MTH 42 NOTES

NIKOS APOSTOLAKIS

Abstract. Notes for Linear Algebra, Fall 2024.

Contents

1. Linear systems

Preliminaries. 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.

Date: Fall 2024.

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* 1.1.** 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.

Definition 1.2. 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.

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

$$ax = c. \tag{1}$$

We have two cases:

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

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

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

(2) **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:

[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.

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.

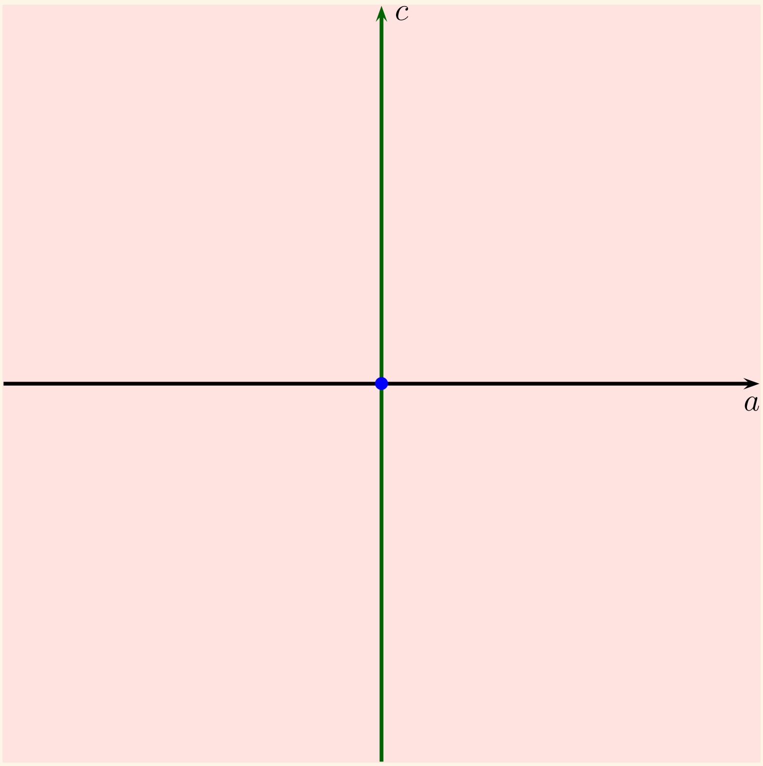

FIGURE 1. The space of linear equations with one variable.

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

$$a\,x + b\,y = c, \tag{2}$$

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

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

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

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

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

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

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

$$\begin{cases} x &= t \\ y &= \frac{2}{3}t \end{cases} \qquad t \in \mathbb{R}. \tag{4}$$

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

$$\begin{pmatrix} x \\ y \end{pmatrix} = t \begin{pmatrix} 1 \\ \frac{2}{3} \end{pmatrix}. \tag{5}$$

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 (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 (5) is therefore just a rewriting of the system of equations (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,

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

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

S is a Vector Space.

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

Theorem 1.3. *The solution set of any* homogeneous *linear equation is closed under scalar multiplication and vector addition.*

Proof. Let

$$a_1x_1 + \ldots + a_nx_n = 0 \tag{6}$$

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

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

Therefore $(v_1 + w_1, \ldots, v_n + w_n)$ is a solution of (6).

Now let t be any number, then

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

therefore $s(v_1, \ldots, v_n)$ is a solution of (6). □

Consider now the equation

$$2x - 3y = 6. \tag{7}$$

Notice that this equation has the same coefficients as Equation (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

$$\begin{cases} x &= t \\ y &= \dfrac{2}{3}t + 2 \end{cases} \qquad t \in \mathbb{R}, \tag{8}$$

and in vector notation:

$$\begin{pmatrix} x \\ y \end{pmatrix} = t\begin{pmatrix} 1 \\ \frac{2}{3} \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix}. \tag{9}$$

Equations (5) and (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.

We see then that $(0, 2)$ is the y–intercept of the line with equation (7). That's where $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$ came from. Geometrically, Equation (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 1.3.2.

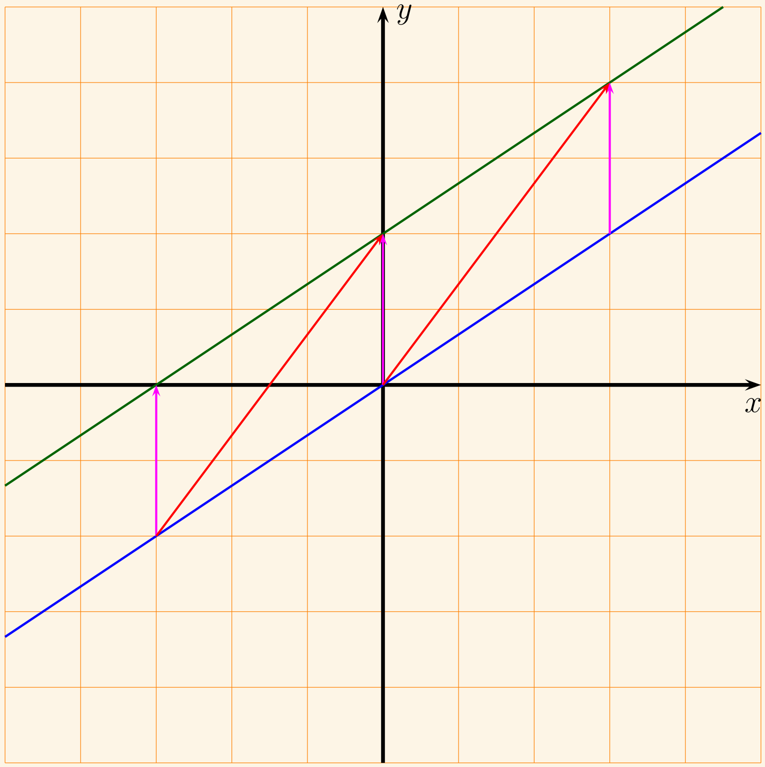

FIGURE 2. The solution sets of Equations (3) and (7).

In general if at least one of the coefficients in non-zero the equation (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

$$\begin{pmatrix} x \\ y \end{pmatrix} = t \begin{pmatrix} \frac{b}{a} \\ 1 \end{pmatrix} + \begin{pmatrix} \frac{c}{a} \\ 0 \end{pmatrix}. \tag{10}$$

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

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

Solving for z we get

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

The solution set

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

has now two free parameters

Using vector notation we have

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

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 $\mathbb{R}^n$.

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. **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.

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

Elementary row operations

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

Theorem 1.4. *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* 1.5 (An often used combination).** In practice we often perform the following combination of the second and third operation:

(1) multiply E_k by a non-zero scalar λ_k,
(2) multiply E_ℓ by a non-zero scalar λ_ℓ,
(3) replace E_k with $E'_k + E'_\ell$,
(4) 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 1.6. 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:

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

Divide the first equation by 7:

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

Now divide the second equation by 27:

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

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

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

Finally we divide the first equation by 2:

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

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 1.6 contains all the ingredients for an algorithm that solves all linear systems.

Example 1.7. 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:

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

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

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

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

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

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

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

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 1.8. 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.

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

Solving the second and third equations for y we get

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

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 1.9 (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

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

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:

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

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

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

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

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

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 1.10 (Determining a quadratic polynomial by three values). For the polynomial $p(x) = ax^2 + bx + c$ 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 1.11. Let's again consider a quadratic binomial $p(x) = ax^2 + bx + c$, 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* 1.12.** 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 1.13. 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

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

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

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

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

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 1.14. Let's solve to solve the system

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

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

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

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

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

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 1.15. 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.

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

(Reduced) Echelon form

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

(1) The zero rows are at the bottom of the matrix.
(2) All the entries below the leading entry of a non-zero row are 0.
(3) 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:

(3) All leading entries are equal to 1.
(4) 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 1.16. 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

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

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:

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

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

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

We finally solve the first equation to get

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

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 1.17. 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

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

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:

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

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 1.18. 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 1.19. *Row equivalence is an equivalence relation. In other words, it enjoys the following properties:*

(1) *It is* reflexive. *This means that every matrix is row equivalent to itself:*
(2) *It is* symmetric. *This means that if A is equivalent to b then B is also equivalent to A:*

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

(3) *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 1.20 (Kill below first, then kill above).

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

(2) *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.

(1) 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.

(2) 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.

(1) If we scan a row and find no non-zero entries, we just discard that row.
(2) If the leading entry is in the last column we stop the procedure and declare that the solution set is $\varnothing$.

***Remark* 1.21.** 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.

Example 1.22. 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.

[6] Or at least we think so

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

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 $c_1, c_2, \ldots, c_n = 0$.

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

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

Notice that $x_1 = x_2 = \cdots = x_n = 0$ is a solution of (12). Therefore homogeneous systems are always consistent, the interesting question then is whether there are other solution besides that obvious one.

Definition 1.23 (Trivial solution of a homogeneous system.). The solution

$$x_1 = 0, \ldots, x_n = 0$$

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* 1.24.** 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.

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

Example 1.25. Consider the homogeneous system:

$$\begin{cases} x_1 + x_2 + x_3 - 2x_4 = 0 \\ 2x_1 - 2x_3 = 0 \\ x_2 + x_3 + 4x_4 = 0 \end{cases}.$$

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

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

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

[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.

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.

Theorem 1.26. *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

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

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 1.26.

Corollary 1.27. *We have:*

(1) Two reduced echelon matrices are row equivalent if and only if they are equal.
(2) The reduced echelon form of any matrix is unique.
(3) 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]

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

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

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

More generally, the number of basic variables, is always equal to the number of non-zero rows of A. In Example 1.25 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 1.28 (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.*

1.3.1. ***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 1.29 (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 1.30 (Vector subspace). A subset $V \subseteq \mathbb{R}^n$ is called a *vector subspace* if the following three conditions hold:

(1) V contains the zero vector, that is $\mathbf{0} \in V$.
(2) V is closed under vector addition, that is

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

(3) 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 1.31 (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*

(1) $A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y}$.
(2) $A(\lambda\mathbf{x}) = \lambda (A\mathbf{x})$.

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

***Remark* 1.33.** We will see later in the course that every vector subspace of $\mathbb{R}$ is the solution set of some homogeneous linear system.

1.3.2. **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,

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".

[11]Do this.

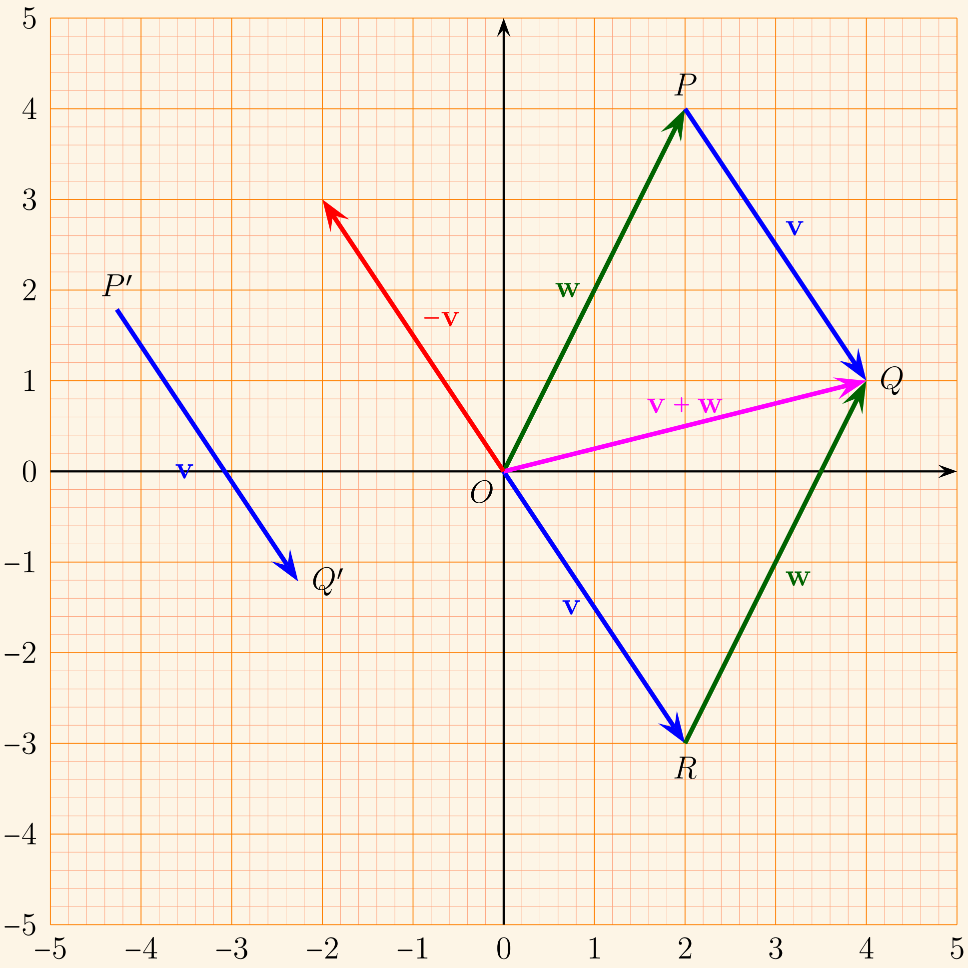

FIGURE 3. Points and vectors in $\mathbb{R}^2$.

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 1.34 (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* 1.35.** 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 1.31. The second is just a reformulation of the first. To prove the third item use the first item and the equation $\mathbf{a} = \mathbf{a}_0 + (\mathbf{a} - \mathbf{a}_0)$. □

[12] Fill the details.

Example 1.36. Consider the system

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

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 1.37 (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 $c\neq 0$.
- *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*

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

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

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. ☐

2. Exercises

(1) 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)

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

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

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

(3) 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.

(4) 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.

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

Hint. Work as in Example 1.9. Show that the system we obtain has non-trivial solutions and all the non-trivial equations differ by a multiplicative constant.

(6) 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 $p(-2) = -21$.

(7) Look at Examples 1.10 and 1.11, there is a geometric reason why in Example refexm:qua2 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 1.11 are *colinear*[15] 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

(8) **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?

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

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 $D = a_1 b_2 - a_2 b_1$[16], and consider two cases: whether D is zero or not.

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

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

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

$$\frac{c_1}{a_1} - \frac{b_1}{a_1} \cdot \frac{a_1 c_2 - a_2 c_1}{D} = \frac{c_1 (a_1 b_2 - a_2 b_1) - b_1 (a_1 c_2 - a_2 c_1)}{a_1 D}.$$

If we set $D_x = c_1 b_2 - c_2 b_1$ and $D_y = a_1 c_2 - a_2 c_1$ 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.

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

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.

3.0.1. *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 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

$$a_1b_2 = a_2b_1. \tag{13}$$

- **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 (3.0.1) 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 (3.0.1) is 0. Since $a_1 \neq 0$ Equation (3.0.1) 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 (3.0.1) 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 (3.0.1) 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 $(1, 1)$.

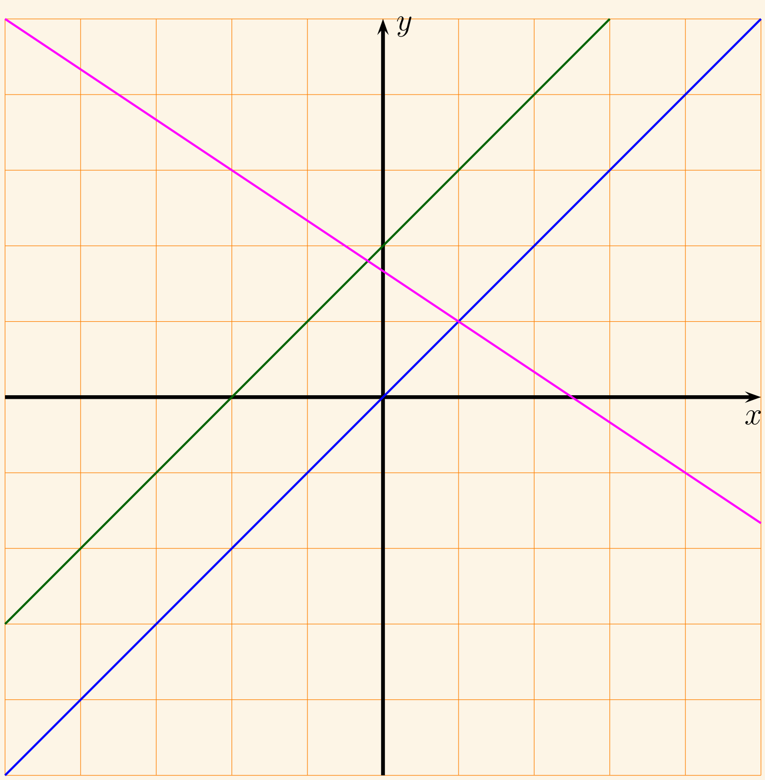

Figure 4. Parallel and intersecting lines.

3.0.2. **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 $\mathbb{R}^n$.

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 3.2 (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.

(1) 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

$$x\,\mathbf{v} + y\,\mathbf{w} = \mathbf{c}. \tag{14}$$

(2) 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 (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 (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.

(1) 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.
(2) 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 $\langle \mathbf{a}, \mathbf{b} \rangle$[17]. $\mathbf{a}$ (or) by itself forms a basis however[18].

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 two non-colinear vectors span $\mathbb{R}^2$ is the same as in the case of the standard basis used in the familiar Cartesian coordinate system.

Let $\mathbf{v}$ be an arbitrary vector. If we take the starting point of $\mathbf{v}$ to be the origin $O(0, 0)$ and if its endpoint is $P(p_1, p_2)$ then draw a line from P, parallel to the y-axis it will intersect the x-axis at a point with coordinates $(p_1, 0)$, and a line parallel to the x-axis intersects the y-axis at a point with coordinates $(0, p_2)$. Then

$$\mathbf{v} = p_1 \mathbf{e}_1 + p_2 \mathbf{e}_2.$$

See the left hand side of Figure 5.

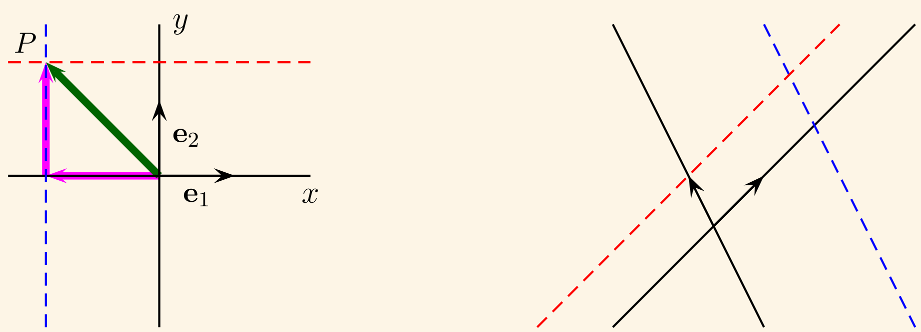

FIGURE 5. Why coordinates work.

[17]Why?

[18]Why?

The same method will work for any two non-parallel lines. Take any two intersecting lines ℓ_1 and ℓ_2, call their intersection O and choose a vector $\mathbf{a}$ in the first and a vector $\mathbf{b}$ in the second. Any vector $\mathbf{v} = \overrightarrow{OP}$ can be written

$$\mathbf{v} = p_1\mathbf{a} + p_2\mathbf{b},$$

because the line from P parallel to ℓ_2 will intersect ℓ_1 and their intersection will determine a vector that is a multiple of $\mathbf{a}$. Similarly the intersection of a line parallel to ℓ_2 will determine a vector that is a multiple of $\mathbf{b}$.

4. 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[19]. 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

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

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

Definition 4.1 (Vector addition and scalar multiplication). Let $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ be two n-vectors and λ a real number. We define

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

and

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

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

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 4.2 (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,

$$\mathbf{a} = a_1\mathbf{e}_1 + \dots + a_n\mathbf{e}_n. \tag{15}$$

Theorem 4.3 (Vector Space Axioms). *The operations of vector addition and scalar multiplication enjoy the following properties:*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(8) 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

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

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 4.4 (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 1.30 in Section 1.3.1), *linear combination*, *linear span*, and *basis* (Section 3.0.2).

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 4.5 (**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 1.30 we have $\lambda\,\mathbf{a}\in V$ and $\mu\,\mathbf{b}\in V$ and therefore by the second property in Definition 1.30 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 $\mathbf{a} \in V$[20], then by the second property we have

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

Condition (2) of Definition 1.30 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 4.6. *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 4.7 (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 4.5.

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

Proof. **Sketch**[21] 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 1.30 are satisfied because:

(1) $\mathbf{0} = 0\,\mathbf{a}$ for any $\mathbf{a} \in S$.
(2) The sum of two sums of multiples of elements S is obviously also a sum of multiples of elements of S.
(3) 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 4.9 (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

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

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

[21] Fill the details.

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 4.10. 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 $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. 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

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

and by Theorem 1.37 we conclude that the system has a unique solution for all $\mathbf{c}$.

Example 4.11. 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 4.12. *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 3.0.2 we show that if two vectors are colinear then they don't form a basis.

***Question* 4.13.** 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 4.14. 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

$$c_1 - 2\,c_2 + c_3 = 0 \iff c_3 = -c_1 + 2\,c_2. \tag{16}$$

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 (16) 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

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

5. Linear dependence, Dimension

Let's take a closer look at Example 4.14. 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 5.1. 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 4.14 the vector $\mathbf{a} = (0, 1, 2)$ is equal to two different linear combinations of vectors of S, namely

$$\mathbf{a} = -2\,\mathbf{v} + \mathbf{u} \text{ and } \mathbf{a} = \mathbf{v} + \mathbf{w}. \tag{17}$$

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 1.37 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 (17).

Claim 5.2. *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

$$x_1\,\mathbf{v} + y_1\,\mathbf{w} = x_2\,\mathbf{v} + y_2\,\mathbf{w} \tag{18}$$

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 (18) 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. 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:

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

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 $\mathbf{0}$, 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 5.3. *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 5.4 (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 5.5. *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 5.6. *The following hold.*

(1) If $\mathbf{0} \in S$ then S is linearly dependent.
(2) If $S = \{\mathbf{v}\}$ then S is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}$.
(3) 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.
(4) 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 $\mathbf{v}_2, \dots, \mathbf{v}_m$.
(5) If $S \subseteq S'$ and S is linearly dependent then S' is linearly dependent as well.
(6) If $S \subseteq S'$ and S' is linearly independent then S is linearly independent as well.

Proof. (1) We have $42\,\mathbf{0} = \mathbf{0}$ an expression of the zero vector as a non-trivial linear combination of vectors from S.

(2) 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.

(3) 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.

(4) 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'.

(5) This is the contra-positive of the previous item.

□

Theorem 5.7. *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 5.2. 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 5.8. *If $B = \{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ is a basis of V and $\mathbf{w}_1 \in V$ is such that*

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

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 $\mathbf{w}, \mathbf{v}_2, \dots, \mathbf{v}_n$:

$$\mathbf{v}_1 = \frac{1}{\lambda_1}\mathbf{w}_1 - \frac{\lambda_2}{\lambda_1}\mathbf{v}_2 - \dots - \frac{\lambda_d}{\lambda_1}\mathbf{v}_d. \tag{19}$$

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 (19) 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* 5.9.** Lemma 5.8 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 5.7. 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 5.8 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 5.8, 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 5.8,

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

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 5.8, 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 5.10 (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[22].

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 5.7, any subset S of B' with d elements would be 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 5.6.

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 5.11 (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*

$$B \subsetneq S \subseteq V \implies S \text{ is linearly dependent.} \tag{20}$$

Proof. Let $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_d\}$ 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 $\lambda_1, \ldots, \lambda_d$ 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 5.6 we conclude that S is linearly dependent.

Conversely, assume that (20) 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 $\lambda, \lambda_1, \ldots, \lambda_d \in \mathbb{R}$ such that

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

with $\lambda, \lambda_1, \ldots, \lambda_d$ 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.

[22] Why?

Theorem 5.12 (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 (20) 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 5.13. 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*.

5.1. 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 5.14. Which of the following subsets of $\mathbb{R}^4$ are vector subspaces?

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

Answer. (1) This set is a subspace. To prove this we will prove that the two conditions in Theorem 4.5 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

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

(2) V is not a vector subspace since $\mathbf{0} \notin V$.

(3) V is a vector subspace. We can proceed as in Item (1) and show that the two properties of Theorem 4.5 are satisfied[23]. An other method is to show that V is the linear span of a subset of $\mathbb{R}^4$. Then by Theorem 4.8 V is a subspace[24].

[23]Do this

[24]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.

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 4.8 that V is a subspace of $\mathbb{R}^4$.

(4) 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 5.15. Find a basis for each of the sets in Example 5.14 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 5.16. 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 5.17. Find a basis and the dimension of the linear span of the vectors

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

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 5.18. We use the same notation as in Example 5.17.

(1) Express each of the "free" vectors $\mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_6$ as a linear combination of the elements of B.
(2) 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. (1) 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 5.17, 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 5.17 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[25]

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & 3 & -24 \\ 0 & 1 & 0 & -2 & 2 & -7 \\ 0 & 0 & 1 & 0 & 0 & 4 \end{array}\right).$$

Therefore,

$$\begin{aligned} \mathbf{v}_3 &= -2\,\mathbf{v}_1 - 2\,\mathbf{v}_2 \\ \mathbf{v}_4 &= 3\,\mathbf{v}_1 + 2\,\mathbf{v}_2 \\ \mathbf{v}_6 &= -24\,\mathbf{v}_1 - 7\,\mathbf{v}_2 + 4\,\mathbf{v}_5. \end{aligned}$$

(2) 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 1.37 this happens if and only if the echelon form of its augmented matrix contains a row of the form

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

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:

[25] Do the calculations and verify this.

(a) Add -2 times the first row to the third.
(b) Add the first row to the fourth.
(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 $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$

(1) Create an $n \times k$ matrix A that has the given vectors as columns

$$A = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots \mathbf{v}_k \end{pmatrix}.$$

(2) Find an echelon form for A.
(3) A basis consists of the columns of A that correspond to the basic columns of the echelon form.

6. Matrices as transformations

We have already introduced the notation $A\mathbf{x}$ where A is an $m \times n$ matrix and$\mathbf{x}$ is an n-vector. We were writing a system of m equations with n variables as

$$A\mathbf{x} = \mathbf{c}, \tag{21}$$

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,

$$\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}. \tag{22}$$

Before proceeding, let's officially define the product of a matrix and a column vector.

Definition 6.1. 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 (22). 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* 6.2.** 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 6.3. 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 $\mathbf{r}_1, \ldots, \mathbf{r}_m$ are the rows of the matrix A then $A\mathbf{x}$ has rows $\mathbf{r}_1\,\mathbf{x}, \ldots, \mathbf{r}_m\,\mathbf{x}$. 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 6.4. Calculate the product $A\mathbf{x}$ if defined.

(1) $A = \begin{pmatrix} 1 & 2 & 0 \\ -2 & 5 & 1 \\ 0 & 6 & 1 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}$

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:

$$A\mathbf{x} = \begin{pmatrix} 1\cdot 3 + 2\cdot(-1) + 0\cdot 2 \\ -2\cdot 3 + 5\cdot(-1) + 1\cdot 2 \\ 0\cdot 3 + 6\cdot(-1) + 2\cdot 2 \end{pmatrix} = \begin{pmatrix} 1 \\ -9 \\ -2 \end{pmatrix}$$

□

(2) $A = \begin{pmatrix} -2 & 5 & 0 & -7 \\ 3 & 4 & -1 & 0 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 0 \\ \pi \\ -2 \end{pmatrix}$

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. □

(3) $A = \begin{pmatrix} -2 & 5 & 0 & -7 \\ 3 & 4 & -1 & 0 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 0 \\ \pi \\ -2 \\ \sqrt{3} \end{pmatrix}$

Answer. The dimensions now match and we have

$$\begin{pmatrix} -2 & 5 & 0 & -7 \\ 3 & 4 & -1 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ \pi \\ -2 \\ \sqrt{3} \end{pmatrix} = \begin{pmatrix} -2\cdot 0 + 5\pi + 0\cdot(-2) - 7\sqrt{3} \\ 3\cdot 0 + 4\pi + (-1)\cdot(-2) + 0\sqrt{3} \end{pmatrix} = \begin{pmatrix} 5\pi - 7\sqrt{3} \\ 2 + 4\pi \end{pmatrix}.$$

□

The product $A\mathbf{x}$ can be also calculated column by column. In Section 4, 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 $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

$A\mathbf{x}$ 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

$$\begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_m \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \left(x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n\right).$$

Example 6.5. 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 (21) 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

$$\mathbf{y} = A\mathbf{x}. \tag{23}$$

and consider $\mathbf{y}$ the *dependent* and $\mathbf{x}$ the *independent* variable.

If we expand Equation (23) 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 (23) as

$$\mathbf{y} = x_1\,\mathbf{a}_1 + \cdots + x_n\,\mathbf{a}_n. \tag{24}$$

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 6.6 (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 6.7 (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 6.8. 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 1.3.1 we proved (see Theorem 1.31) 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 1.31 says that the functions defined by matrices are linear.

Definition 6.9 (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.

(1) *It respects vector addition.* This means that for any two vectors $\mathbf{v}, \mathbf{w} \in \mathbf{R}^n$ we have

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

(2) *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 6.10 (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 6.9 are satisfied by I_n. Indeed,

(1) For $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ we have

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

(2) For $\lambda \in \mathbb{R}$ and $\mathbf{v} \in \mathbf{R}^n$ we have

$$I(\lambda) = \lambda \mathbf{v} = \lambda I \mathbf{v}.$$

Example 6.11 (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.

(1) 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 6.9. 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

$$T(\mathbf{v}+\mathbf{w}) = \begin{pmatrix} 3\left(v_1+w_1\right) - 2\left(v_2+w_2\right) \\ \left(v_1+w_1\right) - 2\left(v_2+w_2\right) + 3\left(v_2+w_2\right) \\ \left(v_2+w_2\right) + \left(v_3+w_3\right) \\ 2\left(v_1+w_1\right) + 3\left(v_2+w_2\right) - \left(v_3+w_3\right) \end{pmatrix}$$

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

$$T(\mathbf{v}) + T(\mathbf{w}) = \begin{pmatrix} 3\left(v_1+w_1\right) - 2\left(v_2+w_2\right) \\ \left(v_1+w_1\right) - 2\left(v_2+w_2\right) + 3\left(v_2+w_2\right) \\ \left(v_2+w_2\right) + \left(v_3+w_3\right) \\ 2\left(v_1+w_1\right) + 3\left(v_2+w_2\right) - \left(v_3+w_3\right) \end{pmatrix} = T(\mathbf{v}+\mathbf{w}).$$

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:

$$T(\lambda\mathbf{v}) = \begin{pmatrix} 3(\lambda v_1) - 2(\lambda v_2) \\ (\lambda v_1) - 2\left(\lambda v_2\right) + 3\left(\lambda v_3\right) \\ \left(\lambda v_2\right) + \left(\lambda v_3\right) \\ 2\left(\lambda v_1\right) + 3\left(\lambda v_2\right) - \left(\lambda v_3\right) \end{pmatrix} = \begin{pmatrix} \lambda\left(3v_1 - 2v_2\right) \\ \lambda\left(v_1 - 2v_2 + 3v_3\right) \\ \lambda\left(v_2 + v_3\right) \\ \lambda\left(2v_1 + 3v_2 - v_3\right) \end{pmatrix} = \lambda \begin{pmatrix} 3v_1 - 2v_2 \\ v_1 - 2v_2 + 3v_3 \\ v_2 + v_3 \\ 2v_1 + 3v_2 - v_3 \end{pmatrix} = \lambda T(\mathbf{v}).$$

Therefore T respects scalar multiplication as well. Thus, T is linear. □

(2) 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 6.12 (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 $\lambda_1, \ldots, \lambda_k$ and any k vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ 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 4.5 and proceed similarly. □

When checking if a function is linear we only need to check that it respects linear combinations of two vectors.

Theorem 6.13 (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 6.14 (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 6.15. None of the following functions is linear:

$$\begin{aligned} f\colon \mathbb{R} &\longrightarrow \mathbb{R}, && x \longmapsto 2\,x - 3 \\ T\colon \mathbb{R}^2 &\longrightarrow \mathbb{R}^3, && (x, y) \longmapsto (x - y + 2, 2\,x + 3\,y, 42\,x) \\ S\colon \mathbb{R}^3 &\longrightarrow \mathbb{R}^2, && (x, y, z) \longmapsto (2\,x - 3\,y + z, 42). \end{aligned}$$

A very useful consequence of Theorem 6.12 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 6.16. 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 6.17. *Let* $T, S\colon \mathbb{R} \longrightarrow \mathbb{R}^m$ *be linear functions and let* $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *be a basis of* $\mathbb{R}^n$. *If*

$$T\,\mathbf{v}_1 = S\,\mathbf{v}_1, \ldots, T\,\mathbf{v}_n = S\,\mathbf{v}_n$$

then we have

$$\forall \mathbf{v} \in \mathbf{R}^n, T\,\mathbf{v} = S\,\mathbf{v},$$

in other words

$$T = S.$$

Proof. Let $\mathbf{v} \in \mathbf{R}^n$ then there are unique $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ so that

$$\mathbf{v} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n.$$

Then we have

$$\begin{aligned} T\,\mathbf{v} &= \lambda_1\,T\,\mathbf{v}_1 + \cdots + \lambda_n\,T\,\mathbf{v}_n \\ &= \lambda_1\,S\,\mathbf{v}_1 + \cdots + \lambda_n\,S\,\mathbf{v}_n \\ &= S\,\mathbf{v}. \end{aligned}$$

□

Now, let's remember that the linear function defined by a matrix A with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ is given by the formula

$$A\,\mathbf{x} = x_1\,\mathbf{a}_1 + \cdots + x_n\,\mathbf{a}_n$$

where $\mathbf{x} = x_1\,\mathbf{e}_1 + \cdots + x_n\,\mathbf{e}_n$. In particular we have that the columns of A are the images of the standard basis, in other words

$$\mathbf{a}_i = A\,\mathbf{e}_i, \quad i = 1, \ldots, n.$$

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 6.18. *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 $T\,\mathbf{e}_1, \ldots, T\mathbf{e}_n$.*

Example 6.19 (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 $\mathbf{e}_1, \ldots, \mathbf{e}_n$. 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 6.20. 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 6.11. 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 6.21. 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[26] 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

[26] Verify all calculations yourself.

$$A = \begin{pmatrix} 2 & 0 & -1 \\ 18/5 & 3/5 & 1/5 \\ 4/5 & 4/5 & 3/5 \\ 6/5 & 1/5 & -3/5 \end{pmatrix}.$$

□

Example 6.22 (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.

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 6.18), 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. 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.

6.1. **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 6.23 (Function Spaces are Vector Spaces). *Addition and scalar multiplication of functions satisfy all the axioms listed in Theorem 4.3, where the role of the zero vector is played by the zero function O. In other words we have the following properties:*

(1) *Function addition is* commutative. *This means that for any two functions A, B we have*
$$A + B = B + A$$
(2) *Function addition is* associative. *This means that for any three functions A, B, and C we have*
$$(A + B) + C = A + (B + C)\,.$$
(3) *O is* neutral *for addition. This means that for any function A we have*
$$O + A = A.$$
(4) *For every function A we have*
$$A + (-A) = O.$$
(5) *The number 1 is neutral for scalar multiplication. This means that for every function A we have*
$$1\,A = A.$$
(6) *Scalar multiplication distributes over function addition. This means that if λ is a scalar and A, B are functions we have*
$$\lambda\,(A + B) = \lambda\,A + \lambda\,B.$$
(7) *Addition of scalars distributes over scalar multiplication. This means that*
$$(\lambda + \mu)\,A = \lambda\,A + \mu\,A.$$
(8) *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 6.24 (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 6.25 (A multiple of a linear function is a linear function). *If A is a linear function then λA is also linear for every $\lambda \in \mathbb{R}$. Furthermore, if the matrix of A is (a_{ij}) then the matrix of λA is*

(λa_{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 6.24 and is left as an exercise.

We can combine Theorems 6.24 and 6.25 into a single theorem.

Theorem 6.26 (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 6.27. 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 6.28 (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,

$$(AB)\,\mathbf{x} = x_1\begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} \\ a_{21}b_{11} + a_{22}b_{21} \\ a_{31}b_{11} + a_{32}b_{21} \end{pmatrix} + x_2\begin{pmatrix} a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{12} + a_{32}b_{22} \end{pmatrix}.$$

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,

$$(AB)\,\mathbf{x} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

So we got that,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{pmatrix}. \tag{25}$$

The same ideas can be used to get the formula for the matrix of AB in the general case where A has dimensions $m \times k$ for $m, k \geq 1$ and B has dimensions $k \times n$[27] 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 $C = AB$, 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 6.29. *Let A be an $m \times k$ matrix and B a $k \times n$ matrix. Then the entries of $C = AB$ are given by*

$$c_{ij} = \mathbf{a}_i^* \cdot \mathbf{b}_j = \sum_{\ell=1}^{k} a_{i\ell}\, a_{\ell j}. \tag{26}$$

Or, if we expand the sum in the RHS,

$$c_{ij} = a_{i1}\, b_{1j} + \cdots + a_{ik}\, b_{kj}.$$

[27]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 can express Equation (26) 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 6.30. Let's compute $A\,B$ and $B\,A$ where

$$A = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 0 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 1 \\ 2 & 0 \\ 4 & 3 \end{pmatrix}.$$

We have

$$A\,B = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 0 & 3 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 2 & 0 \\ 4 & 3 \end{pmatrix} = \begin{pmatrix} -1+4+0 & 1+0+0 \\ 1+0+12 & -1+0+9 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 13 & 8 \end{pmatrix},$$

while

$$B\,A = \begin{pmatrix} -1 & 1 \\ 2 & 0 \\ 4 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ -1 & 0 & 3 \end{pmatrix} = \begin{pmatrix} -1-1 & -2+0 & 0+3 \\ 2+0 & 4+0 & 0+0 \\ 4-3 & 8+0 & 0+9 \end{pmatrix} = \begin{pmatrix} -2 & -2 & 3 \\ 2 & 4 & 0 \\ 1 & 8 & 9 \end{pmatrix}.$$

6.2. **Some Exercises.**

Exercise. Let $P\colon \mathbb{R}^4 \longrightarrow \mathbb{R}^4$ be given by

$$P(x, y, z, w) = (y, z, w, x).$$

(1) Prove that P is linear using Theorem 6.13.
(2) Find the matrix of P.
(3) 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. An 1×1 matrix has only one entry $\begin{pmatrix} a \end{pmatrix}$. 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 $\mathbb{R}$.

(1) What functions $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ arise as the linear functions associated with 1×1 matrices?
(2) When is a linear function $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ one-to-one?
(3) When is a linear function $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ onto?
(4) When is a linear function $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ invertible?
(5) Let $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ be an invertible linear function. What is f^{-1}?
(6) 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. For each of the following functions T

(1) Prove that it is linear.
(2) Find the matrix that gives T.

(1) The function $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ given by

$$T\,\mathbf{x} = \lambda\,\mathbf{x}.$$

(2) The function $T\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ given by

$$T\,(x_1, x_2, x_3) = (x_1, x_2).$$

(3) The function $T\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ given by
$$T(x_1, x_2) = (x_1, x_2, 0).$$
(4) The function $T\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ given by
$$T(x, y, z) = (x + z, x - z).$$

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

(1) Prove that the columns of the matrix form a basis of $\mathbb{R}^3$.
(2) 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.
(3) Find $T\,\mathbf{a}_i^*$, for $i = 1, 2, 3$.

7. 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 7.1 (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 $\mathbf{x} = (x_1, \ldots, x_n)$ is such a solution then

$$x_1\,\mathbf{a}_1 + \cdots + x_n\,\mathbf{a}_n = \mathbf{y},$$

where $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the columns of A. Thus the range of A is the linear span of its columns. So we have the following theorem.

Theorem 7.2 (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 $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the columns of A, then*

$$\mathcal{R}(A) = \langle \mathbf{a}_1, \ldots, \mathbf{a}_n \rangle.$$

Definition 7.3 (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 5.1 as follows.

Theorem 7.4. *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 7.5. 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[28]

$$A \sim \begin{pmatrix} 1 & 0 & 0 & 22 & -21 \\ 0 & 1 & 0 & -5 & 7 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

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 1.32 in Section 1.3.1) that the solution sets of homogeneous systems are subspaces of $\mathbb{R}^m$.

Definition 7.6 (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 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 7.5. The solution of $A\mathbf{x} = \mathbf{0}$ is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -22s + 21t \\ 5s - 7t \\ -s + 2t \\ s \\ t \end{pmatrix} = s\begin{pmatrix} -22 \\ 5 \\ -1 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} 21 \\ -7 \\ 2 \\ 0 \\ 1 \end{pmatrix}, \qquad s, t \in \mathbb{R}. \tag{27}$$

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 (27) 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 $\mathbf{e}_2$.

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

[28] Do the calculations yourself!

$$\begin{cases} x_1 &= -s\,b_{13} - t\,b_{14} - w\,b_{17} \\ x_2 &= -s\,b_{23} - t\,b_{24} - w\,b_{27} \\ x_3 &= s \\ x_4 &= t \\ x_5 &= -s\,b_{53} - t\,b_{54} - w\,b_{57} \\ x_6 &= -s\,b_{63} - t\,b_{64} - w\,b_{67} \\ x_7 &= w \end{cases}$$

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 7.7. *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 7.4 and 7.7 we have the following theorem.

Theorem 7.8 (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 7.9. 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 7.10. 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

$$\ker A = \langle(2,-1,-1,1,0)\rangle\,.$$

8. 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

$$y = f(x). \tag{28}$$

(1) A function $f\colon X \longrightarrow Y$ is *injective* if for all $y \in Y$ Equation (28) has *at most* one solution.
(2) A function $f\colon X \longrightarrow Y$ is *surjective* if for all $y \in Y$ Equation (28) has *at least* one solution.
(3) A function is a *bijection*, i.e. both injective and surjective, if and only if for all $y \in Y$ Equation (28) has a unique solution.
(4) 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 (28).

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 8.1. *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.

(1) A is injective if and only if B has no free columns.

(2) *A is injective if and only if its kernel contains only the zero vector, i.e.*

$$\ker A = \{\mathbf{0}\}$$

.

(3) *A is injective if and only if* $\operatorname{null} A = 0$.
(4) *If A is injective then $n \le m$.*
(5) *A is surjective if and only if its columns span $\mathbb{R}^m$.*
(6) *A is surjective if and only if* $\operatorname{rank} A = m$.
(7) *A is surjective if and only if B has m basic columns.*
(8) *If A is surjective then $n \ge m$.*
(9) *A is invertible if and only if $B = I_n$, the $n \times n$ identity matrix.*
(10) *A is invertible if and only if the columns of A form a basis of $\mathbb{R}^n$.*
(11) *If A is invertible then A^{-1} is linear.*
(12) *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.

(1) This just says that a consistent system has a unique solution if and only if there are no free variables.
(2) This just says that a consistent system has unique solution if and only if the corresponding homogeneous system has only the trivial solution.
(3) The nullity of A is the dimension of its kernel. A subspace has 0 dimension if and only if it equals $\mathbf{0}$.
(4) 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.
(5) 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 $\mathbb{R}^m$.
(6) 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 $\mathbb{R}^m$.
(7) 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.
(8) If there are less variables than equations there are not enough columns to form a basis of $\mathbb{R}^m$.
(9) A is invertible if an only if the system $A\,\mathbf{x} = \mathbf{c}$ has a unique solution, for all $\mathbf{c} \in \mathbb{R}^m$.
(10) 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.
(11) 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

$$\lambda\, A^{-1}\mathbf{x} + \mu\, A^{-1}\,\mathbf{y} = A^{-1}\left(\lambda\,\mathbf{x} + \mu\,\mathbf{y}\right).$$

(12) It follows from (4) and (8).

□

As a consequence of the Rank-Nullity Theorem (see Theorem 7.8), we have the following Theorem.

Theorem 8.2. *Let $A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a linear map. Then the following are equivalent.*

(1) A is injective.
(2) A is surjective.
(3) 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 7.8 we have $\operatorname{rank} A = n$. Thus A is surjective.

If A is surjective then $\operatorname{rank} A = n$ and therefore, again by Theorem 7.8 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* 8.3.** 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 8.4 (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* 8.5.** 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}[29]. □

[29]You should do it!

Example 8.6 (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 8.7 (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

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - dc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{29}$$

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 (29) 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

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \frac{1}{D}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}. \tag{30}$$

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 8.4)

$$\begin{aligned} A\mathbf{x} = \mathbf{c} &\iff A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{c} \\ &\iff \left(A^{-1}A\right)\mathbf{x} = A^{-1}\mathbf{c} \\ &\iff I\mathbf{x} = A^{-1}\mathbf{c} \\ &\iff \mathbf{x} = A^{-1}\mathbf{c}. \end{aligned}$$

Thus, we can recover Crammer's rule (see Section 3). Indeed, we have that the solution of the system is

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{D}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \frac{1}{D}\begin{pmatrix} c_1a_{22} - c_2a_{12} \\ -c_1a_{21} + c_2a_{11} \end{pmatrix}.$$

9. 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 $\mathbf{M}_{m,n}$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* 9.1.** 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 9.2 (Matrix addition and scalar multiplication). For any positive integers m, n we have the operations of addition

$$\mathbf{M}_{mn} \times \mathbf{M}_{mn} \longrightarrow \mathbf{M}_{mn}, \quad (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij}),$$

and scalar multiplication

$$\mathbb{R} \times \mathbf{M}_{mn} \longrightarrow \mathbf{M}_{mn}, \quad \lambda\,(a_{ij}) = (\lambda\, a_{ij}).$$

Of course, these are the "same" operations we've seen in Section 6.1, 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 6.23 hold.

Since we have proved[30] 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 4.3. 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. Prove all the *Vector Space Axioms* (i.e. the properties listed in Theorem 6.23) for matrices in this manner.

The above discussion suggests that we can think of matrices as vectors. In fact an $m \times n$ matrix consists of $m\,n$ 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 $m\,n$-vector, i.e. and element of $\mathbb{R}^{m\,n}$. For example,

[30] We did do the proofs left as an exercise. Didn't we?

$$\mathbf{M}_{23} \ni \begin{pmatrix} 1 & 2 & -1 \\ 3 & -2 & 0 \end{pmatrix} \cong (1, 2, -1, 3, -2, 0) \in \mathbb{R}^6.$$

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 $\mathbf{M}_{23}$ 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 9.3 (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 9.4. 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 9.5 (The standard basis of $\mathbf{M}_{mn}$). 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 9.6. *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}.*

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 AB the *product* of A and B. Let's recall the definition.

Matrix Multiplication

If $A = (a_{ij}) \in \mathbf{M}_{mk}$ and $B = (b_{ij}) \in \mathbf{M}_{kn}$ then their product $C := A\,B \in \mathbf{M}_{mn}$ 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 9.7 (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*[31].

(1) Matrix multiplication is associative.

$$A(B\,C) = (A\,B)\,C.$$

(2) Matrix multiplication distributes over matrix addition on both sides.

$$(A + B)\,C = A\,C + A\,B, \qquad A(B + C) = A\,B + A\,C.$$

(3) Scalar multiplication is compatible with matrix multiplication.

$$(\lambda A)\,B = A(\lambda B) = \lambda\,(A\,B).$$

(4) Multiplication with the identity matrix

$$I\,A = A, \qquad A\,I = A.$$

[31]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.

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 9.7. 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 9.7, the benefit of this being that the proposition will be true whenever those properties (as well as the *vector space axioms*) hold.

Proposition 9.8. *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

$$A\,B = B\,A. \tag{31}$$

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 (31) 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 $\mathbf{M}_{33}$. 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**

$$A\,B = O \implies A = O \text{ or } B = O.$$

9.1. **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 (31) 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 9.9 (Commuting matrices). If Equation (31) 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 9.10. 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[32] we see that

$$A\,B = \begin{pmatrix} -3 & -45 & -3 \\ 15 & 6 & -24 \\ 21 & 12 & -3 \end{pmatrix} = B\,A.$$

Example 9.11 (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

$$M\,A = \begin{pmatrix} x & y \\ z & t \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} x & x+y \\ z & z+t \end{pmatrix}.$$

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:

$$\begin{cases} \quad x \qquad\quad = x \\ \quad x + y \qquad = y + t \\ \qquad\qquad z = z \\ t \qquad\quad + z = t \end{cases}.$$

[32] Do them!

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 9.12 (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

(1) $\lambda A \in \mathbf{A}$.
(2) $A + B \in \mathbf{A}$.
(3) $A B \in \mathbf{A}$.

If in addition any two elements of $\mathbf{A}$ commute, that is, if in addition

(4) $A B = B A$,

then we say that $\mathbf{A}$ is a *commutative algebra of matrices*.

Theorem 9.13. *If $\mathbf{A}$ is an algebra of matrices then:*

(1) $O \in \mathbf{A}$.
(2) $A \in \mathbf{A} \implies -A \in \mathbf{A}$.
(3) If $A \in \mathbf{A}$ and A is invertible then $A^{-1} \in \mathbf{A}$.
(4) 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 9.14 (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 9.15 (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 λI with $\lambda \neq 0$, and of course

$$(\lambda I)^{-1} = \lambda^{-1} I.$$

Example 9.16 (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 $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$ to be the diagonal matrix with $\lambda_1, \dots, \lambda_n$ 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 λI is thus $\operatorname{diag}(\lambda, \dots, \lambda)$.

Notice that the i-th row (as well as the i-th column) of $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$ 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, $\mathbf{a}_1^*, \dots, \mathbf{a}_n^*$ (respectively $\mathbf{a}_1, \dots, \mathbf{a}_m$) be its row (respectively column) vectors. Then,

$$\operatorname{diag}(\lambda_1, \dots, \lambda_n)\, A = \begin{pmatrix} \lambda_1 \mathbf{a}_1^* \\ \vdots \\ \lambda_n \mathbf{a}_n^* \end{pmatrix}, \quad A \operatorname{diag}(\lambda_1, \dots, \lambda_m) = \begin{pmatrix} \lambda_1 \mathbf{a}_1 & \dots & \lambda_m \mathbf{a}_m \end{pmatrix}.$$

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

$$\operatorname{diag}(\lambda_1, \dots, \lambda_n) \operatorname{diag}(\mu_1, \dots, \mu_n) = \operatorname{diag}(\lambda_1 \mu_1, \dots, \lambda_n \mu_n).$$

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

$$\operatorname{diag}(\lambda_1, \dots, \lambda_n) + \operatorname{diag}(\mu_1, \dots, \mu_n) = \operatorname{diag}(\lambda_1 + \mu_1, \dots, \lambda_n + \mu_n),$$

and we established that the set of diagonal matrices is a commutative algebra.

It is rather straightforward[33] to see that a diagonal matrix is invertible if and only if all diagonal entries are non=zero. In that case,

$$\operatorname{diag}(\lambda_1, \dots, \lambda_n)^{-1} = \operatorname{diag}(\lambda_1^{-1}, \dots, \lambda_n^{-1}).$$

Example 9.17 (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

$$i > j \implies a_{ij} = 0.$$

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[34] 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 $A\,B$ is 0.

[33] Is it?

[34] It is easy, right?

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 9.18 (Centralizers). Recall from Example 9.11 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

$$A \in \mathcal{C} \iff \forall X \in S, \quad A\,X = X\,A.$$

I claim that $\mathcal{C}$ is an algebra. The claim follows from the following three facts:

- If A and X commute, so do $\lambda\,A$ and X.

  Proof. Assume A and X commute. Then we have
  $$(\lambda\,A)\;X = \lambda\,(A\,X) = \lambda\,(X\,A) = X\;(\lambda\,A)\,.$$
  □

- If A and X, and B and X commute, the $A + B$ and X and commute.

  Proof. We have
  $$(A + B)\,X = A\,X + B\,X = X\,A + X\,B = X\,(A + B).$$
  □

- If A and X, and B and x commute, the $A\,B$ and X and commute.

  Proof. We have
  $$(A\,B)\,X = A\,(B\,X) = A\,(X\,B) = (A\,X)\,B = (X\,A)\,B = X\,(A\,B).$$
  □

Example 9.19 (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.

Now let A be as above and let $B = \begin{pmatrix} x & y \\ y & x \end{pmatrix}$ be a second element of $\mathbf{A}$. 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 $\mathbf{A}$ is closed under multiplication as well.

We have established then that $\mathbf{A}$ 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 $\mathbf{A}$ is a commutative algebra of matrices.

Now let's find the invertible elements of $\mathbf{A}$. From Example 8.7 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 9.20 (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* 9.21.** Because of the associative property of multiplication (the first property in Theorem 9.7), 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 9.22. *If $A \in \mathbf{M}_n$ and $k, \ell \in \mathbb{N}$ we have*

(1) $A^k\,A^\ell = A^{k+\ell}$.

(2) $\left(A^k\right)^l = A^{k\,l}$.

(3) $(\lambda A)^k = \lambda^k A^k$.
(4) $I^k = I$.
(5) $O^k = O$.

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 9.36.

Theorem 9.23. *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 9.24. 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 9.25. *The following hold. The powers could be positive or negative integers; in the later case we assume that the involved matrices are invertible.*

(1) *Properties (1) through (3) in Proposition 9.22 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*[35].
(2) *All powers of the same matrix commute.*
(3) *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.*
(4) *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 9.26 (Powers of diagonal matrices). Refer to Example 9.16 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.

Let's prove this by induction. It clearly it is true for $k = 0$. Now,

$$\begin{aligned} A^{k+1} &= A^k A \\ &= \operatorname{diag}\left(a_1^k, \ldots, a_n^k\right) \operatorname{diag}\left(a_1, \ldots, a_n\right) \\ &= \operatorname{diag}\left(a_1^k a_1, \ldots, a_n^k a_n\right) \\ &= \operatorname{diag}\left(a_1^{k+1}, \ldots, a_n^{k+1}\right). \end{aligned}$$

Now, if all diagonal entries are non-zero then (see Example 9.16) A is invertible and if $k < 0$ then $-k > 0$ and

$$A^k = \left(A^{-1}\right)^{-k} = \operatorname{diag}\left(a_1^{-1}, \ldots, a_n^{-1}\right)^{-k} = \operatorname{diag}\left(a_1^k, \ldots, a_n^k\right).$$

Example 9.27. Consider the matrix

$$A = \frac{1}{2}\begin{pmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix}.$$

We have

---

[35]Why?

$$A^2 = \frac{1}{4}\begin{pmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix}\begin{pmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1-3 & -\sqrt{3}-\sqrt{3} \\ \sqrt{3}+\sqrt{3} & -3+1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}.$$

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 9.28. 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 9.29. 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 $n \geq 1$[36]

$$A^n = 2^{n-1}\,A.$$

For $n = 1$ the formula clearly holds since both sides are equal to A. Assuming it holds for n we get

$$A^{n+1} = A^n\,A = \left(2^{n-1}\,A\right)\,A = 2^{n-1}\,A^2 = 2^{n-1}\,2\,A = 2^n\,A = 2^{n+1-1}\,A.$$

Example 9.30. 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,

$$A^3 = A^2\,A = I\,A = A, \quad A^4 = A^3\,A = A\,A = I.$$

And therefore[37]

[36]Why the formula doesn't work for $n = 0$?

[37]Give an inductive proof of this.

$$A^n = \begin{cases} I & n \text{ even} \\ A & n \text{ odd} \end{cases}.$$

Now that we have powers, scalar multiplication, and addition we can plug a matrix in any polynomial with real coefficients.

Definition 9.31 (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* 9.32.** 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 9.33. 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

$$\begin{aligned} p\,(A) &= A^3 - 2\,A^2 = 2\,A + 6 \\ &= \begin{pmatrix} -46 & -9 \\ 9 & -46 \end{pmatrix} - 2\begin{pmatrix} -5 & -12 \\ 12 & -5 \end{pmatrix} - 2\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} + 6\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -46 & -9 \\ 9 & -46 \end{pmatrix} + \begin{pmatrix} 10 & 24 \\ -24 & 10 \end{pmatrix} + \begin{pmatrix} -4 & 6 \\ -6 & -4 \end{pmatrix} + \begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix} \\ &= \begin{pmatrix} -34 & 21 \\ -21 & -34 \end{pmatrix}. \end{aligned}$$

And

$$\begin{aligned} q(A) &= A^2 - 4A + 13I \\ &= \begin{pmatrix} -5 & -12 \\ 12 & -5 \end{pmatrix} - 4\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} + 13\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -5 & -12 \\ 12 & -5 \end{pmatrix} + \begin{pmatrix} -8 & 12 \\ -12 & -8 \end{pmatrix} + \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

So A is a root of $q(x)$.

The following theorem is immediate[38].

Theorem 9.34. *If* $\mathbf{A}$ *is an algebra of matrices,* $A \in \mathbf{A}$ *and* $p(x)$ *is any polynomial, then* $p(A) \in \mathbf{A}$.

9.2. **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 9.35 (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

$$AB = I = BA. \tag{32}$$

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 9.36 (Invertible matrices form a group). *We have*

(1) $A \in \mathrm{GL}(n) \implies A^{-1} \in \mathrm{GL}(n)$, *and actually*
$$\left(A^{-1}\right)^{-1} = A.$$
(2) $A, B \in \mathrm{GL}(n) \implies AB \in \mathrm{GL}(n)$, *and actually*[39]
$$(AB)^{-1} = B^{-1}A^{-1}.$$

Proof. The first is obvious since by definition we have
$$AA^{-1} = I = A^{-1}A,$$
and therefor A is the inverse of A^{-1}.

For the second we have:
$$(AB)\left(B^{-1}A^{-1}\right) = A\left(BB^{-1}\right)A^{-1} = AIA^{-1} = AA^{-1} = I.$$
Similarly,
$$\left(B^{-1}A^{-1}\right)(AB) = B^{-1}\left(A^{-1}A\right)B = B^{-1}IB = B^{-1}B = I.$$
□

It turns out that we don't need to check that both products in Equation (32) give the identity matrix. As the following Lemma shows, if one of the products is the identity the other will be as well.

Lemma 9.37. *If* $AB = I$ *then we also have* $BA = I$ *and therefore* $B = A^{-1}$. *Similarly, if* $AB = I$ *then* $B = A^{-1}$.

[38] Provide the proof

[39] Notice the reverse of the order!

Proof. If $AB = I$ 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 8.2 it follows that A is invertible. We have then

$$\begin{aligned} AB = I &\implies A^{-1}(AB) = A^{-1} \\ &\implies \left(A^{-1}A\right)B = A^{-1} \\ &\implies IB = A^{-1} \\ &\implies B = A^{-1}. \end{aligned}$$

□

The properties listed in Theorem 9.36 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 9.38 (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:

(1) The identity function of X is in G.
(2) G is closed under composition of functions.
(3) 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 9.36 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 9.22. We will not pursue this further at this point. We'll come back to these ideas later though.

In Examples 9.27 and 9.30 we have matrices where one of their powers is the identity matrix. Lemma 9.37 implies that such matrices are invertible because if $A^k = I$ then $A^{k-1}A = I$.

Example 9.39. Let

$$A = \frac{1}{9}\begin{pmatrix} 4 & 7 & -4 \\ 1 & 4 & 8 \\ 8 & -4 & 1 \end{pmatrix}.$$

We calculate[40] 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 9.40. *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.*

[40] Do the calculations!

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 9.37, 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 9.41. In Example 9.33 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 9.42. 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[41] 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

$$X^{-1} = -\frac{1}{2}\left(X^3 - 5X^2 + 9X - 7I\right).$$

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[42]

$$X^{-1} = \frac{1}{2}\begin{pmatrix} 3 & -2 & 1 & 1 \\ 8 & -2 & 0 & 0 \\ -1 & -4 & 5 & 3 \\ 7 & -2 & -1 & 1 \end{pmatrix}.$$

Down the road, as a consequence of the Cayley-Hamilton Theorem we will see that the converse of Proposition 9.40 is also true. We state the proposition postponing the proof.

Proposition 9.43. *If X is invertible then it is a root of a polynomial with non-zero constant term.*

Combining this with 9.34 we get the proof of Theorem 9.13.

Proof of Theorem 9.13. If A is invertible then by Propositions 9.43 and 9.40 we have that A^{-1} is a polynomial of A. Theorem 9.34 then implies that $A^{-1} \in \mathbf{A}$. □

[41]Do the calculations!

[42]Do the calculations!

10. 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 10.1 (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 10.2. 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 10.3 (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.

$$\mathbf{a}\colon \mathbb{R}^1 \longrightarrow \mathbb{R}^n, \quad \mathbf{a}\,\mathbf{x} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \begin{pmatrix} x \end{pmatrix} = \begin{pmatrix} a_1\,x \\ \vdots \\ a_n\,x \end{pmatrix} = x\,\mathbf{a}.$$

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

$$\mathbf{a}^*\colon \mathbb{R}^n \longrightarrow \mathbb{R}^1, \quad \mathbf{x} \longmapsto \mathbf{a}\,\mathbf{x} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} x_1\,a_1 + \cdots + x_n\,a_m \end{pmatrix} = \mathbf{x} \cdot \mathbf{a}.$$

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:

(33) $$A^*\,\mathbf{x} = \left(\mathbf{x}^*\,A\right)^*.$$

Using Equation (33) we can prove the following property of the transpose.

Theorem 10.4 (Transpose is an anti-homomorphism). *The following hold.*

1. *Transpose respects scalar multiplication. That is, for any scalar* λ
$$\left(\lambda\,A\right)^* = \lambda\,A^*.$$
2. *Transpose respects matrix addition. That is,*
$$\left(A + B\right)^* = A^* + B^*.$$

(3) If $A\,B$ is defined then $B^*\,A^*$ is also defined and

$$(AB)^* = B^*\,A^*. \tag{34}$$

Proof. The proof of (1) and (2) are straightforward and left as an exercise.

For the third, we will use Equation (33). 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 (34) has the same structure as property (2) in Theorem 9.36. 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 10.5. *We have for $k \in \mathbb{N}$.*

(1) If $A \in \mathbf{M}_n$ then

$$\left(A^k\right)^* = (A^*)^k.$$

(2) If $p(x)$ is any polynomial then

$$p(A^*) = (p(A))^*.$$

Proof. The first follows from Equation (34), 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

$$\begin{aligned}p\,(A^*) &= a_d\,(A^*)^d + \cdots + a_1\,(A^*)^1 + a_0\,(A^*)^0\\ &= a_d\,\left(A^d\right)^* + \cdots + a_1\,A^* + a_0\,I\\ &= \left(a_d\,A^d\right)^* + \cdots + (a_1\,A)^* + (a_0\,I)^*\\ &= \left(a_d\,A^d + \cdots + a_1\,A + a_0\,I\right)^*\\ &= (p(A))^*.\end{aligned}$$

□

Furthermore, as the following Theorem shows, if A is invertible so is its transpose.

Theorem 10.6 (Transposing and Inverting commute). *If A is invertible then so is A^*. Furthermore*

$$(A^*)^{-1} = \left(A^{-1}\right)^*.$$

Proof. We have:

$$\begin{aligned} A\,B = I &\implies (A\,B)^* = I^* \\ &\implies B^*\,A^* = I. \end{aligned}$$

And the result follows from Lemma 9.37. □

Definition 10.7 (Symmetric and orthogonal matrices). A *square* $n \times n$ matrix A is called *symmetric* if

$$A = A^*. \tag{35}$$

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 (35), 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

$$A\,A^* = I. \tag{36}$$

A linear operator $A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ that satisfies Condition (36), is said to be a *(linear) isomometry*, 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 10.2, 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 10.8. *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 10.9. *If A is an $m \times n$ matrix then $A\,A^*$ and $A^*\,A$ are symmetric matrices.*

Proof. We have

$$(A^*\,A)^* = A^*\,(A^*)^* = A^*\,A.$$

Similarly, we have

$$(A\,A^*)^* = A\,A^*.$$

□

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 10.10. *If $\lambda \in \mathbb{R}$ and $A, B \in \mathbf{M}_n$ then:*

(1) $A \in \mathbf{S}_n \implies \lambda A \in \mathbf{S}_n$.
(2) $A, B \in \mathbf{S}_n \implies A + B \in \mathbf{S}_n$.
(3) $A, B \in \mathbf{S}_n \implies (A\,B)^* = B\,A$.
(4) $A, B \in \mathbf{S}_n \implies A\,B + B\,A \in \mathbf{S}_n$.

Proof. If $A^* = A$ and $B^* = B$ we have:

(1) $(\lambda A)^* = \lambda A^* = \lambda A$.
(2) $(A + B)^* = A^* + B^* = A + B$.
(3) $(A\,B)^* = B^*\,A^* = B\,A$.
(4) $(A\,B + B\,A)^* = (A\,B)^* + (B\,A)^* = B^*\,A^* + A^*\,B^* = B\,A + A\,B = A\,B + B\,A$.

□

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 10.11. *If $\lambda \in \mathbb{R}$ and $A, B \in \mathbf{M}_n$ then:*

(1) $A \in \mathrm{O}(n) \implies A^{-1} \in \mathrm{O}_n$.
(2) $A, B \in \mathrm{O}(n) \implies A\,B \in \mathrm{O}(n)$.

Proof. To prove that a matrix is orthogonal we have to prove that its inverse is equal to its transpose.

(1) 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.

(2) Let A and B be two orthogonal matrices. Then we have

$$\begin{aligned}(A\,B)^{-1} &= B^{-1}\,A^{-1}\\ &= B^*\,A^*\\ &= (A\,B)^*\,.\end{aligned}$$

□

10.1. **The rank of the transpose.** In Section 6 we saw that a matrix $A \in \mathbf{M}_{mn}$ 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 10.12. *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 10.13 (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 10.14. *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 10.6.

11. Elementary matrices and row (or column) operations

We have seen two ways of solving systems of linear equations. In Section 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 8.4 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 8.6). 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:

(1) **Type I:** Interchange two rows.
(2) **Type II:** Multiply a row by a non-zero scalar.
(3) **Type III:** Add a row to an other row.

Definition 11.1 (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 11.2. The following are 4×4 elementary matrices of type I, II, and III respectively:

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

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 11.3. *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* 11.4.** 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 9.16.

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 EA 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 EA 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 11.5. 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 11.6. *All elementary matrices are invertible. Namely,*

(1) $P_{k\ell}^{-1} = P_{k\ell}$.
(2) $M_{k;\lambda}^{-1} = M_{k;\lambda^{-1}}$.
(3) $S_{k\ell}^{-1} = M_{k;-1}\,S_{k\ell}M_{k;-1}$.

Proof. By Theorem 11.6 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 1.10[43]

$$\begin{cases} x + y + z = 3 \\ x - y + z = 1 \\ 4x + 2y + z = 10 \end{cases}.$$

The corresponding vector equation is

$$A\,\mathbf{x} = \mathbf{c} \tag{37}$$

where

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 4 & 2 & 1 \end{pmatrix}, \text{ and } \mathbf{c} = \begin{pmatrix} 3 \\ 1 \\ 10 \end{pmatrix}.$$

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

[43] I changed the variables to x, y and z.

row by the sum of the first and second row. In terms of elementary matrices this is equivalent to first multiply Equation (37) 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 (37) by A^{-1}.

The above discussion also gives an algebraic interpretation of the method of finding the inverse of a matrix exposed in Example 8.6. 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 11.7 (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 11.8. 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* 11.9.** 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 1.18) 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 11.10. *Two $m \times n$ matrices A, B are row equivalent if and only if there an invertible $m \times m$ matrix C such that*

$$B = C\, A.$$

11.1. **Column operations.** We introduced row operations in Section 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

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

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[44] 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 11.10.

Definition 11.11 (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.

Theorem 11.12. *Column equivalence is an equivalence relation. In other words if X, Y, Z are $m \times n$ matrices, we have:*

(1) $X \cong X$.
(2) $X \cong Y \implies Y \cong Z$.
(3) $X \cong Y$ and $Y \cong Z \implies X \cong Z$.

Proof. (1) holds because $X = XI$.

(2) follows from the implication $Y = XC \implies X = YC^{-1}$.

To prove (3), notice that if $Y = XC$ and $Z = YD$ then $Z = X(CD)$. Furthermore, if C, D are invertible then so is CD (see Theorem 9.36). □

[44]This is not pure science fiction. There are books where this choice is made.

Theorem 11.13. *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 10.4)

$$B = A\,C \iff B^* = C^*\,A^*.$$

and C is invertible if and only if C^* is invertible (see Theorem 10.6). Thus, if $A \cong B$ then (by Theorem 11.10) $A^* \sim B^*$.

Conversely, if $A^* \sim B^*$ then, again by Theorem 11.10 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 11.14. *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 if 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.

On the other hand, 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 11.15. *We have*

(1) $P^*_{k\ell} = P_{k\ell}$.
(2) $M^*_{k;\lambda} = M_{k;\lambda}$.
(3) $S^*_{k\ell} = S_{\ell k}$.

Theorem 11.16. *Let A be an $m \times n$ matrix. Then*

(1) $A\,P_{k\ell}$ has the same columns as A with the k and ℓ columns interchanged.
(2) $A\,M_{k;\lambda}$ has the same columns as A except the k-th that is equal to λ times the k-th column of A.
(3) $A\,S_{k\ell}$ 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 11.17. 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. Prove that $\approx$ is an equivalence relation.

Theorem 11.18. *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 11.19. 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 10.13 using Theorem 11.18.