)\right)\right) = L\left(T(1)\right) = L(1) = \mathbf{e}_1.$$

Similarly

$$A\,\mathbf{e}_2 = {}_2,\ \ A\,\mathbf{e}_3 = \mathbf{e}_2 + \mathbf{e}_3,\ \ A\,\mathbf{e}_4 = 2\,\mathbf{e}_2 + 3\,\mathbf{e}_3 + \mathbf{e}_4,\ \ A\,\mathbf{e}_5 = 6\,\mathbf{e}_2 + 11\,\mathbf{e}_3 + 6\,\mathbf{e}_4 + \mathbf{e}_5.$$

This means that A is given by the matrix

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 & 6 \\ 0 & 0 & 1 & 3 & 11 \\ 0 & 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

We compute the inverse of A and we find

$$A^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & -1 \\ 0 & 0 & 1 & -3 & 7 \\ 0 & 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

So,

$$T = L^{-1} \circ A \circ L, \quad T = L^{-1} \circ A^{-1} \circ L.$$

What is then $T(9\,x^4 - 3\,x^3 + 2\,x^2 - 4\,x + 7)$?

$$A(L(p(x))) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 & 6 \\ 0 & 0 & 1 & 3 & 11 \\ 0 & 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 9 \\ -3 \\ 2 \\ -4 \\ 7 \end{pmatrix} = \begin{pmatrix} 9 \\ 33 \\ 67 \\ 38 \\ 7 \end{pmatrix}.$$

Therefore

$$T(p(x)) = 9 + 33\,x + 67\,x^2 + 38\,x^3 + 7\,x^4.$$

If we have a linear map

$$T\colon : V \to W$$

between two different vector spaces, say of dimentsion n and m we need two isomorphisms

$$L_1\colon V \to \mathbb{R}^n, \quad L_2\colon W \to \mathbb{R}^m$$

$$\begin{tikzcd}
W \arrow[r, "L_1"] \arrow[d, "T"'] & \mathbb{R}^n \arrow[d, "A"] \\
V \arrow[r, "L_2"'] & \mathbb{R}^m
\end{tikzcd}$$

and we'll get a linear map $A\colon \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$A = L_2 \circ T \circ L_1^{-1}.$$

EXAMPLE 119. Let's look at the linear function $L\colon: \mathbf{M}_3(\mathbb{R}) \to \mathbf{P}_6$ and use the isomorphisms defined by

$$L_1(E_{ij}) = \mathbf{e}_{3\,(i-1)+j}, \quad L_2(x^n) = \mathbf{e}_{n+1}.$$

We get the 6×9 matrix

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

And we see that a basis for the range of A is given by the the columns

$$\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_6, \mathbf{a}_9,\} = \{\mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_5, \mathbf{e}_6, \mathbf{e}_7\}.$$

Thus we get the basis $\{x^2, x^3, x^4, x^5, x^6\}$ for the range of L.

A basis for the kernel of A is given by the free columns with the standard basis of $\mathbb{R}^4$ interpolated. So we get the basis

$$\{(0,-1,0,1,0,0,0,0,0), (0,0,-1,0,1,0,0,0,0), (0,0,-1,0,0,0,1,0,0), (0,0,0,-1,0,0,0,1,0)\}.$$

We can write this basis as

$$\{\mathbf{e}_4 - \mathbf{e}_2, \mathbf{e}_5 - \mathbf{e}_3, \mathbf{e}_7 - \mathbf{e}_3, \mathbf{e}_8 - \mathbf{e}_4\},$$

and thus we get the following basis for the kernel of L:

$$\{E_{21} - E_{12}, E_{22} - E_{13}, E_{31} - E_{13}, E_{32} - E_{21}\},$$

i.e.

$$\left\{ \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right\}.$$

CHAPTER 5

Products

5.1. Inner products

5.1.1. Angles, orthogonality.

5.1.2. Orthogonal complement.

5.1.3. Gram-Schmidt.

5.2. Exterior products

5.2.1. Exterior Algebra.

5.2.2. Determinants and their properties.

5.2.3. Lenght, Area, Volume, Orientation.

CHAPTER 6

Eigenvalues, Eigenvectors, Eigenspaces

6.1. 2×2 linear transformations

6.2. Characteristic polynomial

6.3. Diagonalization

APPENDIX A

Complex Numbers

A.1. $\mathbb{C}$ as an algebra over $\mathbb{R}$

We define $\mathbb{C}$ as a commutative 2-dimensional algebra over the real field $\mathbb{R}$, that contains a square root of -1. Thus we consider the real vector space with basis $\{1, i\}$ and we want to define a commutative, associative multiplication that distributes over vector addition, and such that $i^2 = -1$.

Therefore we define the *set of complex numbers* to be the linear span

$$\mathbb{C} = \{a + b\,i : a, b \in \mathbb{R}\}$$

where $\{1, i\}$ is linearly independent. This means that

$$a + b\,i = c + d\,i \iff a = c \text{ and } b = d,$$

addition is defined by

$$(a + b\,i) + (c + d\,i) = (a + c) + (b + d)\,i,$$

and scalar multiplication is defined by

$$\lambda\,(a + b\,i) = (\lambda\,a) + (\lambda\,b)\,i.$$

The linear span $\mathbb{R}\langle 1\rangle$ is called the *real axis* and we identify its elements with real numbers, and consider $\mathbb{R} \subseteq \mathbb{C}$. The linear span $\mathbb{R}\langle i\rangle$ is called the *imaginary axis* and its elements are called *imaginary numbers*, in particular, $i = 1\,i$ is called the *imaginary unit*.

For a complex number $z = a + b\,i$ we say that a is its *real part* and b its imaginary part, and write

$$\Re z = a, \quad \Im z = b.$$

We identify the vector space $\mathbb{C}$ with the standard real vector space $\mathbb{R}^2$ using the isomorphism

$$T\colon \mathbb{R}^2 \to \mathbb{C}, \quad T\,\mathbf{e}_1 = 1,\ T\,\mathbf{e}_2 = i,$$

and consider complex numbers as standard real 2-dimensional vectors, see Figure 1. We refer to this interpretation as “the Cartesian representation of complex numbers”.

The multiplication in $\mathbb{C}$ is determined by the requirements that it forms a commutative algebra over $\mathbb{R}$ and $i^2 = -1$. Indeed, under these requirements we have

$$\begin{aligned}(a + b i)\,(c + d\,i) &= a\,c + a\,d\,i + b\,i\,c + b\,d i^2\\ &= a\,c + a\,d\,i + b\,c\,i - b\,d\\ &= (a\,c - b\,d) + (a\,d + b\,c)\,i.\end{aligned}$$

You should be able to prove the following theorem, we’ve seen all the necessary ingredients already.

THEOREM A.1.1. *Define multiplication of complex numbers via*

$$(a + b\,i)\,(c + d\,i) = (a\,c - b\,d) + (a\,d + b\,c)\,i.$$

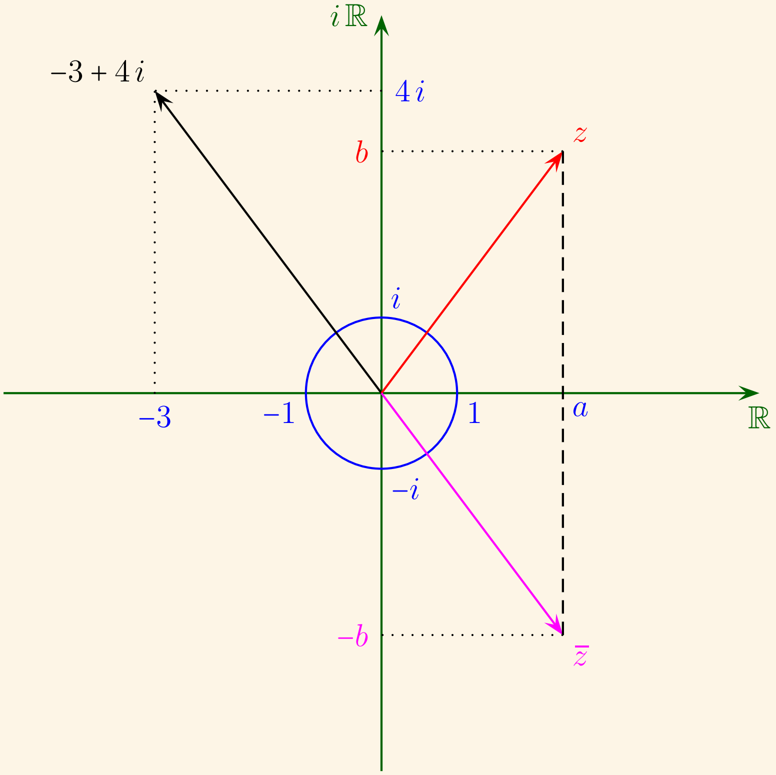


FIGURE 1. The Cartesian representation of complex numbers.

Then $\mathbb{C}$ is a superfield of $\mathbb{R}$. In particular the inverse of $z = a + b\,i$ is

$$z^{-1} = \frac{a - b\,i}{a^2 + b^2}.$$

PROOF. Left as an exercise. □

If $z = a + b\,i$ then we call $\overline{z} = a - b\,i$ the *complex conjugate* of z. Geometrically $\overline{z}$ is given by reflecting z across the real axis, see Figure 1.

The operation of complex conjugation has the following properties.

PROPOSITION 12 (**Properties of complex conjugation**). *The following hold:*

(a) $\overline{\overline{z}} = z$.
(b) $\overline{z + w} = \overline{z} + \overline{w}$.
(c) $\overline{z\,w} = \overline{z}\,\overline{w}$.
(d) $z + \overline{z} = 2\,\Re z, \quad z - \overline{z} = 2\,\Im z$.
(e) $z \in \mathbb{R} \iff \overline{z} = z$.
(f) $z \in i\,\mathbb{R} \iff \overline{z} = -z$.
(g) $z\,\overline{z} = (\Re z)^2 + (\Im z)^2$.
(h) *If $p(x)$ is a polynomial with real coefficients then for all $z \in \mathbb{C}$ we have*

$$p(\overline{z}) = \overline{p(z)}.$$

In particular, complex roots of real polynomials come in conjugate pairs.

PROOF. Exercise. □

The *absolute value* (or *length*, or *modulus*, or *norm*) of a complex number $z = a + b\,i$ is defined to be

$$|\,z\,| = \sqrt{z\,\overline{z}}.$$

Since

$$\sqrt{z\,\overline{z}} = \sqrt{a^2 + b^2},$$

we see that the absolute value of z equals the standard Euclidean norm of z considered as a vector in $\mathbb{R}^2$.

We list some properties of the absolute value.

THEOREM A.1.2 (**Properties of the absolute value**). *The following hold.*

(a) $z^{-1} = \dfrac{\overline{z}}{|z|^2}$.

(b) $|\overline{z}| = |z|$.

(c) $|\mathfrak{R}z| \le |z|, \quad |\mathfrak{I}z| \le |z|$.

(d) $|z| \ge 0$[1]*, and* $|z| = 0$ *if and only if* $z = 0$.

(e) $|z\,w| = |z|\,|w|$. *In particular,*

$$|z| = |w| = 1 \implies |z\,w| = 1.$$

(f) *(*Triangle inequality*)*

$$|z + w| \le |z| + |w|.$$

PROOF. (a) We have

$$\frac{\overline{z}}{|z|} = \frac{\overline{z}}{z\overline{z}} = \frac{1}{z}.$$

(b) We have

$$|\overline{z}|^2 = \overline{z}\,\overline{\overline{z}} = \overline{z}\,z = z\,\overline{z} = |z|^2.$$

(c) Let $z = a + b\,i$, with $a, b \in \mathbb{R}$. Then $a^2, b^2 \ge 0$ and so

$$a^2 \le a^2 + b^2, \quad b^2 \le a^2 + b^2,$$

and the result follows by taking square roots.

(d) For two real numbers a, b we have

$$a^2 + b^2 = 0 \iff a = 0 \text{ and } b = 0.$$

(e) We have

$$\begin{aligned} |z\,w|^2 &= (z\,w)\,\overline{z\,w} \\ &= z\,w\,\overline{z}\,\overline{w} \\ &= z\,\overline{z}\,w\,\overline{w} \\ &= |z|^2\,|w|^2. \end{aligned}$$

(f) We have

$$\begin{aligned} |z + w|^2 &= (z + w)\,\overline{z + w} \\ &= (z + w)\,(\overline{z} + \overline{w}) \\ &= z\,\overline{z} + z\,\overline{w} + w\,\overline{z} + w\,\overline{w} \\ &= |z|^2 + 2\,\mathfrak{R}(z\,\overline{w}) + |w|^2, \end{aligned}$$

[1] There is no order defined for complex numbers. So when we write $z \ge 0$ we mean that z is real and non-negative.

where the last line follows from $\overline{z\,\overline{w}} = \bar{z}\,w$. Now, using Item (c), we continue

$$\begin{aligned} &\leq |\,z\,|^2 + 2\,|\,z\,\overline{w}\,| + |\,w\,|^2 \\ &= |\,z\,|^2 + 2\,|\,z\,|\,|\,\overline{w}\,| + |\,w\,|^2 \\ &= |\,z\,|^2 + 2\,|\,z\,|\,|\,w\,| + |\,w\,|^2 \\ &= (|\,z\,| + |\,w\,|)^2 . \end{aligned}$$

So we've shown,

$$|\,z + w\,|^2 \leq (|\,z\,| + |\,w\,|)^2 ,$$

and the result follows by taking square roots.

□

REMARK 21. Triangle inequality says that the sum of the lengths of two sides of a triangle is always greater than the length of the other side, see Figure 2.

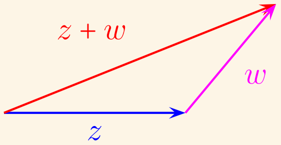


FIGURE 2. Triangle inequality.

The set of complex numbers of length 1 is called the unit circle and is denoted by S^1, that is

$$S^1 = \{z \in \mathbb{C} : |\,z\,| = 1\} .$$

Of course, when we think of complex numbers as points in $\mathbb{R}^2$, the unit circle is the circle of radius 1 centered at the origin $(0,0)$. Indeed, if $z = x + i\,y$ then

$$|\,z\,| = 1 \iff x^2 + y^2 = 1$$

and the latter is the equation a circle of radius 1 and center $(0,0)$.

An important corollary of Item (e) of Theorem A.1.2 is the following theorem.

THEOREM A.1.3. *The unit circle S^1 endowed with multiplication of complex numbers is a commutative group.*

A.1.1. Polar representation of complex numbers. Recall that one definition of the sine and cosine of an angle θ is via the y and x coordinates, respectively of a point in the unit circle, namely the point where the ray from the origin that forms an angle θ with the x-axis intersects the unit circle. It follows then that if $z \in S^1$ then

$$z = \cos\theta + i\,\sin\theta,$$

for some angle θ.

We now prove a formula that allow us to give a nice geometric interpretation of multiplication of complex numbers of length 1.

THEOREM A.1.4 (**de Moivre's formula**). *The following hold.*

(a) For $\varphi, \theta \in \mathbb{R}$ we have

$$(\cos\theta + i\,\sin\theta)\,(\cos\varphi + i\,\sin\varphi) = \cos(\theta + \varphi) + i\,\sin(\theta + \varphi).$$

(b) For $\theta \in \mathbf{R}$ and $n \in \mathbb{N}$ we have

$$(\cos\theta + i\,\sin\theta)^n = \cos n\,\theta + i\,\sin n\,\theta.$$

PROOF. (a) We have

$$\begin{aligned}(\cos\theta + i\,\sin\theta)\,(\cos\varphi + i\,\sin\varphi) &= (\cos\theta\,\cos\varphi - \sin\theta\,\sin\varphi) + i\,(\cos\theta\sin\varphi + \cos\varphi\,\sin\theta)\\ &= \cos(\theta+\varphi) + i\,\sin(\theta+\varphi).\end{aligned}$$

(b) We proceed by induction. For $n = 0$ the formula is true since

$$(\cos\theta + i\,\sin\theta)^0 = 1, \quad \cos 0 = 1, \quad \sin 0 = 0.$$

Assuming now that the formula is true for n we have

$$\begin{aligned}(\cos\theta + i\,\sin\theta)^{n+1} &= (\cos\theta + i\,\sin\theta)^n\,(\cos\theta + i\,\sin n\,\theta)\\ &= (\cos n\,\theta + i\,\sin n\,\theta)\,(\cos\theta + i\,\sin n\,\theta)\\ &= (\cos n\,\theta\,\cos\theta - \sin n\,\theta\,\sin\theta) + i\,(\cos n\,\theta\,\sin\theta - \sin n\,\theta\,\cos\theta)\\ &= \cos(n+1)\,\theta + i, \sin(n+1)\,\theta.\end{aligned}$$

□

Thus, multiplication in S^1 is just addition of angles, see Figure 3.

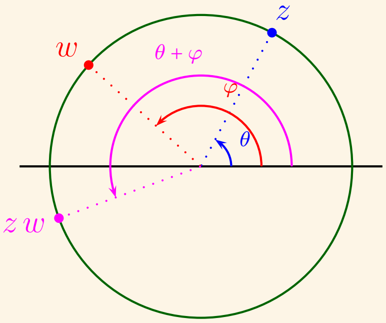


FIGURE 3. Multiplication in S^1.

We now introduce the exponential notation for unit complex numbers. In Equations (A.1) and (A.2), e stands for the base of natural logarithms $e \approx 2.71828182845905\ldots$.

DEFINITION 47 (**Exponential notation for polar form**). For $\theta \in \mathbb{R}$ we define

(A.1) $$e^{i\,\theta} = \cos\theta + i\,\sin\theta.$$

More generally, for $z = a + b\,i \in \mathbb{C}$ we define

(A.2) $$e^{z} = e^{a}\,\cos b + i\,\sin b.$$

With this notation de Moivre's formulas become

de Moivre's formulas

$$e^{i\,(\varphi+\theta)} = e^{i\,\varphi}\,e^{i\,\theta}, \qquad \left(e^{i\,\theta}\right)^n = e^{i\,n\,\theta}.$$

Now let z be a non-zero complex number. Then the linear span $\mathbb{R}\langle z\rangle$ is, geometrically, a line passing through the origin and z. If we write

$$z = |z|\,\frac{z}{|z|},$$

then since

$$\left|\frac{z}{|z|}\right| = \frac{|z|}{|z|} = 1,$$

we have

$$\frac{z}{|z|} = e^{i\theta}$$

for some angle θ. Of course θ is not uniquely determined because for all $k \in \mathbb{N}$

$$e^{2ik\pi} = 1$$

and therefore

$$e^{i\theta} = e^{i(\theta+2k\pi)}.$$

Nevertheless, θ is determined $\mod 2\pi$, in the sense that

$$e^{i\theta} = e^{i\varphi} \implies \theta - \varphi = 2k\pi$$

for some $k \in \mathbb{N}$. We call any such angle an *argument* of z and denote it by $\arg z$.

In summary, every non-zero complex number has a *polar representation*

$$z = |z|\, e^{i\arg z}.$$

If we think of $\mathbb{C}$ as the Euclidean plane $\mathbb{R}^2$, the modulus $|z|$ and the argument $\arg z$ are the so-called *polar coordinates* of z. Two points that lie in the same line through the origin if and only if they have the same argument (modulo 2π), and two points lie on the same circle centered at the origin if and only if they have the same modulus, see Figure 4. Note that the argument of 0 is undefined.

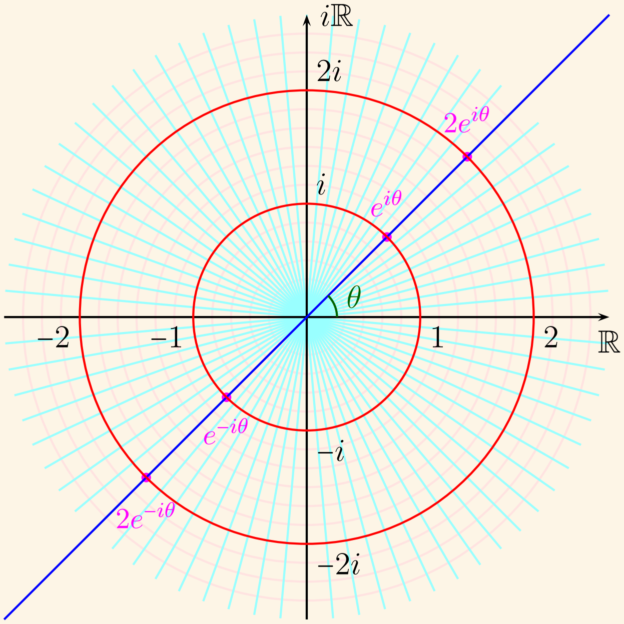


FIGURE 4. Polar representation of complex numbers.

A.2. Roots of complex numbers

The polar representation of complex numbers makes it easy to see that every non-zero complex number z has n distinct n-th roots for all positive integers n. Since an n-th root of z is a root of the polynomial $x^n - z$, by Corollary 7 in Appendix C, there are at most n, n-th roots of z. Thus it suffices to show that z has at least n, n-th roots.

We start by exhibiting n distinct n-th roots of 1. Let Ω_n be the following subset of S^1

$$\Omega_n = \left\{ e^{2\,k\,\pi\,i/n} : k = 0, 1, \ldots, n-1 \right\}.$$

Notice that

$$\Omega_n = \left\{ \omega_0^k : k = 0, 1, \ldots, n-1 \right\}$$

where $\omega_0 = e^{2\,\pi\,i/n}$. Now

$$\omega_0^n = \left(e^{2\,\pi\,i/n}\right)^n = e^{2\,\pi,\text{ß}} = 1,$$

and so for all $k \in \mathbb{N}$, we have

$$\left(\omega_0^k\right)^n = \left(\omega_0^n\right)^k = 1.$$

Thus all elements of Ω_n are n-th roots of 1.

Geometrically, the elements of Ω_n are the corners of a regular n-gon centered at the origin, see for example Figure 5 for the six sixth roots of unity.

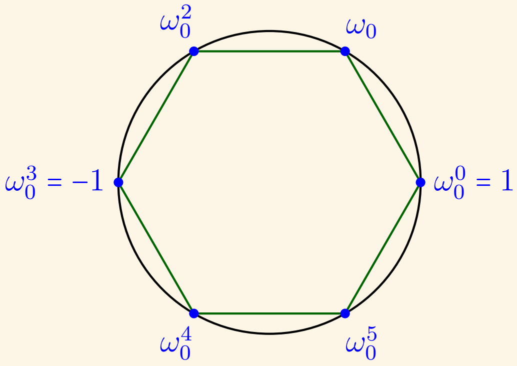


FIGURE 5. The six sixth roots of unity.

We can now prove the general case.

THEOREM A.2.1 (**n-th roots of complex numbers**). *Let z be a non-zero complex numbers and n a positive integer. Then there are n distinct n-th roots of z namely*

$$\sqrt[n]{|\,z\,|}\, e^{\arg z/n}\, \omega_0^k, \qquad k = 0, 1, \ldots, n-1$$

where $\omega_0 = e^{2\,k\,\pi\,i/n}$.

PROOF. We have

$$\left(\sqrt[n]{|\,z\,|}\, e^{\arg z/n}\, \omega_0^k\right)^n = |\,z\,|\; e^{\arg z} = z.$$

□

In particular, every non-zero complex number has two square roots.

THEOREM A.2.2 (**Quadratic formula**). *Every complex quadratic polynomial*

$$p(x) = a\,z^2 + b\,z + c, \quad a \neq 0$$

has two roots

$$z = \frac{-b \pm \sqrt{D}}{2\,a}$$

where $D = b^2 - 4\,a\,c$ *is the* discriminant *of* $p(x)$. *If the discriminant is* 0 *the two roots coincide, and* $p(x)$ *has a double root*

$$z = -\frac{b}{2\,a}.$$

PROOF. Exercise, just substitute and verify. □

It turns out that every polynomial in $\mathbb{C}$ has solutions. This result, known as *The Fundamental Theorem of Algebra*, is one of the reasons that complex numbers are useful. We state the theorem without proof.

THEOREM A.2.3 ($\mathbb{C}$ **is algebraically closed**). *Every degree* n *complex polynomial has exactly* n *roots, counted with multiplicity*[2].

[2] See Appendix C for what this means.

APPENDIX B

A bit of number theory

B.1. Divisibility and primes

We collect some useful facts (mostly without proofs) about the divisibility relation in $\mathbb{Z}$.

DEFINITION 48 (**Divisibility of integers**). Let $m, n \in \mathbb{Z}$. We say that m *divides* n, or that m is a *divisor* of n, or that n is a *multiple* of m, if

$$n = m \cdot k, \text{ for some } k \in \mathbb{Z}.$$

We write $m \mid n$ to mean that m divides n.

We say that a positive integer $p \geq 2$ is prime if

$$m \mid p \implies m = \pm p, \text{ or } m = \pm 1.$$

NOTE. Please do not confuse $\mid$ and division $/$, the former is a relation while the later is an operation. For two integers $m \mid n$ takes two possible values "True" or "False" while m/n is a rational number. Of course, $\mid$ and $/$ are related, we have, for $m \neq 0$:

$$m|n \iff \frac{n}{m} \in \mathbb{Z}.$$

The following properties of $\mid$ can be easily proved (Do it!):

(a) $m \mid m$.
(b) $m \mid n$ and $n \mid m \implies m = \pm n$.
(c) $m \mid n$ and $n \mid k \implies m \mid kn$.
(d) For all $n \in \mathbb{Z}$ we have $0 \mid n$,
(e) on the other hand, $n \mid 0 \implies n = 0$.
(f) $m \mid n \implies m \mid -n$, and
(g) $m \mid n \implies -m \mid n$.
(h) $1 \mid n$ for all $n \in \mathbb{Z}$.
(i) $n \mid 1$ for all $n = \pm 1$.
(j) $m \mid n$ and $m \mid k \implies m \mid nk$.
(k) $m \mid n$ and $m \mid k \implies m \mid n + k$.

We have the following fundamental results. The second item in the theorem below is often referred to, as *The Fundamental Theorem of Arithmetic.*

THEOREM B.1.1. *We have:*

(a) There are infinitely many primes.
(b) Every positive integer is a prime or the product of primes, in essentially one way. That is if

$$P = \{p_1, p_2, \ldots, \}$$

is the set of all primes written in an increasing order, then for every positive integer n, there is a unique sequence of exponents k_i, all but finitely many equal to 0, such that

$$n = \prod_{i=1}^{\infty} p_i^{k_i}.$$

Notice that even though there are infinitely many factors in the product, all but finitely many are equal to 1 *and so the product makes sense.*

PROPOSITION 13 (**Euclidean division**). *If* $m, n \in \mathbb{Z}$ *there are unique integers* q, r *such that*

$$n = mq + r, \quad 0 \le r < |m|.$$

We call q *the* quotient *and* r *the* remainder *of the division of* n *by* m.

PROPOSITION 14 (GCD and LCM). *Given* $m, n \in \mathbb{Z}$

(a) There is a positive common divisor of m, n *that is divided by any other common divisor. We call that divisor the* Greatest Common Divisor *and we denote it by* $\gcd(m,n)$*. So the GCD is characterized by the following two properties*
  (a) $\gcd(m,n) \mid m$*, and* $\gcd(m,n) \mid n$.
  (b) If $d \mid m$ *and* $d \mid n$*, then* $d \mid \gcd(m,n)$.

(b) There is a positive common multiple of m, n *that divides any other common multiple. We call that common multiple the* Least Common Multiple *and we denote it by* $\operatorname{lcm}(m,n)$*. So the LCM is characterized by the following two properties*
  (a) $m \mid \operatorname{lcm}(m,n)$*, and* $n \mid \operatorname{lcm}(m,n)$.
  (b) If $m \mid k$ *and* $n \mid k$*, then* $\operatorname{lcm}(m,n) \mid k$.

(c) The greatest common divisor can be written as a linear combination of m, n *with integer coefficients. That is, there are* $a, b \in \mathbb{Z}$ *such that*

$$\gcd(m,n) = am + bn.$$

Furthermore the GCD of m, n *is the only positive common divisor that can be written as a linear combination of* m, n *with integer coefficients.*

DEFINITION 49. Two integers m, n are called *relatively prime* if $\gcd(m,n) = 1$, that is if there is no non-trivial common divisor.

The following is used often.

PROPOSITION 15. *We have:*

(a) If m, n *are relatively prime then*

$$m \mid n \cdot k \implies m \mid k.$$

In particular if p *is a prime number then*

$$p \mid m \cdot n \implies p \mid m \text{ or } p \mid n.$$

(b) If we divide two integers by their GCD we obtain two relatively prime integers. That is, if

$$\gcd\left(\frac{m}{\gcd(m,n)}, \frac{n}{\gcd(m,n)}\right) = 1.$$

(c) If two integers are relatively prime, then so are any of their powers. That is, if $m, n \in \mathbb{Z}$ *and* $k, \ell \in \mathbb{N}$ *then*

$$\gcd(m,n) = 1 \implies \gcd\left(m^k, n^\ell\right) = 1.$$

Recall that a *rational number* is a number that is obtained as a quotient of two integers, that is a number that can be written as $q = m/n$, with $m, n \in \mathbb{Z}$. Of course, such representation of q as a fraction is by no means unique, for example we could also write $q = \frac{2m}{2n}$.

DEFINITION 50. The fraction $\frac{m}{n}$ is called *reduced* if m and n are relatively prime. Every $q \in \mathbb{Q}$ has a unique (up to signs) expression as a reduced fraction.

As an application of the above we prove the following theorem.

THEOREM B.1.2 (**Rational Root Theorem**). *Let $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ be a polynomial with integer coefficients and r a rational root of $p(x)$. If $r = k/\ell$ is a reduced fraction representing q, then k divides a_0 and ℓ divides a_n. In particular, if $p(x)$ is monic, that is $a_n = 1$, then a rational root of $p(x)$ is actually an integer that divides the constant term a_0.*

PROOF. We have $p(r) = 0$, and so

$$a_0 + a_1 \frac{k}{\ell} + \cdots + a_n \frac{k^n}{\ell^n} = 0 \iff -a_0 \ell^n = a_1 k \ell^{n-1} + \cdots + a_n k^n$$

Now k divides the RHS of the above and therefore

$$k \mid -a_0 \ell^n. \tag{B.1}$$

We assumed that k/ℓ is a reduced fraction, and thus k and ℓ are relatively prime, and thus k, and ℓ^n are relatively prime as well. By (B.1) and Proposition 15(1) it follows that $k \mid a_0$.

Similarly by writing $-a_n k = a_0 l^n + k l^{-1} + \cdots + k^{n-1} l$ we conclude that l divides $-a_n k$ and thus a_n. □

COROLLARY 5 (**Irrationality of roots**). *Let $m \in \mathbb{Z}$ and $n \geq 2$ a positive integer. Then the equation*

$$x^n = m$$

has rational solutions if and only if $m = k^n$ for some integer k. In other words, is irrational, unless m is the n-th power of an integer.

PROOF. By the Rational Root Theorem, a rational root of $x^n - m = 0$ is an integer k that divides $-m$. Now by definition, $\sqrt[n]{m}$ is a solution of that equation and thus if $\sqrt[n]{m} \in \mathbb{Q}$ then $\sqrt[n]{m} = k \in \mathbb{Z}$, or equivalently $m = k^n$. □

Thus all of the following real numbers are irrational:

$$\sqrt{2}, \sqrt{3}, \sqrt{6}, \sqrt{110}, \sqrt[3]{42}, \sqrt[7]{56}.$$

In particular all roots of prime numbers are irrational[1].

B.2. Modular arithmetic

DEFINITION 51 (**Congruence modulo m**). Let m be a positive integer, and $a, b \in \mathbb{Z}$. We say that *a is congruent to b modulo m*, and write $a \equiv b \pmod{m}$, if $m \mid a - b$.

The following characterization of congruence modulo m is often very useful.

THEOREM B.2.1. *Let m be a positive integer. Then for all $a, b \in \mathbb{Z}$ we have $a \equiv b \pmod{m}$ if and only if a and b leave the same remainder when divided by m.*

PROOF. Let $a = m q_1 + r_1$, and $b = m q_2 + r_2$ be the quotients and remainders of the Euclidean division (see Proposition 13). Then

$$a - b = m(q_1 - q_2) + (r_1 - r_2). \tag{B.2}$$

Now, if $r_1 = r_2$ then Equation B.2 becomes $a - b = m(q_1 - q_2)$ and so $m \mid a - b$.

Conversely, since $|r_1 - r_2| < |m|$, Equation B.2 says that $a - b$ leaves remainder $|r_1 - r_2|$ when divided by m. Thus if $m \mid a - b$ then $|r_1 - r_2| = 0$, i.e. $r_1 = r_2$. □

Using Theorem B.2.1 we can prove the following theorem.

THEOREM B.2.2 (**Congruence modulo m is an equivalence relation)**. *The following hold for any modulus $m \geq 1$.*

[1] Why? Prove this.

(a) For all $a \in \mathbb{Z}$*,*

$$a \equiv a \pmod{m}.$$

(b) For all $a, b \in \mathbb{Z}$*,*

$$a \equiv b \pmod{m} \implies b \equiv a \pmod{m}.$$

(c) For all $a, b, c \in \mathbb{Z}$*,*

$$a \equiv b \pmod{m} \text{ and } b \equiv c \pmod{m} \implies a \equiv c \pmod{m}.$$

PROOF. Exercise. □

Now, notice that each $a \in \mathbb{Z}$ is congruent $\pmod{m}$ to exactly one of the possible remainders $\{1, \dots, m-1\}$.

DEFINITION 52 (**Congruence classes modulo** m). For a positive integer m, and $k \in \{0, 1, \dots, m-1\}$ we define the *congruence class of* $k \pmod{m}$, denoted by $[k]_m$, to be the set of integers that are congruent to $k \pmod{()m}$, i.e. leave reminder k when divided by m. Thus,

$$[m]_k = \{a \in \mathbb{Z} : a \equiv k \pmod{m}\},$$

or equivalently,

$$[m]_k = \{m\,\ell + k : \ell \in \mathbb{Z}\}.$$

The set of all congruence classes $\pmod{m}$ is denoted $\mathbb{Z}/m$. Thus

$$\mathbb{Z}/m = \{[k]_m : k = 0, \dots, m-1\}.$$

Elements of $\mathbb{Z}/m$ are called *integers modulo* m.

More generally, if $a \in \mathbb{Z}$ we write $[a]_m$ to stand for the set of all integers that are congruent to $a \pmod{m}$. Thus,

$$[a]_m = \{b \in \mathbb{Z} : a \equiv b \pmod{m}\}.$$

So,

$$[a]_m = [b]_m \iff a \equiv b \pmod{m}.$$

For $m = 2$, there are two possible remainders, namely $\{0, 1\}$, and so we have two congruence classes $[0]_2$ and $[1]_2$ and

$$\mathbb{Z} = [0]_2 \cup [1]_2.$$

Notice that $[0]_2$ is the set of even integers and $[1]_2$ is the set of odd integers. We have then that, for example, $[-4]_2 = [42]_2 = [0]_2$ and $[11]_2 = [-59]_2 = [1]_2$.

NOTE. To simplify notation, we often identify $[k]_m$ with k and so we write

$$\mathbb{Z}/m = \{0, \dots, m\}.$$

THEOREM B.2.3 (**Addition and multiplication respect congruences**). *Let* $m \geq 1$*, and* a, b, c, d *integers such that* $a \equiv c \pmod{m}$ *and* $b \equiv d \pmod{m}$*. Then,*

$$a + b \equiv c + d \pmod{m} \quad \text{and} \quad a\,b \equiv c\,d \pmod{m}.$$

PROOF. We have

$$(a + b) - (c + d) = (a - c) + (b - d)$$

and so $m \mid (a + b) - (c + d)$.

And,

$$a\,b - c\,d = a\,b - a\,d + a\,d - c\,d = a\,(b - d) + (a - c)\,d.$$

Therefore $m \mid a\,b - c\,d$. □

Theorem B.2.3 makes the following definition possible.

DEFINITION 53 (**Modular addition and multiplication**). In $\mathbb{Z}/m$ we define operations of addition and multiplication as follows:

$$[a]_m + [b]_m = [a + b]_m$$

and

$$[a]_m [b]_m = [a b]_m.$$

In words, the sum (respectively product) of the congruence classes of two integers is the congruence class of their sum (respectively product).

Theorem B.2.3 ensures that no matter what elements of the congruence classes we chose we will always get the same answer.

EXAMPLE 120. Consider $\mathbb{Z}/12$. We have

$$[7]_{12} + [8]_{12} = [15]_{12} = [3]_{12}$$

while,

$$[5]_{12} \times [7]_{12} = [35]_{12} = [11]_{12}.$$

Now, $7 \equiv -5 \pmod{12}$ and $8 \equiv 32 \pmod{12}$ and so $[7]_{12} = [-5]_{12}$ and $[8]_{12} = [32]_{12}$. Of course,

$$[-5]_{12} + [32]_{12} = [27]_{12} = [3]_{12}$$

as we expect by Theorem B.2.3.

Similarly, $[5]_{12} = [41]_{12}$ and $[7]_{12} = [19]_{12}$ and

$$[41]_{12} [19]_{12} = [779]_{12} = [11]_{12}$$

.

Before we continue we give the following definition.

DEFINITION 54. A set R with two operations + and · that satisfy all the axioms of a field except (possibly) (5) and (8) is called a *(unital) ring*[2].

If axiom (5) is also satisfied then we say that R is a *commutative* ring.

EXAMPLE 121. The set of integers is a commutative ring $\mathbb{Z}$. The set of square matrices with real entries is a non-commutative ring.

We then have the following theorem.

THEOREM B.2.4. *We have:*

(a) *$\mathbb{Z}/m$ endowed with the operations of addition and multiplication as given by Definition 53 is a commutative ring. The zero element is the class $[0]_m$ and $-[k]_m = [m = k]_m$.*
(b) *If p is a prime then $\mathbb{Z}/p$ is a field.*
(c) *If m is not prime then $\mathbb{Z}/m$ is not a field.*

PROOF. (a) This follows from the fact that $\mathbb{Z}$ is a commutative ring. As an example I prove the distributive property, and leave the other properties as an exercise.

[2] Sometimes people consider rings that don't have 1, that is axiom (7) does not hold. When we want to emphasize that the rings we consider have 1 we use the term *unital*.

We have, omitting the subscript m,

$$\begin{aligned} [a]\,([b]+[c]) &= [a]\,[b+c] \\ &= [a\,(b+c)] \\ &= [a\,b + a\,c] \\ &= [a\,b] + [a\,c] \\ &= [a]\,[b] + [a]\,[c]. \end{aligned}$$

(b) Let $k \in \{1, \dots, p-1\}$ and consider the $p-1$ products

$$[1]_p\,[k]_p, \dots, [p-1]_p\,[k]_p.$$

If two of these products are equal, say

$$[a]_p\,[k]_p = [b]_p\,[k]_p$$

then

$$([a]_p - [b]_p)\;[k]_p = [0]_p$$

and so

$$p \mid (a-b)\,k$$

and since $k < p$, and $k \neq 0$, we have that p does not divide k. Therefore $p \mid a-b$ and it follows that $[a]_p = [b]_p$.

It follows that the products $[a]_p\,[k]_p$, with $a = 1, \dots, p-1$ are all different. Since there are $p-1$ possible values for this products, namely $[1]_p, \dots, [p-1]_p$ we conclude that one of these products is equal to $[1]_p$. Thus for for some $[a]_p \in \mathbb{Z}/p \smallsetminus \{[0]_p\}$ we have

$$[a]_p\,[k]_p = [1]_p.$$

Thus $[k]_p$ has an inverse.

(c) If m is composite (i.e. not prime) then $m = a\,b$ with $1 < a, b < m$. But then

$$[a]_m\,[b]_m = [m]_m = [0]_m$$

while $[a]_m \neq [0]_m$ and $[b]_b m \neq [0]_m$. But this cannot happen in a field, see Item (g) of Theorem 4.1.1.

□

APPENDIX C

Polynomials

C.1. The algebra of polynomials

We are mostly interested in polynomials with coefficients in a field, but sometimes we may consider more general coefficients. For example we may consider polynomials with coefficients in $\mathbb{Z}$, and (see Example 121 of Appendix B), $\mathbb{Z}$ is only a *ring*.

If R is a *ring* then $R[x]$ denotes the set of all polynomials with coefficients in R and indeterminate x. A *polynomial* with coefficients in R is an algebraic expression of the form

$$p(x) = a_0 + a_1\,x + \cdots + a_d\,x^d, \tag{C.1}$$

where x is an indeterminate, i.e. a variable, and $a_0, \ldots, a_d \in R$ are the coefficients.

The set of all polynomials of one variable x with coefficients in R is denoted by $R[x]$. It is sometimes convenient to write polynomials as a sum of infinitely many terms, with only finitely many of them non-zero. In other words we think of a polynomial as having infinitely many coefficients, one for each power x^n, but after a certain power of x all coefficients are 0. Thus we may write any of the following expressions

$$p(x) = \sum_{k\in\mathbb{N}} a_k\,x^k = a_0 + a_1\,x + \cdots + a_n\,x^n + \cdots = \sum a_k\,x^k$$

and we assume that $a_n = 0$ for all but finitely many $n \in \mathbb{N}$.

Two polynomials are considered equal if they have the same coefficients. That is,

$$\sum_{k\in\mathbb{N}} a_k\,x^k = \sum_{k\in\mathbb{N}} b_k\,x^k \iff \forall k \in \mathbb{N},\ a_k = b_k.$$

We emphasize that, at the outset, the operations in Equation (C.1) are *formal*, $a_n\,x^n$ is not really[1] a product, it's only an *symbolic expression*. Similarly $\sum a_n\,x^n$ is not really a sum. Of course eventually, after we have introduced addition and multiplication, we will be able to interpret the formal sums and products as *actual* sums and products.

The *zero polynomial*, denoted simply by 0, is the polynomial with all coefficients equal to $0 \in R$. That is,

$$0 = \sum_{n\in\mathbb{N}} 0\,x^n.$$

More generally, for $a \in R$ we define the *constant polynomial with value* a, to be the polynomial with the 0-th coefficient equal to a and all other coefficients equal to 0. We simply write a for the constant polynomial with value a. Thus

$$a = \sum_{n\in\mathbb{N}} \delta_{0n} a\,x^n.$$

The *leading term* of $p(x)$ is defined to be the largest non-zero term, provided that such a term exists (that is provided that $p(x) \neq 0$). If $a_d\,x^d$ is the leading term, then we say that a_d

[1] Not yet, at least.

is the *leading coefficient* of $p(x)$ and that $p(x)$ has *degree* d. We denote the degree of $p(x)$ by $\deg p(x)$. Thus, for $p(x) = \sum a_n x^n$ we have

$$\deg p(x) = d \iff a_d \neq 0, \text{ and } a_k = 0 \text{ for } k > d.$$

Notice that since the zero polynomial has no non-zero terms its degree is not defined. By convention we define $\deg 0 = -\infty$, so that the degree of the zero polynomial is less than the degree of all other polynomials.

We employ the usual shorthands when writing a polynomial as the sum of finitely many terms: we write x^n instead of $1\,x^n$, and $a_m x^m - a_n x^n$ instead of $a_m x^m + (-a_n)\,x^n$, etc. Thus for example,

$$3\,x^2 - 5\,x^3 + x^5$$

is the polynomial $p(x) = \sum a_n x^n$ with

$$a_n = \begin{cases} 3 & n = 2 \\ -5 & n = 3 \\ 1 & n = 5 \\ 0 & \text{otherwise} \end{cases}.$$

Of course, a polynomial's purpose in life is to be *evaluated* for various values of the variable x. In elementary algebra variables such a x stand for unknown, or indeterminate numbers[2] and then a polynomial stands for the result of some algebraic operations applied to that unknown number. Thus if x stands for a unknown number, $x^2 - 3\,x$ stands for the difference of the square of that number and three times that number. Once x is known or fixed we can find what number $x^2 - 3\,x$ stands for by *evaluating*, i.e. *substituting* that value for x.

DEFINITION 55 (**Evaluating polynomials**). Let $p(x) = \sum a_n x^n \in R[x]$ and $a \in R$. Then the evaluation of $p(x)$ at a, denoted by $p(a)$, is defined to be

$$p(a) = \sum a_n a^n,$$

where, as usual we consider the sum of infinitely many zeros to be zero. The notation

$$p(x)|_{x=a}$$

is also used occasionally for $p(a)$.

If $p(a) = 0$ then we say that a is a *root* or *zero* of $p(x)$.

Thus $p(x)$, via evaluation, defines a function

$$p \colon R \to R, \quad a \mapsto p(a).$$

We say that p is the *polynomial function defined by* $p(x)$.

We now want to define addition and multiplication of polynomials in a way that respects evaluation. That means we want,

$$(p(x) + q(x))\,|_{x=a} = p(a) + q(a) \tag{C.2}$$

$$(p(x)\,q(x))\,|_{x=a} = p(a)\,q(a), \tag{C.3}$$

for all $a \in R$.

[2] The terms "unknown" and "indeterminate" or "arbitrary" have different connotations. If x is an unknown number then x is a certain number, we just don't know which number it is. On the other hand if x is an indeterminate then x can vary, it could be any number. In practice though the distinction is not that important because in both cases we manipulate x and expressions involving it, using only properties that hold for all numbers.

Let then $p(x) = \sum a_n\, x^n$ and $q(x) = \sum b_n\, x^n$ be two polynomials, and let $a \in R$ be arbitrary. Then, using the properties of addition and multiplication in the commutative ring R we have,

$$p(a) + q(a) = \sum a_n\, a^n + sumb_n\, x^n = \sum \left(a_n\, a^n + b_n\, a^n\right) = \sum \left(a_n + b_n\right)\, a^n.$$

Thus Equation (C.2) is satisfied if we define

(C.4) $$p(x) + q(x) = \sum \left(a_n + b_n\right)\, x^n.$$

It is straightforward to verify then that addition is commutative and associative, that the zero polynomial is neutral for addition and that $p(x) + (-p(x)) = 0$ where

$$-p(x) = \sum(-a_n)\, x^n.$$

To define a multiplication that satisfies Equation (C.3) we start by declaring that $c\, x^m$, viewed as a polynomial, is actually the product of the constant polynomial c, and the polynomial x^m. That is

$$\left(\sum \left(\delta_{0n}\, c\right)\, x^n\right)\, \left(\sum \delta_{mn}\, x^n\right) = \sum \left(\delta_{mn}\, c\right)\, x^n.$$

Similarly, since $a^m\, a^n = a^{m+n}$ we declare that

$$x^m\, x^n = x^{m+n}.$$

REMARK 22 (**Formal sums are actual sums**). Notice that with our definitions so far mean that a polynomial is actually the sum of its terms[3], and each term $a_n\, x^n$ is the actual product of its coefficient, the constant polynomial a_n, and the polynomial x^n.

Now given that we want multiplication to distribute over addition we define

(C.5) $$p(x)\, q(x) = \sum_{n\in\mathbb{N}} \left(\sum_{\substack{0\le k,\ell\le n \\ k+\ell=n}} \right) x^n.$$

That is, the coefficient of x^n in the product of two polynomials, is the sum of the products of all coefficients of terms with degrees that add up to n.

THEOREM C.1.1 (**Polynomials form a ring**). *$R[x]$ endowed with addition as defined by (C.4), and multiplication defined by (C.5) is a commutative ring. The role of zero is played by the zero polynomial and the role of one is played by the constant polynomial 1.*

If we identify $a \in R$ with the constant polynomial $a \in R[x]$ then $R \subseteq R[x]$ is a subring of $R[x]$.

PROOF. Exercise. □

Recall from Definition 25 that we can evaluate a polynomial $p(x) \in \mathbb{R}[x]$ at a square matrix $A \in \mathbf{M}_n(\mathbb{R})$ to obtain $p(A) \in \mathbb{M}_n(\mathbb{R})$. Clearly this definition make sense for any field K[4]. More generally, if $\mathbf{A}$ is an algebra over a field K, and $a \in \mathbf{A}$ we can define $p(a) \in \mathbf{A}$.

Now $K[x]$ is an algebra over K and therefore we can evaluate $p(x)$ at any polynomial $q(x)$ to obtain a polynomial $p\left(q(x)\right) \in K[x]$. In particular, if we evaluate $p(x)$ at the polynomial x we obtain the polynomial $p(x)$.

[3] With the usual caveat regarding sums of infinitely many zero terms

[4] Or any ring really, but our main interest is polynomials over a field.

C.1.1. Polynomials vs Polynomial functions. Often polynomials are defined as functions. For example a polynomial with real coefficients is defined as a function

$$P\colon \mathbb{R} \to \mathbb{R}, \quad P(x) = \sum_{k=0}^{n} a_k\, x^k$$

with $a_k \in \mathbb{R}$. This is fine when we work in a field like $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$ because in such fields[5] two polynomials are equal if and only if the functions they define are equal.

THEOREM C.1.2. *Let $P, Q\colon \mathbb{R} \to \mathbb{R}$ be defined via*

$$P(x) = \sum_{k=0}^{n} a_k\, x^k, \quad Q(x) = \sum_{k=0}^{n} b_k\, x^k,$$

where $a_k, b_k \in \mathbb{R}$. Then $P = Q$ if and only if, $a_k = b_k$ for all $k = 0, \ldots n$.

The proof follows from the following lemma from Calculus.

LEMMA 4. *The function*

$$P\colon \mathbb{R} \to \mathbb{R}, \quad P(x) = \sum_{k=0}^{n} a_k\, x^k$$

has derivatives of all orders and for $k = 1, \ldots, n$

$$a_n = \frac{P^{(n)}(0)}{n!}.$$

PROOF. For $n = 0$ the lemma says that $a_0 = P(0)$ which is true. Now we have

$$p'(x) = n\, a_n\, x^{n-1} + \cdots + 3\, a_3\, x^2 + 2\, a_2\, x + a_1.$$

Evaluating at $x = 0$ then gives

$$P'(0) = a_1.$$

Differentiating again gives

$$P''(x) = n\,(n-1)\, x^{n-2} + \cdots + 6\, a_3 x + 2\, a_2.$$

Evaluating at $x = 0$ then gives

$$P''(0) = 2\, a_2 \implies a_2 = \frac{f''(0)}{2}.$$

Continuing in the same manner we get

$$P^{(3)}(x) = n\,(n-1)\,(n-2)\, x^{n-2} + \cdots + 24\, a_4\, x + 6\, a_3,$$

and so

$$a_3 = \frac{P^{(3)}(0)}{6}.$$

In general we can prove by induction that for $0 \le k \le n$ we have

$$P^{(k)}(x) = n\,(n-1) \cdots (n-k+1)\, x^{n-k} + \cdots + k!\, a_k,$$

and the lemma follows by evaluating at $x = 0$. □

PROOF OF THEOREM C.1.2. Clearly, if for all $a_k = b_k$ for all k then $P = Q$. Conversely, if $P = Q$ then all the derivatives of P and Q are equal and therefore, by Lemma 4, for all k we have $a_k = b_k$. □

[5] These are fields of *characteristic* 0. Explaining what this means would take us far afield.

By considering real and imaginary parts of a polynomial we can see that Theorem C.1.2 holds for complex polynomials as well[6]. But in other fields we may have different polynomials (in the sense that they have different coefficients) that define the same function. For example in $\mathbb{Z}/5$ the polynomial $x^5 - x$ defines the same function as the zero polynomial, see Claim 3. Here is another example.

EXAMPLE 122. The polynomials $p(x) = x^2 - x$ and 0 both induce the same function $\mathbb{Z}/2 \to \mathbb{Z}/2$, namely the zero function.

However, $p(x)$ does not induce the zero function $\mathbf{M}_2(\mathbb{Z}/2) \to \mathbf{M}_2(\mathbb{Z}/2)$. Indeed for

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

we have

$$A^2 = I$$

and therefore

$$p(A) = A^2 - A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

C.1.2. Euclidean Division, Factors, roots. In what follows, unless we explicitly say otherwise, we assume that K is a field and that polynomials have coefficients in K.

In general, division of polynomials

$$p(x) \div d(x)$$

is not defined, in the sense that there is no polynomial $q(x)$ such that $p(x) = d(x)\,q(x)$. However there is an analogue of Euclidean division (see Proposition 13) for polynomials.

EXAMPLE 123. Let's try to divide $p(x) = 3\,x^4 - 14\,x^3 + 9\,x^2 - 11\,x + 70$ by $d(x) = x - 4$, i.e. to perform the division

$$\frac{3\,x^4 - 14\,x^3 + 9\,x^2 - 11\,x + 70}{x - 4}.$$

Thus we want to find a polynomial $q(x)$ such that

$$p(x) = d(x)\,q(x).$$

We won't succeed because no such polynomial exists, but we'll try our best. The leading term of $p(x)$ $(3\,x^4)$ has to be the leading term of $d(x)$ (x) times the leading term of $q(x)$. Thus the leading term of $q(x)$ has to be $3\,x^3$. So we have

$$p(x) = d(x)\,(3\,x^3 + q_1(x))$$

where $q_1(x)$ is to be determined.

Now

$$p(x) = d(x)\,(3\,x^3 + q_1(x)) \iff p(x) - 3\,x^3\,d(x) = d(x)\,q_1(x),$$

or, setting $p_1(x) = p(x) - 3\,x^3\,d(x)$

$$p_1(x) = d(x)\,q_1(x).$$

Thus $q_1(x)$ is the quotient of a division with the same denominator but as we'll see shortly numerator of smaller degree. Indeed we calculate

[6] Can you prove this?

$$\begin{aligned} p_1(x) &= p(x) - 3\,x^3\,d(x) \\ &= 3\,x^4 - 14\,x^3 + 9\,x^2 - 11\,x + 70 - 3\,x^3\,(x-4) \\ &= 3\,x^4 - 14\,x^3 + 9\,x^2 - 11\,x + 70 - 3\,x^4 + 12\,x^3 \\ &= -2\,x^3 + 9\,x^2 - 11\,x + 70. \end{aligned}$$

Thus, as before we have that $-2\,x^3$, the leading term of $p_1(x)$, has to be the product of x, the leading term of $d(x)$, and the leading term of $q_1(x)$. Thus the leading term of $q_1(x)$ is $-2\,x^2$. So at the second step we have

$$p(x) = d(x)\,(3\,x^3 - 2\,x^2 + q_2(x))$$

where $q_2(x)$ is to be determined. As before, we have that q_2 is the quotient of the division

$$p_2(x) = d(x)\,q_2(x),$$

where

$$p_2(x) = p_1(x) + 2\,x^2\,d(x) = x^2 - 11\,x + 70.$$

So the leading term of $q_2(x)$ is x and we have

$$p(x) = d(x)\,(3\,x^3 - 2\,x^2 + x + q_3(x))$$

with $q_3(x)$ is a constant to be determined. Now,

$$p_2(x) - (x-4)\,x = -7\,x + 70$$

forcing $q_3(x) = -7$. So $q(x)$ has been determined:

$$q(x) = 3\,x^3 - 2\,x^2 + x - 7.$$

But this doesn't quite work, because

$$d(x)\,q(x) = (x-4)\,(3\,x^3 - 2\,x^2 + x - 7) = 3\,x^4 - 14\,x^3 + 9\,x^2 - 11\,x + 28 \neq p(x).$$

Still we got the first four terms of $p(x)$ right, and that's the best we can do. For, as the above calculations demonstrate, if the first four coefficients are correct then the constant term has to be 28.

$$3\,x^4 - 14\,x^3 + 9\,x^2 - 11\,x + 70 = (x-4)\,(3\,x^3 - 2\,x^2 + x - 7) + 42.$$

Euclidean division algorithm for polynomials

The input is two polynomials $p(x), d(x) \in K[x]$ with $d(x) \neq 0$ and $\deg d(x) < \deg p(x)$. Let $b_m\,x^m$ be the leading term of $d(x)$. Set

$$q(x) := 0.$$

```
**While** $\deg d(x) \leq \deg p(x)$:
    (a) Set $a_n\,x^n$ to be the leading term of $p(x)$.
    (b) Set $c(x) := \frac{a_n}{b_m}\,x^{n-m}$.
    (c) Set $q(x) := q(x) + c(x)$.
    (d) Set $p(x) := p(x) - c(x)\,d(x)$.
**Return** $q(x)$ as the quotient and $p(x)$ as the remainder.
```

EXAMPLE 124. Let's also divide two polynomials in $\mathbb{Z}/11$. Perform the division

$$(x^4 + 5\,x^3 + 8\,x^2 + 5) \div (5\,x^2 + 10\,x + 2),$$

in $\mathbb{F}_{11}[x]$.

We have

$$\begin{array}{rl}
x^4 + 5\,x^3 + 8\,x^2 + 5 & = (5\,x^2 + 10\,x + 2)\,(9\,x^2 + 5\,x + 10) + 7 \\
-x^4 - 2\,x^3 - 7\,x^2 & \\
3\,x^3 + x^2 + 5 & \\
-3\,x^3 - 6\,x^2 - 10\,x & \\
6\,x^2 + x + 5 & \\
-6\,x^2 - x - 9 & \\
7 &
\end{array}$$

In the first step of the algorithm we divided the leading term of $p(x)$ by the leading term of $d(x)$

$$\frac{x^4}{5\,x^2} = \frac{1}{5}\,x^2 = 9\,x^2$$

because in $\mathbb{Z}/11$ we have $5^{-1} = 9$. Indeed in $\mathbb{Z}$ we have $5 \cdot 9 = 45$ and 45 leaves remainder 1 when divided by 11 since $45 = 11 \cdot 4 + 1$.

Then we multiplied $9\,x^2$ with $d(x)$ and subtracted it from $p(x)$. So we first calculated

$$-(9\,x^2\,d(x)) = -\big(9\,x^2\,(5\,x^2 + 10\,x + 2)\big) = -(x^4 + 2\,x^3 + 7\,x^2) = -x^4 - 2\,x^3 - 7\,x^2$$

and so

$$p_1(x) = p(x) - 9\,x^2\,d(x) = 3\,x^3 + x^2 + 5.$$

Since still $d(x)$ has lower degree than $p_1(x)$ we have $q(x) = 9\,x^2$ and repeat the procedure, with $p(x)$ replaced by $p_1(x)$. The leading term of $p_1(x)$ is $3\,x^3$ and

$$\frac{3\,x^3}{5\,x^2} = \frac{3}{5}\,x = 5\,x$$

because in $\mathbb{Z}/11$ we have $5^{-1} = 9$ and $3 \cdot 9 = 5$. Indeed in $\mathbb{Z}$ we have $3 \cdot 9 = 27$ and $27 = 2 \cdot 11 + 5$. Then

$$-\,(5\,x\,d(x)) = -\big(3\,x^3 + 6\,x^2 + 10\,x\big) = -3\,x^3 - 6\,x^2 - 10\,x$$

and so

$$p_2(x) = 6\,x^2 + 6\,x + 5$$

and the quotient is updated to

$$q(x) = 9\,x^2 + 5\,x.$$

We still have $\deg d(x) \leq \deg p(x)$ and so we continue.

$$\frac{6\,x^2}{5\,x^2} = \frac{6}{5} = 10.$$

Then

$$-\,(10\,d(x)) = -6\,x^2 - x - 9$$

and so

$$p_3 = 5 - 9 = 7$$

and the quotient is updated to

$$q(x) = 9\,x^2 + 5\,x + 10.$$

Since $\deg p_3(x) < \deg d(x)$ we stop and return $q(x)$ as the quotient and $p_3(x) = 7$ as the remainder.

We have then the following theorem.

THEOREM C.1.3 (**Long division**). *Let $p(x) \in K[x]$ and $d(x) \in K[x]$ with $d(x) \neq 0$ and $\deg d(x) < \deg p(x)$. Then there exist unique polynomials $q(x), r(x) \in K[x]$ with $\deg r(x) < \deg d(x)$ such that*

$$p(x) = d(x)\,q(x) + r(x). \tag{C.6}$$

PROOF. The existence of the quotient and the remainder are guarantied by the algorithm we just described. Division is defined in any field so the steps can be performed at every field. The algorithm eventually terminates because the degree of $p(x)$ decreases strictly at every step and so eventually it will become less than the degree of $d(x)$.

To see that the quotient and remainder are uniquely defined notice that Equation (C.6) implies that the leading term of $q(x)$ is the quotient of leading term of $p(x)$ by the leading term of $d(x)$. Let $q_1(x)$ stand for the lower degree terms of $q(x)$ so that

$$p(x) = d(x)\left(\frac{a_n}{b_m}\,x^{n-m} + q_1(x)\right) + r(x).$$

Therefore, in analogy with the calculations in Example 123 we have

$$p(x) - d(x)\,\frac{a_n}{b_m}\,x^{n-m} = d(x)\,q_1(x) + r(x),$$

and so $q_1(x)$ is the quotient of

$$p_1(x) = p(x) - d(x)\,\frac{a_n}{b_m}\,x^{n-m}$$

by $d(x)$. Thus the leading term of $q_1(x)$ is unique. Inductively, all the terms of $q(x)$ are unique. But then

$$r(x) = p(x) - d(x)\,d(x),$$

and so the remainder is also uniquely determined. □

A particularly interesting case is when we divide by a linear polynomial of the form $x - a$, as in Example 123. We have $\deg(x - a) = 1$ and therefore the remainder has degree 0 or is the zero polynomial, and so $r(x)$ is a constant c. In other words, we have

$$p(x) = (x - a)\,q(x) + c. \tag{C.7}$$

Now substituting $x = a$ in Equation (C.7) we get

$$p(a) = (a - a)\,q(a) + c = 0\,q(a) + c = 0 + c = c.$$

Thus the reminder is the value of $p(x)$ at a.

We have then proved the following theorem.

THEOREM C.1.4 (**The remainder theorem**). *Let $a \in K$ and $p(x) \in K[x]$. Then the remainder of the division $p(x) \div (x - a)$ is $p(a)$.*

REMARK 23. Theorem C.1.4 provides an efficient way to evaluate polynomials. Dividing by $x - a$ can be done much more efficiently than actually evaluating $p(a)$ by actually plugging it in.

An immediate corollary is the following important theorems.

THEOREM C.1.5 (**Root-Factor correspondence**). *Let $a \in K$ and $p(x) \in K[x]$. Then a is a root of $p(x)$ if and only if $x - a$ is a factor of $p(x)$.*

PROOF. $x - a$ is a factor of $p(x)$ if and only if the remainder of the division $p(x) \div (x - a)$ is 0, if and only if $p(a) = 0$. □

COROLLARY 6 (**The number of roots does not exceed the degree**). *Let $p(x) \in K[x]$. Then $p(x)$ has at most $\deg p(x)$ roots.*

PROOF. A product of k factors of the form $x - a$ has leading term x^k. So if $p(x)$ has k roots, it has a factor of degree k, and therefore $k \leq \deg p(x)$. □

DEFINITION 56 (**Multiplicity of roots**). Let $a \in K$ be a root of the polynomial $p(x)$. If $(x - a)^m$ is a factor of $p(x)$ but $(x - a)^{m+1}$ is not, we say that a is a root of *multiplicity* m.

For example the polynomial $p(x) = x^6 + 3\,x^5 - 4\,x^3$ factors like this

$$x^6 + 3\,x^5 - 4\,x^3 = x^3\,(x+3)\,(x-1)^2.$$

Therefore, 0 is a root of multiplicity 3, -3 a root of multiplicity 1, and 1 a root of multiplicity 2. We say then that $p(x)$ has six roots, *counting with multiplicity*.

C.1.3. Roots of real and complex polynomials. We end this appendix by mentioning some important results regarding roots of real and complex polynomials.

THEOREM C.1.6. *If $p(x) \in \mathbb{R}[x]$ has odd degree then $p(x)$ has at least one real root.*

PROOF. $p(x) \to \pm\infty$ as $x \to \pm\infty$ and thus $p(x)$ takes both negative and positive values. The result follows from Intermediate Value Theorem. □

Of course, there are real polynomials of even degree that have no real roots. For example

$$p(x) = x^2 + 1$$

has no real roots, because for all $x \in \mathbb{R}$, we have $x^2 \geq 0$ and thus $x^2 + 1 > 0$.

THEOREM C.1.7 (**Fundamental Theorem of Algebra**). *Every $p(x) \in \mathbb{C}[x]$ has at least one root.*

But, if $p(a) = 0$ then $x - a$ is a factor of $p(x)$ and the quotient $q(x) = p(x)/(x - a)$ is also a complex polynomial. Therefore, if $\deg q(x) > 0$ then $q(x)$ has a root, and this root will also be a root of $p(x)$. Continuing this way we see that $p(x)$ has d roots, where $d = \deg p(x)$.

COROLLARY 7. *Let $p(x) \in \mathbb{C}[x]$ be a degree d polynomial. Then $p(x)$ has exactly d linear factors.*

PROPOSITION 16. *Let $p(x) \in \mathbb{R}[x]$. If $z \in \mathbb{C}$ is a root of p then so is its complex conjugate $\overline{z}$.*

C.2. Useful polynomial identities

- **Binomial Theorem**

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k}.$$

- **Difference of powers**

$$a^n - b^n = (a-b) \sum_{k=0}^{n-1} a^k\, b^{n-k}.$$

- **Sum of odd powers** If n is odd then for all real numbers a, b we have

$$a^n + b^n = (a+b) \sum_{k=0}^{n-1} (-1)^k\, a^k\, b^{n-k}.$$

APPENDIX D

On Bases

Let V be a vector space, how can we find a basis for V? We can proceed as in the proof of Theorem 2.2.6: we build the basis one vector at the time.

If $V = \{0\}$ is a trivial vector space then by convention we have that the empty set is a basis and we are done. In other words we start with

$$B_0 = \varnothing$$

and check if $\langle B_0 \rangle = V$. If this is the case we are done, and $\langle B_0 \rangle = \{0\}$ we know that in that case V is the zero vector space.

If not, then there is a non-zero vector $v_1 \in V$, so we add it to B_0 and get

$$B_1 = B_0 \cup \{v_1\} = \{v_1\} .$$

Notice that B_1 is linearly independent so if $\langle B_1 \rangle = V$ we are again done, B_1 is a basis. If not, then there are vectors in V that are not multiples of v_1, so we choose one of them, say v_2 and we add it to B_1 to obtain

$$B_2 = B_1 \cup \{v_2\} = \{v_1, v_2\} .$$

Notice that B_2 is linearly independent, because

$$\lambda_1 v_1 + \lambda_2 v_2 = 0, \tag{D.1}$$

cannot hold unless $\lambda_1 = \lambda_2 = 0$. To see this note that if $\lambda_2 \neq 0$ then the Equation (D.1) gives

$$v_2 = -\frac{\lambda_1}{\lambda_2} v_1,$$

a contradiction because $v_2 \notin \langle v_2 \rangle$.

Thus $\lambda_2 = 0$ and then Equation (D.1) becomes

$$\lambda_1 v_1 = 0$$

and since $v_1 \neq 0$ we conclude that $\lambda_1 = 0$.

We now check whether $V = \langle B_2 \rangle$. If yes then B_2 is a basis and we are done. If not, we choose $v_3 \notin \langle B_2 \rangle$ and we add it to B_2 to obtain

$$B_3 = B_2 \cup \{v_3\} = \{v_1, v_2, v_3\} .$$

Again it's easy to see that B_3 is linearly independent, for, if

$$\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = 0, \tag{D.2}$$

we cannot have $\lambda_3 \neq 0$ otherwise

$$\lambda_1 v_1 + \lambda_2 v_2 = 0,$$

impossible since $v_3 \notin \langle B_2 \rangle$. Thus $\lambda_3 = 0$ and Equation (D.2) together with the fact that B_2 is linearly independent gives that $\lambda_1 = \lambda_2 = 0$.

And we continue as before: if $V = \langle B_3 \rangle$ we are done, ...

Thus we have a procedure that for every natural number n, either gives a basis B_n or produces a linear independent set $B_{n+1} \supseteq B_n$ by choosing $v_{n+1} \notin \langle B_n \rangle$ and defining

$$B_{n+1} = B_n \cup \{v_{n+1}\} = \{v_1, v_2, \ldots, v_n, v_{n+1}\}.$$

In the case where V is a subspace of $\mathbb{R}^n$, this procedure will stop at some step $k \leq n$ because we can't chose $n+1$ linearly independent vectors. However there is no guarantee that it will for a general vector space. For example if $V = K[x]$, the space of polynomials with coefficients in K, we could at every stage choose $v_k = x^k$ and then we get an infinite sequence of linearly independent sets

$$B_n = \left\{x^k : k = 0, 1, \ldots, n\right\}$$

without ever stopping.

But we don't have to stop either. We can take the union of all the B_n and call it B_ω

$$B_\omega = \bigcup_{n \in \mathbf{N}} B_n$$

and notice that B_ω is linearly independent. To see this notice that, the sequence B_n is *an increasing sequence of sets*, i.e.

$$k < \ell \implies B_k \subsetneq B_\ell,$$

because, by the way we constructed these sets we have

$$B_0 \subsetneq B_1 \subsetneq B_2 \subsetneq \cdots \subsetneq B_n \subsetneq B_{n+1} \subsetneq \cdots. \tag{D.3}$$

Now consider a linear dependency in B_ω

$$\lambda_1 v_{n_1} + \lambda_2 v_{n_2} + \cdots + \lambda_{n_k} v_{n_k} = 0$$

where $n_1 < n_2 < \cdots < n_k$. Then by Equation (D.3) we have that

$$v_{n_1}, \ldots, v_{n_k} \in B_{n_k}$$

and we have a linear dependency relation in B_{n_k} a contradiction.

So we can ask whether $\langle B_\omega \rangle = V$. If so, as it happens in the case of polynomials, we conclude that B_ω is a basis and we are done.

If not we start again, we choose $v_{\omega+1} \notin \langle B_\omega \rangle$ and set

$$B_{\omega+1} = B_\omega \cup \{v_{\omega+1}\} = \{v_1, v_2, \ldots, v_n, \ldots, v_{\omega+1}\}.$$

If $B_{\omega+1}$ is a basis we are done. Otherwise we have $B_{\omega+2}, \ldots$ and so on and so forth. If some $B_{\omega+n}$ is a basis we stop. Otherwise we define $B_{\omega+\omega}$ and check whether that is a basis. If not we continue ...

The claim is then that eventually[1] this process will stop and we have our basis.

For example let's reexamine the process of finding a basis for $K[x]$. Aw we indicated above if at the k step we chose $v_k = x^k$ then we get the standard basis of $K[x]$ as B_ω. But we could have made other choices, for example we could chose $v_1 = x^2$, $v_2 = x^4$, and so on: at the k step we choose $v_k = x^{2k}$, then we would get

$$B_\omega = \left\{x^{2k} : k \in \mathbb{N}\right\}$$

and B_ω is not a basis.

We could then proceed and choose $v_{\omega+k} = x^{2k+1}$ and then we would obtain the standard basis as $B_{\omega+\omega}$.

"But", I hear you ask, "how do we know that this process will really stop?" "Well", I answer smugly, "it's because we put this as an axiom." You see, so far we have been dealing with sets in a *naive* way, as if their properties are obvious. However this is dangerous and if

[1] This may really take loooooong, looooong time!

we are not careful it leads to contradictions. Mathematicians had to deal with such issues at the beginning of the previous century. This a rather long story, which I would love to go into but alas, that would take us far afield. So I'm only giving you the moral: "in order to avoid paradoxes we have to work within an *axiomatic set theory*".

There is a (more or less) universally accepted set of axioms that sets are supposed to satisfy. One of those axioms, the so called *Axiom of Choice* guarantees that the above procedure will eventually terminate thus producing a basis of V[2].

The inductive procedure that we outlined above is an informal application of *transfinite induction*, a generalization of the ordinary induction over the set of natural numbers.

One final remark, the existence of basis is usually given as an application of the so called *Zorn's Lemma*, a statement equivalent to the axiom of choice. Roughly speaking, Zorn's Lemma guarantees that if we have a class of sets $\mathcal{C}$ that is closed under unions of increasing families, then there is a maximal element in that class, i.e. a set that is not a subset of any other set in $\mathcal{C}$. If we take $\mathcal{C}$ to be the set of all linearly independent subsets of V, then $\mathcal{C}$ is closed under unions of increasing families and so there is a maximal element in $\mathcal{C}$, i.e. there exists a maximal linearly independent subset of V. But by Item (d) of Theorem 4.3.2, a maximal linearly independent subset is a basis.

A quick note on dimension. We note that Theorem 2.2.3 does not hold in general. If V is infinite dimensional with basis B, and B' is a linearly independent subset of V with $|B'| = |B'|$ it does not follow that B' is a basis. For example,

$$B' = \left\{x^{2k} : k \in \mathbb{N}\right\} \subseteq K[x]$$

is linearly independent and has the same cardinality as the standard basis of $K[x]$ but clearly[3] B' is not a basis.

Dealing seriously with the cardinality of infinite sets is beyond the scope of these notes, as it requires a deeper excursion into the wild world of set theory. However, Lemma 2, holds for all vector spaces and this allow us to prove that if B and B' are two bases of the vector space V then

$$|B'| \leq |B|.$$

Indeed, each $v \in B'$ can replace some element of B and to give us a new basis B'' that has the same cardinality as as B and $B' \subseteq B''$.

But similarly,

$$|B| \leq |B'|,$$

and therefore

$$|B'| = |B'|.$$

We remark, that in the case of infinite sets the seemingly "obvious" statement

$$|B'| \leq |B| \text{ and } |B| \leq |B'| \implies |B'| = |B'|,$$

is called the Cantor–Bernstein Theorem[4] and is rather subtle, not obvious by any means.

In any case, even though we haven't really rigorously proved Theorem 4.3.3 in the case of infinitely dimensional spaces, I hope that the discussion in this Appendix gives you enough confidence to believe that a proof does indeed exist.

[2] Actually, given the other axioms of set theory, the axiom of choice is equivalent to the statement that every vector space has a basis.

[3] Is it clear to you?

[4] The name Cantor–Schröder–Bernstein Theorem is also often used.

APPENDIX E

Homework

In this appendix we collect the assigned homework and provide solutions and answers.

E.1. Homework Set 1

Exercise E.1 Solve each of the following systems:

(a)

$$\begin{cases} x + 2y + 3z = 0 \\ 3x + y + 2z = 0 \\ 2x + 3y + z = 0 \end{cases}.$$

ANSWER. This is a homogeneous system. A row echelon form of the coefficient matrix is

$$A \begin{pmatrix} 1 & 0 & 11 \\ 0 & 1 & 5 \\ 0 & 0 & 18 \end{pmatrix}.$$

Since there are no free columns we conclude that the system has only the trivial solution. The solution set is therefore $\{(0,0,0)\}$. □

(b)

$$\begin{cases} x - y + z = 0 \\ -x + 3y + z = 5 \\ 3x + y + 7z = 2 \end{cases}.$$

ANSWER. The system is inconsistent. The solutions set is $\varnothing$. □

(c)

$$\begin{cases} x_1 + 3x_2 - 2x_3 + 2x_5 = 0 \\ 2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 = -1 \\ \phantom{2x_1 + 6x_2 - {}} 5x_3 + 10x_4 + 15x_6 = 5 \\ 2x_1 + 6x_2 + 8x_4 + 4x_5 + 18x_6 = 6 \end{cases}.$$

ANSWER. The reduced echelon form (after discarding a zero row) of the augmented matrix is

$$\left(\begin{array}{cccccc|c} 1 & 3 & 0 & 4 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/3 \end{array}\right).$$

So the solution is

$$x_1 = -3\,s - 4\,t - 2\,w,\ x_2 = s,\ x_3 = -2\,t,\ x_4 = t,\ x_5 = w,\ x_6 = \frac{1}{3}.$$

□

Exercise E.2 Find conditions on the real numbers a, b, c, if any, so that the system

$$\left\{\begin{array}{ccccccc} x & + & y & & & = 0 \\ & & y & + & z & = 0 \\ x & & & - & z & = 0 \\ ax & + & by & + & cz & = 0 \end{array}\right.$$

(a) is inconsistent.
(b) Has a unique solution.
(c) Has more than one solution.

ANSWER. We can immediately answer the first part. This is a homogeneous system and is therefore consistent. Thus there exists no conditions on a, b, c that the system is inconsistent.

To answer parts (b) and (c) we proceed to reduce the matrix of the system to an echelon form.

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & -1 \\ a & b & c \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & -1 \\ 0 & b-a & c \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & b-a & c \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & b-a & c \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 0 \\ 0 & b-a & c \\ 0 & 0 & 1 \end{pmatrix}.$$

If $b \neq a$ then the system has a unique solution. If $b = a$ then the second column is free and thus the system has more than one solutions. □

Exercise E.3 Consider the 2×2 matrix

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

where $a, b, c, d \in \mathbb{R}$.

(a) Prove that if $a\,d - b\,c \neq 0$ then the reduced row echelon form of A is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

SOLUTION. Set $D = a\,d - b\,c$. We are given $D \neq 0$. We distinguish two cases:

Case I: $a = 0$. Then we interchange the rows and we get

$$A = \begin{pmatrix} 0 & b \\ c & d \end{pmatrix} \sim \begin{pmatrix} c & d \\ 0 & b \end{pmatrix}.$$

Since, $D \neq 0$ we have $b\,c \neq 0$ and therefore $b \neq 0$ and $c \neq 0$. So we divide the second row by b, and the first row by c and we have

$$A \sim \begin{pmatrix} 1 & d/c \\ 0 & 1 \end{pmatrix}.$$

Finally we add $-d/c$ times the second row to the first and we get

$$A \sim \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

.

Case II: $a \neq 0$. We first divide the first row by a, then add $-c$ times the first row to the second, and get

$$A \sim \begin{pmatrix} 1 & b/a \\ c & d \end{pmatrix} \sim \begin{pmatrix} 1 & b/a \\ 0 & d-(b\,c)/a \end{pmatrix} = \begin{pmatrix} 1 & b/a \\ 0 & (a\,d - b\,c)/a \end{pmatrix} = \begin{pmatrix} 1 & b/a \\ 0 & D/a \end{pmatrix}.$$

Since $D \neq 0$ we can multiply the second row by a/D, and then add $-a/b$ times the second row to the first:

$$A \sim \begin{pmatrix} 1 & b/a \\ c & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

□

(b) Prove that if $a\,d - b\,c \neq 0$ then the system

$$\begin{cases} ax + by &= k \\ cx + dy &= l \end{cases}$$

has a unique solution, for all real numbers k, l.

SOLUTION. This follows from Part (a) and the second Item of Theorem 1.2.8. □

REMARK 24. In Section refsec:2x2 we treat 2×2 systems in detail.

Exercise E.4 Prove that there is a unique line passing through any two *distinct* points of the plane.

SOLUTION. A line is a set of points in $\mathbb{R}^2$ whose coordinates (x, y)satisfy a linear equation of the form

$$a\,x + b\,y + c = 0 \tag{E.1}$$

where $a, b, c \in \mathbb{R}$ and at least one of a, b is non-zero. A non-zero multiple of Equation (E.1) defines the same line.

Let (x_1, y_1) and (x_2, y_2) two points, to find all lines that pass through these two points we solve the system

$$\begin{cases} a\,x_1 + b\,y_1 + c &= 0 \\ a\,x_2 + b\,y_2 + c &= 0 \end{cases}$$

for a, b, c.

Set $\Delta x = x_2 - x_1$, and $\Delta y = y_2 - y_1$. Since the points are distinct at least one of Δx, Δy is non-zero. Without loss of generality we assume that $\Delta x \neq 0$. Subtracting the two equations we get

$$a\,\Delta x + b\,\Delta y = 0 \implies a = -\frac{\Delta y}{\Delta x}\,b.$$

Substituting in the first equation we get

$$-\frac{x_1\,\Delta y}{\Delta x}\,b + y_1\,b + c = 0 \implies c = \left(\frac{x_1(y_2 - y_1) - y_1(x_2 - x_1)}{x_2 - x_1}\right) b \implies c = \frac{x_1 y_2 - x_2 y_1}{x_2 - x_1}\,b$$

So the solution is

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = t \begin{pmatrix} \dfrac{x_1\,y_2 - x_2\,y_1}{x_2 - x_1} \\ 1 \\ \dfrac{y_2 - y_1}{x_2 - x_1} \end{pmatrix}, \quad t \in \mathbb{R}.$$

Thus all equations of the form (E.1) that are satisfied by the coordinates of both points are multiples of the same equation and therefore determine the same line. □

Exercise E.5 Find the cubic polynomial

$$p(x) = a\,x^3 + b\,x^2 + c\,x + d$$

given that $p(1) = 0$, $p(2) = 3$, $p(-1) = -6$, and $p(-2) = -21$.

ANSWER. Substituting the given values we get the system

$$\begin{cases} a + b + c + d = 0 \\ 8a + 4b + 2c + d = 3 \\ -a + b - c + d = -6 \\ -8a + 4b - 2c + d = -21 \end{cases}.$$

Passing to the augmented matrix we have

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 8 & 4 & 2 & 1 & 3 \\ -1 & 1 & -1 & 1 & -6 \\ -8 & 4 & -2 & 1 & -21 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right).$$

Therefore the polynomial is

$$p(x) = x^3 - 2x^2 + 2x - 1.$$

□

Exercise E.6 Look at Examples 5 and 6. There is a geometric reason why in Example 6 the polynomial we got was not quadratic. The graph of a quadratic polynomial is a parabola so in these examples we were trying to find a parabola that passes through three distinct points. But the points of Example 6 are *colinear* and so there is no parabola that passes through all three of them.

(a) Prove that given any three *distinct* real numbers x_1, x_2, x_3 and any three real numbers y_1, y_2, y_3 we can always find a polynomial $p(x) = a\,x^2 + b\,x + c$ such that $p(x_1) = y_1$, $p(x_2) = y_2$, and $p(x_3) = y_3$.

SOLUTION. Let $a, b, c \in \mathbb{R}$ be the coefficients of p. Then (a, b, c) is a solution of the system

$$\begin{cases} c + x_1\,b + x_1^2\,a & = y_1 \\ c + x_2\,b + x_2^2\,a & = y_2 \\ c + x_3\,b + x_3^2\,a & = y_3 \end{cases}$$

The augmented matrix of the system is

$$\left(\begin{array}{ccc|c} 1 & x_1 & x_1^2 & y_1 \\ 1 & x_2 & x_2^2 & y_2 \\ 1 & x_3 & x_3^2 & y_3 \end{array}\right).$$

Subtracting the first row from the other two we get

$$\left(\begin{array}{ccc|c} 1 & x_1 & x_1^2 & y_1 \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & y_2 - y_1 \\ 0 & x_3 - x_1 & x_3^2 - x_1^2 & y_3 - y_1 \end{array}\right).$$

Since x_1, x_2 and x_3 are distinct, $x_2 - x_1$ and $x_3 - x_1$ are non-zero. We can also assume that $x_1 \neq 0$, for if it is zero then x_2 is non-zero and we just rename our numbers. So we can divide each row by its leading entry to get[1]

$$\left(\begin{array}{ccc|c} 1 & 1 & x_1 & y_1/x_1 \\ 0 & 1 & x_2 + x_1 & (y_2 - y_1)/(x_2 - x_1) \\ 0 & 1 & x_3 + x_1 & (y_3 - y_1)/(x_3 - x_1) \end{array}\right).$$

Now subtract the second row from the third to get

[1] Remember "Difference of Squares"?

$$\left(\begin{array}{ccc|c} 1 & 1 & x_1 & y_1/x_1 \\ 0 & 1 & x_2+x_1 & (y2-y1)/(x_2-x_1) \\ 0 & 0 & x_3-x_2 & (y_3-y_1)/(x_3-x_1)-(y_2-y_1)/(x_2-x_1) \end{array}\right).$$

Since, $x_3 - x_2 \neq 0$ we conclude that the system has a unique solution. □

(b) The polynomial in part (a) is quadratic (i.e. $a \neq 0$) if and only if the points (x_1, y_1), (x_2, y_2), and (x_3, y_3) are not colinear.

SOLUTION. From the echelon form from part one we see that $a = 0$ if and only if

$$\frac{y_3 - y_1}{x_3 - x_1} = \frac{y_2 - y_1}{x_2 - x_1}. \tag{E.2}$$

The fraction on the LHS of Equation (E.2) is the slope of the line through the points (x_1, y_1) and (x_3, y_3) and the one on the RHS is the slope of the line through (x_1, y_1) and (x_2, y_2). Since these lines share the point (x_1, y_1) they are the same line if and only if they have the same slope. Now (x_1, y_2), (x_2, y_2), and (x_3, y_3) are colinear if and only if these lines are the same line, and we conclude that $a = 0$ if and only if the three points are colinear. □

E.2. Second Homework

Exercise E.1 Solve the system

$$\begin{cases} 2x - 5y + 2z - 4s + 2t &= 4 \\ 3x - 7y + 2z - 5s + 4t &= 9 \\ 5x - 10y - 5z - 4s + 7t &= 22 \end{cases}$$

by first solving the corresponding homogeneous system and then finding a particular solution. Refer to Example 13 in Section 1.2.3.

ANSWER. The corresponding homogeneous system is

$$\begin{cases} 2x - 5y + 2z - 4s + 2t &= 0 \\ 3x - 7y + 2z - 5s + 4t &= 0 \\ 5x - 10y - 5z - 4s + 7t &= 0 \end{cases}$$

We find the reduced echelon for of its matrix:

$$\begin{pmatrix} 2 & -5 & 2 & -4 & 2 \\ 3 & -7 & 2 & -5 & 4 \\ 5 & -10 & -5 & -4 & 7 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & \frac{11}{5} & \frac{42}{5} \\ 0 & 1 & 0 & \frac{8}{5} & \frac{16}{5} \\ 0 & 0 & 1 & -\frac{1}{5} & \frac{3}{5} \end{pmatrix}.$$

The general solution of the homogeneous system is therefore

$$\begin{pmatrix} x \\ y \\ z \\ s \\ t \end{pmatrix} = \frac{a}{5}\begin{pmatrix} 11 \\ 8 \\ -1 \\ 5 \\ 0 \end{pmatrix} + \frac{b}{5}\begin{pmatrix} 42 \\ 16 \\ 3 \\ 0 \\ 5 \end{pmatrix}.$$

To find a particular solution of the original system we try to guess: we substitute values to some of the variables and solve for the others until we find a solution that works. If we put $z = s = 0$ and $t = 1$ we find that $x = 11$ and $y = 4$ works for all equations. So $(11, 4, 0, 1, 0)$ is a particular solution and so the general solution of the original system is

$$\begin{pmatrix} x \\ y \\ z \\ s \\ t \end{pmatrix} = \frac{a}{5}\begin{pmatrix} 11 \\ 8 \\ -1 \\ 5 \\ 0 \end{pmatrix} + \frac{b}{5}\begin{pmatrix} 42 \\ 16 \\ 3 \\ 0 \\ 5 \end{pmatrix} + \begin{pmatrix} 11 \\ 4 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

□

Exercise E.2 Express the vector $\mathbf{c} = 3\,\mathbf{e}_1 - 2\,\mathbf{e}_2 - \mathbf{e}_3$ as a linear combination of the vectors

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{e}_1 + 2\,\mathbf{e}_2 + 3\,\mathbf{e}_3 \\ \mathbf{v}_2 &= 2\,\mathbf{e}_1 + 3\,\mathbf{e}_2 + \mathbf{e}_3 \\ \mathbf{v}_3 &= 3\,\mathbf{e}_1 + \mathbf{e}_2 + 2\,\mathbf{e}_3. \end{aligned}$$

ANSWER. The coefficients x, y, z will be solutions of the system

$$\begin{cases} x + 2y + 3z = 3 \\ 2x + 3y + z = -2 \\ 3x + y + 2z = -1 \end{cases}.$$

Working with the augmented matrix we get

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 3 \\ 2 & 3 & 1 & -2 \\ 3 & 1 & 2 & -1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & -\frac{4}{3} \\ 0 & 1 & 0 & -\frac{1}{3} \\ 0 & 0 & 1 & \frac{5}{3} \end{array}\right).$$

So,

$$\mathbf{c} = -\frac{4}{3}\,\mathbf{v}_1 - \frac{1}{3}\,\mathbf{v}_2 + \frac{5}{3}\,\mathbf{v}_3.$$

□

Exercise E.3 Express the vector $\mathbf{c} = 5\,\mathbf{e}_1 - \mathbf{e}_2 + 3\,\mathbf{e}_3$ as a linear combination of the vectors

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{e}_1 - 2\,\mathbf{e}_2 + 3\,\mathbf{e}_3 \\ \mathbf{v}_2 &= 4\,\mathbf{e}_1 + \mathbf{e}_2 \\ \mathbf{v}_3 &= \mathbf{e}_1 - 11\,\mathbf{e}_2 + 15\,\mathbf{e}_3, \end{aligned}$$

in three different ways.

ANSWER. The augmented matrix of the system we get reduces to

$$\left(\begin{array}{ccc|c} 1 & 0 & 5 & 1 \\ 0 & 1 & -1 & 1 \end{array}\right).$$

and the solution is

$$(x, y, z) = t\,(-5, 1, 1) + (1, 1, 0).$$

Setting arbitrarily, $t = 0, \pm 1$, we get three different solutions:

$$\begin{aligned}\mathbf{c} &= \mathbf{v}_1 + \mathbf{v}_2\\ &= -4\,\mathbf{v}_1 + 2\,\mathbf{v}_2 + \mathbf{v}_3\\ &= 6\,\mathbf{v}_1 - \mathbf{v}_3.\end{aligned}$$

□

Exercise E.4 Find a vector $\mathbf{c}$ that cannot be expressed as a linear combination of the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ of the previous exercise.

SOLUTION. From the previous question we know that reduced echelon form of the matrix A with columns $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ has a zero row. If $\mathbf{c}$ is such that the matrix A augmented by $\mathbf{c}$ has at that point transformed to a matrix that has a non-zero entry in that row, the system is inconsistent and therefore $\mathbf{c}$ cannot be expressed as a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.

Let's apply the Gauss-Jordan procedure then until we get the zero row.

$$\begin{pmatrix}1 & 4 & 1\\ -2 & 1 & -11\\ 3 & 0 & 15\end{pmatrix} \sim \begin{pmatrix}1 & 4 & 1\\ 0 & 9 & -9\\ 0 & -12 & 12\end{pmatrix} \sim \begin{pmatrix}1 & 4 & 1\\ 0 & 1 & -1\\ 0 & -1 & 1\end{pmatrix} \sim \begin{pmatrix}1 & 4 & 1\\ 0 & 1 & -1\\ 0 & 0 & 0\end{pmatrix}.$$

We applied, in order, the following row operations:

(a) Added 2 times the first row to the second.
(b) Added -3 times the first row to the third.
(c) Divided the second row by 3.
(d) Divided the third row by 4.
(e) Added the second row to the third.

If we start with any vector $\mathbf{c}'$ with non-zero third coordinate and apply the *reverse* of the above operations, *in reverse order*, we will get a vector $\mathbf{c}$ that is not in the span of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. For example starting with $\mathbf{c}' = (1, 3, -2)$ we get

$$(1,3,-2) \sim (1,3,-5) \sim \left(1,3,-\frac{5}{4}\right) \sim \left(1,1,-\frac{5}{4}\right) \sim \left(1,1,\frac{7}{4}\right) \sim \left(1,-1,\frac{7}{4}\right).$$

So, one such $\mathbf{c}$ is

$$\mathbf{c} = \mathbf{e}_1 - \mathbf{e}_2 + \frac{7}{4}\mathbf{e}_3.$$

□

REMARK 25. In the above solution I chose a random vector to illustrate the idea. However there is a much easier choice, I could have chosen $\mathbf{c}' = (0, 0, 4)$. Then only the fourth of the operations affect $\mathbf{c}'$ and as a result I would get $\mathbf{c} = \mathbf{e}_3$.

Note that since $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is not spanning we know that at least one vector from the standard basis (or any basis) of $\mathbb{R}^3$ is not in $\langle \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\rangle$[2].

Exercise E.5 Let

$$A = \begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}, \qquad B = \begin{pmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix}x_1\\ x_2\end{pmatrix}.$$

Let $\mathbf{y} = B\,\mathbf{x}$. Find the vector $\mathbf{z} = A\,\mathbf{y}$.

SOLUTION. We first find $\mathbf{y}$:

$$\mathbf{y} = \begin{pmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}b_{11}\,x_1 + b_{12}\,x_2\\ b_{21}\,x_1 + b_{22}\,x_2\end{pmatrix}.$$

[2] Why?

Now $\mathbf{z}$:

$$\mathbf{z} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11}\,x_1 + b_{12}\,x_2 \\ b_{21}\,x_1 + b_{22}\,x_2 \end{pmatrix} = \begin{pmatrix} a_{11}\,(b_{11}\,x_1 + b_{12}\,x_2) + a_{12}\,(b_{21}\,x_1 + b_{22}\,x_2) \\ a_{21}\,(b_{11}\,x_1 + b_{12}\,x_2) + a_{22}\,(b_{21}\,x_1 + b_{22}\,x_2) \end{pmatrix}.$$

We now factor x_1 and x_2 to get

$$\mathbf{z} = \begin{pmatrix} (a_{11}\,b_{11} + a_{12}\,b_{21})\;x_1 + (a_{11}\,b_{12} + a_{12}\,b_{22})\;x_2 \\ (a_{21}\,b_{11} + a_{22}\,b_{21})\;x_1 + (a_{21}\,b_{12} + a_{22}\,b_{22})\;x_2 \end{pmatrix}.$$

□

Exercise E.6 Find a 2×2 matrix A that interchanges $\mathbf{e}_1$ and $\mathbf{e}_2$, in other words such that

$$A\,\mathbf{e}_1 = \mathbf{e}_2 \quad \text{and} \quad A\,\mathbf{e}_2 = \mathbf{e}_1.$$

SOLUTION. If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then

$$A\,\mathbf{e}_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \cdot 1 + b \cdot 0 \\ c \cdot 1 + d \cdot 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix},$$

and

$$A\,\mathbf{e}_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} a \cdot 0 + b \cdot 1 \\ c \cdot 0 + d \cdot 1 \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}.$$

So we have the following two vector equations

$$\begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} b \\ d \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Thus the matrix is

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

□

Exercise E.7 Prove that if

$$a_1\,b_2 - a_1\,b_3 + a_2\,b_3 - a_3\,b_2 + a_3\,b_1 - a_2\,b_1 \neq 0$$

then the system

$$\begin{cases} x + a_1 y + b_1 z & = c_1 \\ x + a_2 y + b_2 z & = c_2 \\ x + a_3 y + b_3 z & = c_3 \end{cases}$$

has a unique solution for all real numbers c_1, c_2, c_3.

ANSWER. We proceed to get an echelon form of the augmented matrix of the system.

$$\left(\begin{array}{ccc|c} 1 & a_1 & b_1 & c_1 \\ 1 & a_2 & b_2 & c_2 \\ 1 & a_3 & b_3 & c_3 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & a_1 & b_1 & c_1 \\ 0 & a_2 - a_1 & b_2 - b_1 & c_2 - c_1 \\ 0 & a_3 - a_1 & b_3 - b_1 & c_3 - c_1 \end{array}\right).$$

Now notice that the condition given implies that at least one of $a_2 - a_1$, $a_3 - a_1$ is non-zero. For, if both were zero then we would have $a_1 = a_2 = a_3$ and the LHS of the condition would be zero. Assume then that $a_2 - a_1 \neq 0$ so we can divide the second row by it. Do that and then add $(a_1 - a_3)$ times the second row to the third. The second entry of the third row will then be 0 while the third is

$$b_3 - b_1 + \frac{(a_1 - a_3)(b_2 - b_1)}{a_2 - a_1}.$$

Combining and expanding this will give a fraction with numerator the LHS of the given inequality. Thus the third entry of the third row is non-zero. It follows that the system has a unique solution. □

E.3. Homework 3

Exercise E.1 Let S_1 and S_2 be two subsets of $\mathbb{R}^n$ with $S_1 \subseteq S_2$. Prove

(a) If S_1 is spanning then S_2 is also spanning.

(b) If S_1 is linearly dependent then S_2 is also linearly dependent.

(c) If S_2 is linearly independent then S_1 is also linearly independent.

SOLUTION. Note that (b) and (c) are logically equivalent: (c) is the contrapositive of (b). So we'll prove (a) and (b). Both follow from the fact that a linear combination of elements of S_1 is also a linear combination of elements of S_2.

(a) If S_1 is spanning then every vector $\mathbf{v} \in \mathbb{R}^n$ can be expressed as a linear combination of elements of S_1, and hence as a linear combination of elements of S_2. Therefore S_2 is spanning.

(b) If S_1 is linearly dependent then $\mathbf{0}$ can be expressed as a non-trivial linear combination of elements of S_1, and hence as a non-trivial linear combination of elements of S_2. Therefore S_2 is linearly dependent.

□

Exercise E.2 Decide whether each of the following subsets is a vector subspace of the given standard real vector space.

(a) $\{(x, 3x + y, 0, y - z) : x, y, z \in \mathbb{R}\} \subseteq \mathbb{R}^4$.

(b) $\{(x, y, z) \in \mathbb{R}^3 : x, y, z \in \mathbb{R}$ and $3\,x - 4\,y = 11\,z\}$.

(c) The set of points in $\mathbb{R}^2$ that lie in the parabola $y = x^2$.

(d) The set of points in $\mathbb{R}^3$ that lie in the plane with equation $2x - 3y + 4z = 0$.

(e) The set of points in $\mathbb{R}^3$ that lie in the plane with equation $2x - 3y + 4z = 8$.

(f) $\{(x, 2, 3x + 4y, y - z) : x, y, z \in \mathbb{R}\} \subseteq \mathbb{R}^4$.

(g) $\{(3\,w, 2\,z - 5\,t, x - 4\,y + 5\,t, -2x + z - 3t + 4w) : x, y, z, w, t \in \mathbb{R}\} \subseteq \mathbb{R}^4$.

ANSWER. (a) Yes, this is a vector subspace of $\mathbb{R}^4$. It is nonempty since it contains the zero vector. Simple calculations show that

$$\begin{aligned}&\lambda\,(x_1, 3\,x_1 + y_1, 0, y_1 - z_1) + \mu\,(x_2, 3\,x_2 + y_2, 0, y_2 - z_2)\\ &= (\lambda\,x_1 + \mu\,x_2, 3\;(\lambda\,x_1 + \mu\,x_2) + (y_1 + y_2), 0, (y_1 + y_2) - (z_1 + z_2))\end{aligned}$$

So both conditions of Theorem 4.5 hold.

Alternatively we can show that this subset is the linear span of the vectors

$$(1, 3, 0, 0), (0, 1, 0, 1), (0, 0, 0, -1).$$

(b) Yes, this subset is a vector subspace of $\mathbb{R}^3$. Perhaps the easiest way to see this is to note that this subset is the solution set of the homogeneous linear equation $3\,x - 4\,y - 11\,z = 0$.

(c) No this is not a vector subspace. For example it contains $\mathbf{v} = (1, 1)$ but not $2\,\mathbf{v} = (2, 2)$.

(d) Yes, it's the solution set of a homogeneous linear equation.

(e) No. This subset does not contain the zero vector.

(f) No. This subset does not contain the zero vector.

(g) Yes. We can either use Theorem 4.5 or note that this subset is the linear span of the vectors

$$(0, 0, 1, -2), (0, 0, -4, 0), (0, 2, 0, 1), (0, -5, 5, 4), (3, 0, 0, 4).$$

□

Exercise E.3 For those subsets in Question E.2 that are subspaces find a basis and the dimension.

ANSWER. For (a) we have the spanning set

$$S = \{(1,3,0,0),(0,1,0,1),(0,0,0,-1).\}\,.$$

The reduced echelon form is:

$$\begin{pmatrix}1&0&0\\3&1&0\\0&0&0\\0&1&-1\end{pmatrix}\sim\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\end{pmatrix}.$$

There are no free columns so these vectors are linearly independent, and form a basis. Therefore the dimension is 3.

For (b) we first find a spanning set. We have that the solution is

$$(x,y,z) = \left(\frac{1}{3}\left(4t+11s\right),t,s\right) = t\left(\frac{4}{3},1,0\right)+s\left(\frac{11}{3},0,1\right).$$

The set $\left\{(\frac{4}{3},1,0),(\frac{11}{3},0,1)\right\}$ is thus spanning and since it's linear independent if forms a basis. The dimension is therefore 2.

(d) is similar to (c). There are two free variables, y,z, and we again get a 2-dimensional subspace. A basis is

$$\left\{\left(\frac{3}{2},1,0\right),(2,0,1)\right\}.$$

(g) is similar to (a). The basic columns are the first, second, third, and fifth. So a basis is

$$\{(0,0,1,-2),(0,0,-4,0),(0,2,0,1),(3,0,0,4)\}\,,$$

and the dimension is 4. □

Exercise E.4 Which of the following subsets of $\mathbb{R}^3$ are a basis?

(a) $\{(1,2,3),(3,2,1)\}$.
(b) $\{(1,1,2),(1,-2,0),(2,0,1)\}$
(c) $\{(1,2,3),(3,1,2),(2,3,1)\}$.
(d) $\{(1,2,3),(1,1,0),(0,3,1),(1,0,0)\}$.

ANSWER. A basis of $\mathbb{R}^3$ contains exactly 3 vectors, so (a) and (d) are not bases. Both (b) and (c) are bases since the reduced echelon form of the matrices with columns those vectors is the identity matrix. □

Exercise E.5 Let

$$B = \{(1,1,1,1,1),(0,1,1,1,1),(0,0,1,1,1),(0,0,0,1,1),(0,0,0,0,1)\}\,.$$

(a) Prove that B is a basis of $\mathbb{R}^5$.
(b) Express the elements of the standard basis of $\mathbb{R}^5$ as linear combinations of elements of B.

ANSWER. As indicated in the hint, if we succeed in completing part (b), that is if we express every vector of the standard basis as a linear combination of elements of B then it follows that B is a basis[3].

[3] Why?

Let $\mathbf{v}_i$, $i = 1, \ldots, 5$ be the elements of B in the order given. We augment the matrix with columns the vectors of B with the five vectors of the standard basis and find its reduced echelon form.

$$\left(\begin{array}{ccccc|ccccc} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccccc|ccccc} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 1 \end{array}\right).$$

Therefore

$$\begin{aligned} \mathbf{e}_1 &= \mathbf{v}_1 - \mathbf{v}_2 \\ \mathbf{e}_2 &= \mathbf{v}_2 - \mathbf{v}_3 \\ \mathbf{e}_3 &= \mathbf{v}_3 - \mathbf{v}_4 \\ \mathbf{e}_4 &= \mathbf{v}_4 - \mathbf{v}_5 \\ \mathbf{e}_5 &= \mathbf{v}_5. \end{aligned}$$

□

Exercise E.6 Let $T\colon \mathbb{R}^3 \to \mathbb{R}^4$ be defined by

$$T(x, y, z) = (x + 2y + z, x + y, y - 3z, 4x - 3y + 2z).$$

(a) Prove that T is linear.
(b) Find the matrix of T.

ANSWER. (a) We need to check that for $\lambda_i \in \mathbb{R}$ and $\mathbf{v}_i = (x_i, y_i, z_i) \in \mathbb{R}^3$, where $i = 1, 2$ we have:

$$T(\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2) == \lambda_1 T\mathbf{v}_1 + \lambda_2 T\mathbf{v}_2.$$

This is a straightforward calculation.

(b) We have

$$T\mathbf{e}_1 = (1, 1, 0, 4), \quad T\mathbf{e}_2 = (2, 1, -1, -3), \quad T\mathbf{e}_3 = (1, 0, -3, 2).$$

So the matrix is

$$T = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 0 \\ 0 & -1 & -3 \\ 4 & -3 & 2 \end{pmatrix}.$$

□

Exercise E.7 Let

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

(a) Prove that the columns of A form a basis of $\mathbb{R}^3$.
(b) Express each of the vectors in the standard basis of $\mathbb{R}^3$ as linear combinations of the columns of A.
(c) Let T be the linear function that sends the i-th column to the i-th row of A. That is if $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are the columns of A then T is defined by

$$T\mathbf{a}_1 = (1, 2, 3), \quad T\mathbf{a}_2 = (0, 2, 1), \quad T\mathbf{a}_3 = (1, 0, 3).$$

Find the matrix of T.

ANSWER. (a) The reduced echelon form of A is the identity matrix.

(b) We have

$$\left(\begin{array}{ccc|ccc}1 & 2 & 3 & 1 & 0 & 0\\ 0 & 2 & 1 & 0 & 1 & 0\\ 1 & 0 & 3 & 0 & 0 & 1\end{array}\right) \sim \left(\begin{array}{ccc|ccc}1 & 0 & 0 & 3 & -3 & -2\\ 0 & 1 & 0 & 1/2 & 0 & -1/2\\ 0 & 0 & 1 & -1 & 1 & 1\end{array}\right).$$

And so, letting $\mathbf{a}_i$ be the columns of A, we have:

$$\begin{aligned}\mathbf{e}_1 &= 3\,\mathbf{a}_1 + \frac{1}{2}\,\mathbf{a}_2 - \mathbf{a}_3\\ \mathbf{e}_2 &= -3\,\mathbf{a}_1 + \mathbf{a}_3\\ \mathbf{e}_3 &= -2\,\mathbf{a}_1 - \frac{1}{2}\,\mathbf{a}_2 + \mathbf{a}_3.\end{aligned}$$

(c) We need to find $T\,\mathbf{e}_i$ for $i = 1, 2, 3$. We will use the linearity of T and the expressions of $\mathbf{e}_i$ as linear combinations of $\mathbf{a}_i$ from Part (b).
We have:

$$\begin{aligned}T\,\mathbf{e}_1 &= T\left(3\,\mathbf{a}_1 + \frac{1}{2}\,\mathbf{a}_2 - \mathbf{a}_3\right)\\ &= 3\,T\,\mathbf{a}_1 + \frac{1}{2}\,T\,\mathbf{a}_2 - T\,\mathbf{a}_3\\ &= (3, 6, 9) + \left(0, 1, \frac{1}{2}\right) - (1, 0, 3)\\ &= \left(2, 7, \frac{13}{2}\right).\end{aligned}$$

Similarly,

$$T\,\mathbf{e}_2 = (-2, -6, -6), \quad T\,\mathbf{e}_3 = \left(-1, -5, -\frac{7}{2}\right).$$

Thus T is induced by the matrix

$$T = \begin{pmatrix} 2 & -2 & -1\\ 7 & -6 & -5\\ \frac{13}{2} & -6 & -\frac{7}{2}\end{pmatrix}.$$

□

Exercise E.8 Find a polynomial of degree at most 4 that satisfies the following conditions

$$p(0) = -5, \quad p(-1) = -10, \quad p(1) = 0, \quad p(2) = 29, \quad p(-2) = -15.$$

PROOF. The polynomial is

$$p(x) = x^4 + 2\,x^3 - x^2 + 3\,x - 5.$$

□

Exercise E.9 Let V be the subspace of $\mathbb{R}^5$ spanned by the vectors

$$\begin{aligned}\mathbf{v}_1 &= (1, 2, -1, 3, 4), & \mathbf{v}_2 &= (2, 4, -2, 6, 8),\\ \mathbf{v}_3 &= (1, 3, 2, 2, 6), & \mathbf{v}_4 &= (1, 4, 5, 1, 8),\\ \mathbf{v}_5 &= (2, 7, 3, 3, 9), & \mathbf{v}_6 &= (4, 9, -1, 11, 18).\end{aligned}$$

Find a basis and the dimension of V.

ANSWER. We have:

$$\begin{pmatrix}1 & 2 & 0 & -1 & 0 & 3\\ 0 & 0 & 1 & 2 & 0 & 1\\ 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix} \sim \begin{pmatrix}1 & 2 & 0 & -1 & 0 & 3\\ 0 & 0 & 1 & 2 & 0 & 1\\ 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix}$$

So $\{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_5\}$ is a basis, and the dimension of the subspace is 3. □

;

E.4. Homework 4

Exercise E.1 Let A be an $m \times k$ matrix and B a $k \times n$ matrix. Prove that

$$\ker B \subseteq \ker A B.$$

SOLUTION. We have

$$\begin{aligned} \mathbf{x} \in \ker B &\implies B\mathbf{x} = \mathbf{0} \\ &\implies A(B\mathbf{x}) = \mathbf{0} \\ &\implies (AB)\mathbf{x} = \mathbf{0} \\ &\implies \mathbf{x} \in \ker(AB). \end{aligned}$$

Therefore,

$$\ker B \subseteq \ker A B.$$

□

Exercise E.2 Let A be a 4×3 matrix and B a 3×4 so that AB is a square 4×4 matrix. Prove that AB is not invertible.

SOLUTION. We will prove that

$$\ker(AB) \neq \{\mathbf{0}\},$$

and so AB is not injective. To do that we will prove that $\ker B \neq \{\mathbf{0}\}$ and the result follows from Question 1.

By the *Rank-Nullity Theorem* (see Theorem 3.2.4 in the notes) we have

$$\operatorname{rank} B + \operatorname{null} B = 4,$$

which gives

$$\operatorname{null} B = 4 - \operatorname{rank} B. \tag{E.3}$$

But $\operatorname{rank} B \leq 3$ and therefore Equation (E.3) gives

$$\operatorname{null} B \geq 1,$$

that is

$$\dim(\ker B) \geq 1.$$

It follows that

$$\ker B \neq \{\mathbf{0}\}.$$

□

Exercise E.3 Find a basis and the dimension of the solution set of the following homogeneous system:

$$\begin{cases} x_1 - 3x_2 \qquad\qquad + x_4 + x_5 = 0 \\ 2x_1 - 6x_2 + 2x_3 + 4x_4 + 2x_5 = 0 \\ -3x_1 + 4x_2 \qquad\qquad\qquad + x_5 = 0 \\ \qquad x_2 + x_3 + x_4 \qquad = 0 \end{cases}$$

ANSWER. We get the reduced echelon form of the matrix of the system:

$$A = \begin{pmatrix} 1 & -3 & 0 & 1 & 1 \\ 2 & -6 & 2 & 4 & 2 \\ -3 & 4 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 & -1/3 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -4/3 \\ 0 & 0 & 0 & 1 & 4/3 \end{pmatrix}$$

The solution set is therefore

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = t \begin{pmatrix} 1/3 \\ 0 \\ 4/3 \\ -4/3 \\ 1 \end{pmatrix}.$$

Thus the solution set is the linear span $\langle(1/3, 0, 4/3, -4/3, 1)\rangle$, and has, therefore, dimension 1. □

Exercise E.4 For each of the following two transpose matrices:

$$A = \begin{pmatrix} 1 & 2 & 3 & 1 & 2 \\ 2 & 1 & 2 & 3 & 1 \\ 3 & 3 & 5 & 4 & 3 \\ 1 & -1 & -1 & 2 & -1 \end{pmatrix}, \quad A^* = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 2 & 1 & 3 & -1 \\ 3 & 2 & 5 & -1 \\ 1 & 3 & 4 & 2 \\ 2 & 1 & 3 & -1 \end{pmatrix}.$$

(a) Find a basis for their ranges and state their rank.

ANSWER. We have

$$A \sim \begin{pmatrix} 1 & 0 & 1/3 & 5/3 & 0 \\ 0 & 1 & 4/3 & -1/3 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus a basis for $\mathcal{R}(A)$ is $\{(1, 2, 3, 1), (2, 1, 3, -1)\}$ and so $\operatorname{rank} A = 2$.
For A^* we have

$$A^* \sim \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus a basis for $\mathcal{R}(A^*)$ is $\{(1, 2, 3, 1, 2), (2, 1, 2, 3, 1)\}$, and so $\operatorname{rank} A^* = 2$. □

(b) Find a basis for their kernels and state their nullity.

ANSWER. From the reduced echelon forms in Part (a) we have that a basis for the kernel of A is

$$\{(-1/3, -4/3, 1, 0, 0), (-5/3, 1/3, 0, 1, 0), (0, 1, 0, 0, 1)\}.$$

Thus $\operatorname{null} A = 3$.

Similarly, a basis for the kernel of A^* is

$$\{(-1,-1,1,0),(1,-1,0,1)\}.$$

Thus $\operatorname{null} A^* = 2$. □

Exercise E.5 Find the inverse of each of the following matrices:

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 2 & 4 \\ 1 & 3 & -3 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & -1 & 1 \\ 1 & 3 & 1 & -2 \\ 1 & 4 & -2 & 4 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 2 & 0 & -2 \\ 0 & 2 & 1 & 4 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & -8 \end{pmatrix}.$$

ANSWER. We have

$$\left(\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 2 & 2 & 4 & 0 & 1 & 0 \\ 1 & 3 & -3 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 9 & -3/2 & -5 \\ 0 & 1 & 0 & -5 & 1 & 3 \\ 0 & 0 & 1 & -2 & 1/2 & 1 \end{array}\right).$$

Therefore,

$$A^{-1} = \begin{pmatrix} 9 & -3/2 & -5 \\ -5 & 1 & 3 \\ -2 & 1/2 & 1 \end{pmatrix}.$$

Similarly,

$$B^{-1} = \begin{pmatrix} -10 & -20 & 4 & 7 \\ 3 & 6 & -1 & -2 \\ 5 & 8 & -2 & -3 \\ 2 & 3 & -1 & -1 \end{pmatrix}, \quad C^{-1} = \begin{pmatrix} 1 & -1 & 1/3 & -3/4 \\ 0 & 1/2 & -1/6 & 1/4 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & -1/8 \end{pmatrix}.$$

□

Exercise E.6 Let $\mathbf{C}$ be the following set of 2×2 matrices.

$$\mathbf{C} = \left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} : a, b \in \mathbb{R} \right\}.$$

Prove the following:

(a) $\mathbf{C}$ is closed under matrix addition. That is,

$$A, B \in \mathbf{C} \implies A + B \in \mathbf{C}.$$

(b) $\mathbf{C}$ is closed under scalar multiplication. That is,

$$\lambda \in \mathbb{R},\ A \in \mathbf{C} \implies \lambda A \in \mathbf{C}.$$

In particular,

$$A \in \mathbf{C} \implies -A \in \mathbf{C}.$$

(c) $\mathbf{C}$ is closed under matrix multiplication. That is,

$$A, B \in \mathbf{C} \implies AB \in \mathbf{C}.$$

(d) All non-zero elements of $\mathbf{C}$ are invertible, and

$$A \in \mathbf{C},\ A \neq O \implies A^{-1} \in \mathbf{C}.$$

(e) Any two elements of $\mathbf{C}$ commute. That is,

$$A, B \in \mathbf{C} \implies AB = BA.$$

(f) Let

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Verify that

$$J^2 = -I.$$

(g) what is J^{503}?

(h) Every element of **C** can be uniquely expressed as a linear combination of I and J. In other words, if $A \in \mathbf{C}$ then there exist two unique real numbers a, b such that

$$A = a\, I + b\, J.$$

(i) Let $A \in \mathbf{C}$ with $A = a\, I + b\, J$ with at least one of a, b non-zero. Express A^{-1} as a linear combination of I and J.

(j) Prove that every non-zero element of **C** has exactly two *square roots*. That is, prove that if $A \neq O$ is an element of **C** then there are exactly two elements $B \in \mathbf{C}$ such that $B^2 = A$.

(k) Prove the that every quadratic equation

$$A\, X^2 + B\, X + C = O$$

where $A, B, C \in \mathbf{C}$ and $A \neq 0$, has two solutions (that may coincide) in **C**.

Answer. Parts (a) through (e) are straightforward calculations, very similar to Example 52 in the notes. Let

$$A_1 = \begin{pmatrix} a_1 & -b_1 \\ b_1 & a_1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_2 & -b_2 \\ b_2 & a_2 \end{pmatrix},$$

be two elements of **C** and $\lambda \in \mathbb{R}$. We have

(a)

$$A_1 + A_2 = \begin{pmatrix} a_1 + a_2 & -b_1 - b_2 \\ b_1 + b_2 & a_1 + a_2 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 & -(b_1 + b_2) \\ b_1 + b_2 & a_1 + a_2 \end{pmatrix} \in \mathbb{C}.$$

(b)

$$\lambda\, A_1 = \begin{pmatrix} \lambda\, a_1 & \lambda\, (-b_1) \\ \lambda\, b_1 & \lambda\, a_1 \end{pmatrix} = \begin{pmatrix} \lambda\, a_1 & -(\lambda\, b_1) \\ \lambda\, b_1 & \lambda\, a_1 \end{pmatrix} \in \mathbf{C}.$$

(c)

$$A_1\, A_2 = \begin{pmatrix} a_1\, a_2 + b_1\, b_2 & -\,(a_1\, b_2 + a_2\, b_1) \\ a_1\, b_2 + a_2\, b_1 & a_1\, a_2 + b_1\, b_2 \end{pmatrix} \in \mathbf{C}.$$

(d) Using the result of Example 43, we need to prove that if $A_1 \in \mathbf{C}$ is non-zero then the determinant of A is non-zero. But the determinant of A_1 is

$$a_1\, a_1 - (-b_1)\, b_1 = a_1^2 + b_1^2.$$

But if $A_1 \neq O$ then at least one of a_1, b_1 is non-zero and therefore the determinant $a_1^2 + b_1^2 \neq 0$ and thus A_1 is invertible.

Then using Equation (3.9) we have

$$A_1^{-1} = \frac{1}{a_1^2 + b_1^2} \begin{pmatrix} a_1 & b_1 \\ -b_1 & a_1 \end{pmatrix}.$$

Now

$$\begin{pmatrix} a_1 & b_1 \\ -b_1 & a_1 \end{pmatrix} = \begin{pmatrix} a_1 & -(-b_1) \\ -b_1 & a_1 \end{pmatrix} \in \mathbf{C}$$

and therefore by Part (b) it follows that A^{-1} in **C**.

(e) This is a straightforward calculation.

(f) Again this is a straightforward calculation.

(g) Since $J^2 = -I$ we have $J^3 = -J$ and $J^4 = I$. It follows then that for $k, \ell \in \mathbb{Z}$ we have

$$J^{4k+\ell} = \left(J^4\right)^k J^\ell = I^k J^\ell = J^\ell.$$

Now $503 = 4 \cdot 125 + 3$ and therefore

$$J^{503} = J^3 = -J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

(h) We have

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}. = a \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = a\, I + b\, J.$$

This representation is unique because

$$a_1\, I + b_1\, J = a_2\, I + b_2\, J \implies \begin{pmatrix} a_1 & -b_1 \\ b_1 & a_1 \end{pmatrix} = \begin{pmatrix} a_2 & -b_2 \\ b_2 & a_2 \end{pmatrix},$$

and so $a_1 = a_2$ and $b_1 = b_2$.

(i) From the calculation in Part (d) we have

$$(a\, I + b\, J)^{-1} = \frac{a}{a^2 + b^2}\, I - \frac{b}{a^2 + b^2}\, J.$$

(j) Following the hint assume that

$$B = \begin{pmatrix} x & -y \\ y & x \end{pmatrix} \in \mathbf{C}$$

satisfies $B^2 = A$.

Now

$$B^2 = \begin{pmatrix} x^2 - y^2 & -2\,x\,y \\ 2\,x\,y & x^2 - y^2 \end{pmatrix}$$

and so if $B^2 = A$ we have the system[4]

$$\begin{cases} x^2 - y^2 & = a \\ 2\,x\,y & = b \end{cases} . \tag{E.4}$$

Now we consider two cases:

- **Case I:** $b \neq 0$. Then the second equation gives that $x \neq 0$ and $y \neq 0$ and

$$y = \frac{b}{2\,x}. \tag{E.5}$$

Substituting in the first equation then gives

$$x^2 - \frac{b^2}{4\,x^2} = a \iff 4\,x^4 - 4\,a\,x^2 - b^2 = 0.$$

Using the quadratic formula we have

$$x^2 = \frac{4\,a \pm \sqrt{16\,a^2 + 16\,b^2}}{8} = \frac{a \pm \sqrt{a^2 + b^2}}{2}$$

and since $a^2 + b^2 > a^2$ we have $\sqrt{a^2 + b^2} > a$ and therefore the

$$\frac{a - \sqrt{a^2 + b^2}}{2} < 0.$$

[4] Notice that this is not a linear system.

Since x is a real number, $x^2 \geq 0$ and therefore only the solution

$$x^2 = \frac{a + \sqrt{a^2+b^2}}{2}$$

is acceptable. Thus we have

$$x = \pm\sqrt{\frac{a + \sqrt{a^2+b^2}}{2}}.$$

Substituting into Equation (E.5) we get

$$y = \pm\frac{b}{\sqrt{a+\sqrt{a^2+b^2}}}.$$

Thus the System (E.4) has two solutions and therefore A has two square roots.

REMARK 26. Alternatively, we could substitute x^2 into the first equation of the System (E.4) to get

$$\frac{a+\sqrt{a^2+b^2}}{2} - y^2 = a \implies y^2 = \frac{-a+\sqrt{a^2+b^2}}{2} \implies y = \pm\sqrt{\frac{-a+\sqrt{a^2+b^2}}{2}}.$$

The two square roots of A are then

$$\sqrt{A} = \pm\frac{\sqrt{2}}{2}\begin{pmatrix} \sqrt{a+\sqrt{a^2+b^2}} & -\sqrt{-a+\sqrt{a^2+b^2}} \\ \sqrt{-a+\sqrt{a^2+b^2}} & \sqrt{a+\sqrt{a^2+b^2}} \end{pmatrix}.$$

- **Case II:** $b = 0$. Then from the second equation we get $x = 0$ or $y = 0$. Since $A \neq O$ we have $a \neq 0$ and from the first equation of System (E.4) we have that exactly one of the x, y is 0: If $a > 0$ then $y = 0$ and $x \neq 0$ and in that case we get solutions

  $$x = \pm\sqrt{a}, \quad y = 0.$$

  If $a < 0$ then $x = 0$ and $y \neq 0$ and in that case we get solutions

  $$x = 0, \quad y = \pm\sqrt{-a}.$$

REMARK 27. When $b = 0$, $A = a\,I$ a scalar matrix. Our conclusion is that for $a > 0$ the scalar matrix has two square roots $\pm\sqrt{a}\,I$, while if $a < 0$ then we have the square roots $\pm\sqrt{a}\,J$.

(k) For $Z \in \mathbf{C}$ let us denote by $\pm\sqrt{Z}$ the two square roots of Z whose existence was proven in Part (j). Then I claim that the equation

$$A\,X^2 + B\,X + C = O$$

has the solutions

$$X = \frac{1}{2}\left(-B \pm \sqrt{B^2 - 4\,A\,C}\right) A^{-1}.$$

Indeed, since by Part (e) any two elements of $\mathbf{C}$ commute we have

$$X^2 = \frac{1}{4}\left(B^2 \mp 2\,B\sqrt{B^2-4\,A\,C} + B^2 - 4\,A\,C\right) A^{-2}.$$

Now we calculate:

$$\begin{aligned}
A\,X^2 + B\,X &= A\left(\frac{1}{4}\left(B^2 \mp 2\,B\,\sqrt{B^2 - 4\,A\,C} + B^2 - 4\,A\,C\right) A^{-2}\right) \\
&\quad + B\left(\frac{1}{2}\left(-B \pm \sqrt{B^2 - 4\,A\,C}\right) A^{-1}\right) \\
&= \frac{1}{2}\,A^{-1}\left(B^2 \mp B\,\sqrt{B^2 - 4\,A\,C} - 2\,A\,C\right) \\
&\quad + \frac{1}{2}\,A^{-1}\left(B^2 \pm B\,\sqrt{B^2 - 4\,A\,C}\right) \\
&= \frac{1}{2} A^{-1}\,(-2\,A\,C) \\
&= -C.
\end{aligned}$$

□

Exercise E.7 For each of the following *permutation matrices* P:

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}, \quad \begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}, \quad \begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}, \quad \begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}, \quad \begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}, \quad \begin{pmatrix}0&0&1\\0&1&0\\1&0&0\end{pmatrix}.$$

Compute $P\,A$ and $A\,P$, where

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

ANSWER. The first matrix is the identity matrix and so $P\,A = A\,P = A$.
If

$$P = \begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}$$

then

$$P\,A = \begin{pmatrix}a_{11}&a_{12}&a_{13}\\a_{31}&a_{32}&a_{33}\\a_{21}&a_{22}&a_{23}\end{pmatrix}, \quad A\,P = \begin{pmatrix}a_{11}&a_{13}&a_{12}\\a_{21}&a_{23}&a_{22}\\a_{31}&a_{33}&a_{32}\end{pmatrix}.$$

If

$$P = \begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}$$

then

$$P\,A = \begin{pmatrix}a_{21}&a_{22}&a_{23}\\a_{11}&a_{12}&a_{13}\\a_{31}&a_{32}&a_{33}\end{pmatrix}, \quad A\,P = \begin{pmatrix}a_{12}&a_{11}&a_{13}\\a_{22}&a_{21}&a_{23}\\a_{32}&a_{31}&a_{33}\end{pmatrix}.$$

If

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

then

$$PA = \begin{pmatrix} a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}, \quad AP = \begin{pmatrix} a_{12} & a_{13} & a_{11} \\ a_{22} & a_{23} & a_{21} \\ a_{32} & a_{33} & a_{31} \end{pmatrix}.$$

If

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

then

$$PA = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \end{pmatrix}, \quad AP = \begin{pmatrix} a_{13} & a_{12} & a_{11} \\ a_{23} & a_{22} & a_{21} \\ a_{33} & a_{32} & a_{31} \end{pmatrix}.$$

Finally, if

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

then

$$PA = \begin{pmatrix} a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \end{pmatrix}, \quad AP = \begin{pmatrix} a_{13} & a_{12} & a_{11} \\ a_{23} & a_{22} & a_{21} \\ a_{33} & a_{32} & a_{31} \end{pmatrix}.$$

□

Exercise E.8 Let

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and let A be as in the previous question. Compute XA, AX, YA, and AY.

ANSWER. We have

$$XA = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ -2a_{21} & -2a_{22} & -2a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad AX = \begin{pmatrix} a_{11} & -2a_{12} & a_{13} \\ a_{21} & -2a_{22} & a_{23} \\ a_{31} & -2a_{32} & a_{33} \end{pmatrix}.$$

And,

$$YA = \begin{pmatrix} a_{11} + 3a_{31} & a_{12} + 3a_{32} & a_{13} + 3a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad AY = \begin{pmatrix} a_{11} & a_{12} & a_{11} + 3a_{13} \\ a_{21} & a_{22} & a_{23} + 3a_{21} \\ a_{31} & a_{32} & a_{33} + 3a_{31} \end{pmatrix}.$$

□

Exercise E.9 Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

(a) Compute A^n for $n = 0, 1, 2, 3, 4$.
(b) what pattern do you observe? Conjecture a formula for A^n based on that pattern.
(c) Prove your conjecture.

ANSWER. (a) We have

$$A^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A^1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad A^2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad A^3 = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}, \quad A^4 = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}.$$

(b) We conjecture that

$$A^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}.$$

(c) We proceed by induction. For $n = 0$ our conjecture is true. Assume then that

$$A^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}.$$

Then,

$$A^{n+1} = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + n \cdot 0 & 1 \cdot 1 + n \cdot 1 \\ 0 \cdot 1 + 1 \cdot 0 & 0 \cdot 1 + 1 \cdot 1 \end{pmatrix} = \begin{pmatrix} 1 & n+1 \\ 0 & 1 \end{pmatrix},$$

and we established that our conjecture holds for $n + 1$ as well. By induction then, our conjecture is true for all $n \in \mathbb{N}$.

□

Exercise E.10 Let

$$A = \begin{pmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 & 3 \\ 5 & 2 & 6 \\ -2 & -1 & -3 \end{pmatrix}.$$

(a) Find A^{42}.

ANSWER. We calculate:

$$A^2 = \begin{pmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{pmatrix} = A.$$

It follows that $A^n = A$ for all $n \geq 1$. Therefore

$$A^{42} = A = \begin{pmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{pmatrix}.$$

□

(b) Find B^{101}.

ANSWER. We calculate:

$$B^2 = \begin{pmatrix} 0 & 0 & 0 \\ 3 & 3 & 9 \\ -1 & -1 & -3 \end{pmatrix}, \quad B^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = O.$$

It follows that $B^n = O$ for $n \geq 3$. Therefore

$$B^{101} = O.$$

□

Exercise E.11 Let $p(x) = x^3 - 3x^2 + x - 3$ and let

$$A = \begin{pmatrix} 5 & 0 & 13 \\ 1 & 3 & 14 \\ -2 & 0 & -5 \end{pmatrix}.$$

Evaluate $p(A)$.

ANSWER. We have

$$p(A) = \begin{pmatrix} -9 & 2 & 1 \\ 7 & -12 & -2 \\ 5 & -5 & -8 \end{pmatrix}.$$

□

Exercise E.12 Let $A = \begin{pmatrix} 5 & 2 \\ 0 & a \end{pmatrix}$. Find the real number a if A is a root of the polynomial $p(x) = x^2 - 7x + 10$.

ANSWER. We have

$$A^2 = \begin{pmatrix} 25 & 2a+10 \\ 0 & a^2 \end{pmatrix},$$

and so

$$\begin{aligned} p(A) &= A^2 - 7A + 10I \\ &= \begin{pmatrix} 25 & 2a+10 \\ 0 & a^2 \end{pmatrix} - 7\begin{pmatrix} 5 & 2 \\ 0 & a \end{pmatrix} + 10\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 25 & 2a+10 \\ 0 & a^2 \end{pmatrix} + \begin{pmatrix} -35 & -14 \\ 0 & -7a \end{pmatrix} + \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 2a-4 \\ 0 & a^2-7a+10 \end{pmatrix}. \end{aligned}$$

So if A is a root of $p(x)$ we have

$$\begin{pmatrix} 0 & 2a-4 \\ 0 & a^2-7a+10 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

So we must have

$$2a - 4 = 0, \text{ and } a^2 - 7a + 10 = 0.$$

Thus $a = 2$. □

Exercise E.13 Let

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

and let $p(x) = x^3 - 2x^2 - 2x + 6$.

(a) Verify that A is a root of $p(x)$.
(b) Express A^{-1} as a polynomial in A.
(c) Use Part (b) to find A^{-1}.

ANSWER. (a) We calculate

$$A^2 = \begin{pmatrix} 3 & 1 & -1 \\ -1 & 3 & -1 \\ 4 & 2 & 2 \end{pmatrix}, \quad A^3 = \begin{pmatrix} 2 & 4 & -2 \\ 2 & 0 & -4 \\ 14 & 2 & 0 \end{pmatrix}.$$

And then with straightforward calculations we verify that

$$A^3 - 2A^2 - 2A + 6I = O.$$

(b) We have

$$\begin{aligned} A^3 - 2A^2 - 2A + 6I = O &\implies A^3 - 2A^2 - 2A = -6I \\ &\implies A\left(A^2 - 2A - 2I\right) = -6I \\ &\implies A\left(-\frac{1}{6}\left(A^2 - 2A - 2I\right)\right) = I. \end{aligned}$$

Therefore,

$$A^{-1} = -\frac{1}{6}\left(A^2 - 2\,A - 2\,I\right).$$

(c) We have

$$\begin{aligned} A^{-1} &= -\frac{1}{6}\left(A^2 - 2\,A - 2\,I\right) \\ &= -\frac{1}{6}\left(\begin{pmatrix} 3 & 1 & -1 \\ -1 & 3 & -1 \\ 4 & 2 & 2 \end{pmatrix} - 2\begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & -1 \\ 3 & -1 & 1 \end{pmatrix} - 2\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) \\ &= -\frac{1}{6}\begin{pmatrix} -1 & -1 & -1 \\ -5 & 1 & 1 \\ -2 & 2 & -2 \end{pmatrix} \\ &= \frac{1}{6}\begin{pmatrix} 1 & 1 & 1 \\ 5 & -1 & -1 \\ 2 & -2 & 2 \end{pmatrix}. \end{aligned}$$

□

Exercise E.14 Find all 2×2 matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ that commute with $B = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$.

SOLUTION. We have

$$A\,B = \begin{pmatrix} a & 2\,a + b \\ c & 2\,c + d \end{pmatrix}, \quad B\,A = \begin{pmatrix} a + 2\,c & b + 2\,d \\ c & d \end{pmatrix}.$$

So we have the system

$$\begin{cases} a & = a + 2c \\ 2a + b & = b + 2d \\ \quad c & = c \\ \quad 2c + d & = d \end{cases}.$$

The first equation gives $c = 0$ and the second $a = d$. Thus the centralizer of B is

$$\left\{\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} : a, b \in \mathbb{R}\right\}.$$

□

Exercise E.15 Let A and B be two symmetric $n \times n$ matrices. Prove that $A\,B$ is symmetric if and only if A and B commute.

SOLUTION. Since A and B are symmetric we have $A^* = A$ and $B^* = B$ and so

$$(A\,B)^* = B^*\,A^* = B\,A.$$

Thus

$$(A\,B)^* = A\,B \iff B\,A = A\,B.$$

In words, $A\,B$ is symmetric if and only if A and B commute. □

Exercise E.16 Let $A \in \mathrm{M}_n$. Prove that $A + A^*$ is symmetric, where A^* is the transpose of A.

SOLUTION. We have

$$(A + A^*)^* = A^* + (A^*)^* = A^* + A = A + A^*.$$

□

Exercise E.17 A square matrix A is called *nilpotent* if $A^k = O$ for some positive integer k. Prove that if A is nilpotent then A is not invertible.

SOLUTION. Assume,to get a contradiction, that A is invertible. Then A^k is also invertible and $= \left(A^{-1}\right)^k$. But then

$$\begin{aligned} A^k = O &\implies A^k \left(A^k\right)^{-1} = O \left(A^k\right)^{-1} \\ &\implies I = O, \end{aligned}$$

a contradiction. Therefore A is not invertible. □

Exercise E.18 A square matrix A is called *idempotent* if $A^2 = A$. Find all the matrices that are both idempotent and invertible.

SOLUTION. If we multiply both sides of the equation

$$A^2 = A$$

with A^{-1} we get

$$A = I.$$

Thus the only idempotent and invertible matrix is the identity matrix I. □

Exercise E.19 A square matrix is said to be *antisymmetric* if $A^* = -A$, in other words if for all i, j we have

$$a_{ji} = -a_{ij}.$$

Prove that if A and B are symmetric matrices then $A\,B - B\,A$ is antisymmetric.

SOLUTION. We have

$$\left(A\,B - B\,A\right)^* = \left(A\,B\right)^* - \left(B\,A\right)^* = B^*\,A^* - A^*\,B^*.$$

Now if A and B are symmetric we have

$$B^*\,A^* - A^*\,B^* = B\,A - A\,B = -\left(A\,B - B\,A\right).$$

So if A and B are symmetric matrices then

$$\left(A\,B - B\,A\right)^* = -\left(A\,B - B\,A\right),$$

and so $A\,B - B\,A$ is antisymmetric. □

Exercise E.20 Prove that all permutation matrices in Question E.7 are orthogonal.

SOLUTION. Let P be any permutation matrix. Then the columns of P are obtained from the standard basis of $\mathbb{R}^3$ by applying a permutation. Now for the standard basis of $\mathbb{R}^3$ we have

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}, \quad i, j = 1, 2, 3.$$

Therefore, if $\mathbf{p}_i$, $i = 1, 2, 3$ are the columns of P we have

$$\mathbf{p}_i \cdot \mathbf{p}_j = \delta_{ij}, \quad i, j = 1, 2, 3.$$

By Proposition (8) we have then that P is orthogonal. □

E.5. Homework 5

Exercise E.1 Let

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{pmatrix}.$$

(a) Find a matrix A such that

$$A, X = \begin{pmatrix} 3\,x_{41} - 2\,x_{21} & 3\,x_{42} - 2\,x_{22} & 3\,x_{43} - 2\,x_{23} & 3\,x_{44} - 2\,x_{41} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \end{pmatrix}.$$

(b) Find a matrix B such that

$$X\,B = \begin{pmatrix} x_{14} - 2\,x_{13} & x_{12} & x_{11} & x_{14} \\ x_{24} - 2\,x_{23} & x_{22} & x_{21} & x_{24} \\ x_{34} - 2\,x_{33} & x_{32} & x_{31} & x_{34} \\ x_{44} - 2\,x_{43} & x_{42} & x_{41} & x_{44} \end{pmatrix}.$$

Exercise E.2 Let

$$A = \begin{pmatrix} 2 & -1 & 1 \\ 6 & -3 & 4 \\ 3 & -2 & 3 \end{pmatrix}.$$

(a) Verify that A is a root of the polynomial

$$p(x) = x^2 - 3\,x + 2.$$

(b) Find A^{-1}.

Exercise E.3 Let

$$\mathbb{Q}(\sqrt{3}) = \left\{ a + b\sqrt{3} : a, b \in \mathbb{Q} \right\}.$$

(a) Prove that $\mathbb{Q}(\sqrt{3})$ is a subfield of the field of real numbers R.

(b) Give an explicit formula for $(a + b\sqrt{3})^{-1}$.

Exercise E.4 Consider the set $\mathbb{F} = \{0, 1, a, b\}$ where $a \neq b$. Define addition and multiplication via the following tables

$+$	0	1	a	b
0	0	1	a	b
1	1	2	b	a
a	a	b	0	1
b	b	a	1	0

$\cdot$	0	1	a	b
0	0	0	0	0
1	0	1	a	b
a	0	a	b	1
b	0	b	1	a

.

Prove that $\mathbb{F}$ is a field.

Exercise E.5 Consider $\mathbb{R}^2$ with the usual addition

$$(a, b) + (c, d) = (a + c, b + d),$$

and multiplication given by

$$(a, b)\,(c, d) = (a\,c, b\,d).$$

Is $\mathbb{R}^2$ with these operations a field? Fully justify your answer.

Exercise E.6 Consider the following matrices[5] with complex entries

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

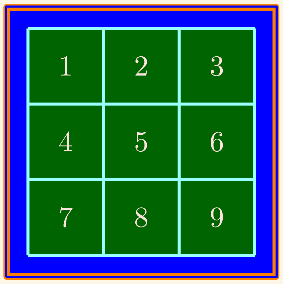
Prove that these matrices satisfy the following relations:

(a) $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = I$.

(b) $\sigma_x \sigma_y = i\,\sigma_z, \quad \sigma_y \sigma_z = i\,\sigma_z, \quad \sigma_z \sigma_x = i\,\sigma_y$.

(c) $\sigma_x \sigma_y = -\sigma_y \sigma_x, \quad \sigma_x \sigma_z = -\sigma_z \sigma_x, \quad \sigma_y \sigma_z = -\sigma_z \sigma_y$.

Exercise E.7 Consider a 3×3 grid of squares, each either green or red. When we touch a square its color and the color of its neighbors change, where the neighbors of a square are all squares that share an edge with it.



Thus for example, if we touch the square numbered 1 the squares numbered 1, 2, and 4 change color, if we touch square 5 then all squares except 1, 3, 7, and 9 change color, and if we touch 8 then 5, 7, 8, and 9 change colors.

We start with all squares green. Find, if possible, a sequence of squares to touch so that all squares turn red.

Exercise E.8 Consider the following vectors in $\mathbb{C}^4$:

$$\mathbf{v}_1 = (1, i, 0, -i) \qquad \mathbf{v}_2 = (2+i, 3i, i, 1-4i)$$
$$\mathbf{v}_3 = (5+i, 2+6i, 1+2i, 7-9i) \qquad \mathbf{v}_4 = (0, 3-i, 1+1, 0).$$

Find a basis and state the dimension of the linear span $\mathbb{C}\langle \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4 \rangle$.

Exercise E.9 Consider $\mathbb{R}$ as a vector space over $\mathbb{Q}$. Prove that

$$\left\{\sqrt{2}, \sqrt{3}, \sqrt{5}\right\}$$

is linearly independent. You may consider Item (a) of Example 93 in the notes known.

Exercise E.10 Let $\mathbf{S}_n$ denote the set of $n \times n$ symmetric matrices over $\mathbb{R}$ (see Definition 29 in the notes).

(a) Prove that $\mathbf{S}_n$ is a vector subspace of $\mathbf{M}_n$.

(b) Find a basis and the dimension of $\mathbf{S}_n$.

Exercise E.11 Consider the vector space $\mathbb{R}[x]$ of polynomials with real coefficients. Which of the following subsets is a vector subspace of $\mathbb{R}[x]$?

(a) $V = \{p(x) \in \mathbb{R}[x] : p(42) = 0\}$.

(b) $U = \{p(x) \in \mathbb{R}[x] : p(42) \geq 0\}$.

(c) $W = \{p(x) \in \mathbb{R}[x] : p(42) = p(0)\}$.

(d) $X = \{p(x) \in \mathbb{R}[x] : \deg p(x) = 8\}$.

Fully justify your answers.

Exercise E.12 Let $\mathbf{P}_3$ be the set of real polynomials of degree at most 3:

$$\mathbf{P}_3 = \{p(x) \in \mathbb{R}[x] : \deg p(x) \leq 3\}.$$

[5] These matrices are called *Pauli spin matrices*. They are used in Quantum Mechanics to compute the spin of an electron.

Prove that

$$B = \left\{1, x-1, (x-1)^2, (x-1)^3\right\},$$

is a basis of $\mathbf{P}_3$.

Exercise E.13 Let $S = \{A, B, C, D\} \subseteq \mathbf{M}_2$ where

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

(a) Prove that S is a basis of $\mathbf{M}_2$.

(b) Express $X = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ as a linear combination of elements of S.

Exercise E.14 Find conditions on the complex number z so that the vectors

$$\mathbf{v}_1 = (z, 0, 1), \quad \mathbf{v}_1 = (0, 1, z^3), \quad \mathbf{v}_3 = (z, 1, 1+z)$$

form a basis of $\mathbb{C}^3$.

Exercise E.15 Consider the vector space M_n of real $n \times n$ matrices, and let B be a basis of M_n. Prove that

$$B^* = \{X^* : X \in B\}$$

is also a basis of M_n, where X^* stands for the transpose of a matrix X.