MATH2210
Linear Algebra Notes

by

Özgür Esentepe

Contents

Part I

Preliminaries

1. Systems of Linear Equations

1.1. Addition is easy, multiplication is difficult

In this first chapter, we will deal with systems of linear equations and related topics. Therefore, the first question we should ask is what we mean by *linear equations* and why we care about them. We shall start with some small examples.

A linear equation in one variable x is an equation of the form $ax = b$ where a and b are some scalars. In this chapter, we will choose a and b to be real numbers but scalars may come from different number systems, too. A linear equation in two variables looks like $ax + by = c$. As you very well know this equation defines a *line* in the plane. Similarly, a linear equation in three variables x, y, z looks like $ax + by + cz = d$ where a, b, c, d are scalars.

From the examples, one may guess that a linear equation in variables $x_1, \dots, x_n$ is an equation involving these variables where one is only allowed the two operations: *addition and scalar multiplication*. We are not allowed to have x_1x_2 or x_3^2 or $x_1^2x_2^3$. In the rest of this section, I will try to argue why we only allow these operations.

Exercise 1.1.1. What is $236 + 457$? Compute by hand and keep track of how long it takes you to do this computation.

Exercise 1.1.2. What is 236×457? Compute by hand and keep track of how long it takes you to do this computation.

From these easy exercises, it is easy to convince yourself that addition is a much easier operation than multiplication in general.

Extra Exercise 1.1.3. Find places in mathematics where multiplication is easier than addition.

The next attempt to convince you is from geometry. We have said that a linear equation in two variables define a line in the plane and if you have taken a course in multivariable calculus you know that a linear equation in three variables gives you a plane in the three dimensional space. The purpose of the next exercise is to show you that adding a term which is *not linear* makes the geometric object much more complicated very very quickly.

Exercise 1.1.4. For each of the following equations, sketch the set of points satisfying the equation.

1. $3x = 3$
2. $2x + y = 4$
3. $xy + y^2 = 1$
4. $x + xy^3 = 3$
5. $x + y + z^2 = 7$.

From the first two exercises, you also see that allowing scalar multiplication is not making things much more complicated both in terms of algebraic computations and geometric picture. The equation $x + y = 4$ defines a line but the equation $2x + y = 4$ also defines a line. It is not too complicated. But what does the third equation define, how about the fourth one? The point is that they define something more complicated than a line.

The next attempt to convince you is from calculus. Consider two differentiable functions f and g. For example, let $f(x) = \cos x$ and $g(x) = x^2$.

Exercise 1.1.5.

1. Compute the derivative of $2f + g$.
2. Compute the derivative of fg.

The point of this exercise is to remind you that addition and scalar multiplication operations are usually not that bad but the product rule for derivatives is much more complicated.

You can convince yourself that the same holds integration.

Exercise 1.1.6. Find antiderivatives for $2f + g$ and fg.

Okay, after all this discussion, let's say we are convinced that addition is easier than multiplication. What does this mean? Are we doing linear algebra just because it is easy? Well, of course the answer is not yes. It is easy but also it gives you a lot of information. Remember again from your calculus class that differentiation gives you the best linear approximation of your function at a given point. This loses you a lot of information, yes. But it also keeps the information about important things such as increasing/decreasing behaviour of the function. Looking at a linear approximation is easy *and* useful.

1.2. Systems of linear equations

In this section, we are going to introduce the main character of our course: a system of linear equations.

Definition 1.2.1. A *system of linear equations* with m equations and n variables is a system of equations

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad &\vdots \quad \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

with n variables $x_1, \ldots, x_n$ appearing in m equations.

As an example, let us consider

$$\begin{aligned} 2x + 3y + 4z &= 9 \\ 5x + 2y + 4z &= 11 \end{aligned}$$

which is a system of 2 equations in three variables. We say that a triple (a, b, c) *solves this system* or *is a solution for this system* if

$$\begin{aligned} 2a + 3b + 4c &= 9 \\ 5a + 2b + 4c &= 11. \end{aligned}$$

That is, if (a, b, c) satisfies all the equations in the system, we say that it satisfies the system. In the future, we will ask "find the solution set to a system" by which we will mean "find *all* the solutions to a system".

Exercise 1.2.2. Verify that $(1,1,1)$ is a solution for the example system.

Throughout this section, we will make a couple important observations. We will make these observations through the example. However, these observations will hold true for all systems of linear equations. Recall that our running example is

$$\begin{aligned} 2x + 3y + 4z &= 9 \\ 5x + 2y + 4z &= 11. \end{aligned}$$

Suppose that we have a solution (a,b,c) for this system. Recall that this means $2a+3b+4c = 9$ and $5a+2b+4c = 11$. Therefore, we must have

$$(2a + 3b + 4c) + (5a + 2b + 4c) = 9 + 11 = 20$$

which gives us

$$7a + 5b + 8c = 20.$$

In other words, (a,b,c) is a solution to $7x+5y+8z = 20$. If you were following closely, we just showed that *if* (a,b,c) is a solution to $2x + 3y + 4z = 9$ and $5x+2y+4z = 11$, *then* it is also a solution to the equation $7x+5y+8z = 20$ which is the sum of the two equations. This was our first observation: if we have a solution to a system of two equations, that solution also is a solution for the sum of the two equations.

Exercise 1.2.3. Verify that $(1,1,1)$ is a solution to $7x + 5y + 8z = 20$.

Our second observation is simpler than this. Consider $2x + 3y + 4z = 9$ and $4x + 6y + 8z = 18$. You must have already realized that the second equation is just twice the first equation. If (a,b,c) is a solution to the first equation, we have $2a + 3b + 4c = 9$ and therefore

$$2(2a + 3b + 4c) = 2 \times 9$$

which gives us

$$4a + 6b + 8c = 18.$$

Thus, (a,b,c) satisfies $4x + 6y + 8z = 18$. Therefore, if we have a solution for an equation, it is a solution for all scalar multiples for the same equation. (Note that 2 here was arbitrary, the same argument would hold for 3 or 5 or any $k \in \mathbb{R}$.)

Exercise 1.2.4. Verify that $(1, 1, 1)$ is a solution for $4x + 6y + 8z = 18$.

Exercise 1.2.5. Using ideas similar to above, write a convincing paragraph showing that if you have a solution for a system of two equations E_1 and E_2, then it is also a solution for $kE_1 + lE_2$ for any choice of k and l.

Now, if you were able to solve this exercise, then you should be convinced that the two systems

$$\begin{aligned} 2x + 3y + 4z &= 9 \\ 5x + 2y + 4z &= 11 \end{aligned}$$

and

$$\begin{aligned} 2x + 3y + 4z &= 9 \\ 9x + 8y + 12z &= 29 \end{aligned}$$

have the *same solution set*. That is, any solution to the first system is also a solution for the second system and any solution to the second system is also a solution for the first system. Let us explain this a little bit further. Let us put E_1 and E_2 to be the first and second equations in the first system and pick a solution (a, b, c) to the first system. Then, by definition we know that (a, b, c) is a solution to E_1 and by the exercise we know that (a, b, c) is a solution to $2E_1 + E_2$. Note that E_1 and $2E_1 + E_2$ are the two equations in the second system. Therefore, (a, b, c) satisfy the second system as well!

Now, let us name F_1 and F_2 the equations in the second system so that we have $F_1 = E_1$ and $F_2 = 2E_1 + E_2$. Note that from here, we see that $E_2 = F_2 - 2F_1$. Therefore, if we have a solution to the second system, again by the same reasoning, it is a solution to the first system!

Exercise 1.2.6. Make sure you understand the previous two paragraphs. This is essential to our algorithm for solving systems of linear equations.

Of course, the previous example seemed like a useless thing to do. We took a system of two equations, manipulated a little bit and we ended up with some system which does not look any better. But in future we will make this manipulation in a clever way so that we end up in a *simpler* system to solve.

Exercise 1.2.7. Which of the following two systems is *simpler* to solve?

1.

$$\begin{aligned} 1x + 3y + 2z &= 0 \\ 2x + 3y + 1z &= 0 \\ 3x + 1y + 3z &= 0 \end{aligned}$$

2.

$$\begin{aligned} 1x + 0y + 0z &= 0 \\ 0x + 1y + 0z &= 0 \\ 0x + 0y + 1z &= 0 \end{aligned}$$

We make the habit of not writing the variable at all if the coefficient is zero. Therefore, the second system in the exercise can be written as

$$\begin{aligned} x &= 0 \\ y &= 0 \\ z &= 0. \end{aligned}$$

Note that there is nothing you need to do here to solve the system! The system is already in the simplest form it can be! There is only one solution and it is $(0, 0, 0)$. Our purpose is going to be to bring every system into a simpler system like this. We will see that it is not always going to be as simple as this one but we will try our best to bring it to its simplest form.

1.3. Consistent and inconsistent systems

1.4. Coefficient matrix and row operations

1.5. Echelon form

2. Different Languages

In Chapter 1, we have seen properties of systems of linear equations and how to solve them. In this chapter, we are going to be interested in representing questions about systems of linear equations in different languages. The primary purpose of this chapter is to introduce new vocabulary and showcase the importance of language.

2.1. Vector equations

In Chapter 1, we have discussed that the *names* of the variables do not play a role in our theory. If our variables are called x, y, z or x_1, x_2, x_3 or y_1, y_2, y_3, it does not matter. We will do the same thing to solve the system: consider the coefficient matrix and reduce it to an echelon form. We said that what matters is the coefficients.

In this section, we are going to see a different representation of a system of linear equations. Namely, a vector representation.

Definition 2.1.1. A *(column) vector in* $\mathbb{R}^n$ is an ordered n-tuple of real numbers.

Given a system of linear equations

$$\begin{array}{c} a_{11}x_1 + \ldots + a_{1n}x_n = b_1 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = b_m \end{array}$$

we can group the coefficients of each variable and represent this system as

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \ldots + \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Note that the coefficients formed vectors in $\mathbb{R}^m$ in this case. If we like, we can rearrange the same representation as

$$x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \ldots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

We can also give names to our vectors! Let us put

$$v_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} \quad , \quad \ldots \quad , \quad v_n = \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

Then, we get that our system of linear equations can be written as

$$x_1 v_1 + \ldots + x_n v_n = b$$

where

$$b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Again, recall that our purpose is to rewrite things in different format. Summarizing our discussion, we see that a system of linear equations

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\ \vdots \quad \vdots \quad \vdots \quad & \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

can also be represented as a vector equation

$$x_1 v_1 + \ldots + x_n v_n = b.$$

Exercise 2.1.2. What are some advantages of this representation at a first glance?

Example 2.1.3. Consider a system of linear equations

$$\begin{aligned} 2x + 4y + 5z &= 0 \\ 6x + 2y + 3z &= 0 \\ 2x + 3y + 4z &= 0. \end{aligned}$$

Then the corresponding vector equation is

$$x \begin{bmatrix} 2 \\ 6 \\ 2 \end{bmatrix} + y \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix} + z \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Now, there are some remarks that I have to make.

Remark 2.1.4. Do you realize that the "role" of the x, y, z has changed? At least it *looks like* the role has changed. In the first representation they look like they are the variables whereas in the second representation they look like they are coefficients in front of our vectors. Think this in terms of our main goal: writing things in a different language. With a simple rewriting trick, we managed to change our point of view. This is a very common and useful thing to do in mathematics.

Remark 2.1.5. Some people like to put arrows on top their vectors. I do not prefer to do this. I believe that if we do enough practice, we can understand from the context what letter represents a scalar and what vector represents a vector.

Remark 2.1.6. Continuing from the previous remark, I will represent *the zero vector* whose entries all zeroes by 0. So, the vector equation in the previous example will be written as

$$x \begin{bmatrix} 2 \\ 6 \\ 2 \end{bmatrix} + y \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix} + z \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix} = 0.$$

It may take some time to get used to this but from the context, you should be able to understand this 0 is the zero vector in $\mathbb{R}^3$.

Remark 2.1.7. The geometry of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ is important and we will revisit this from time to time. However, it is important to also change your point of view from a geometric object to a purely algebraic object. At this point, we will want to see a vector in $\mathbb{R}^3$ as a list of three real numbers as opposed to a point (or an arrow) in the three dimensional space.

2.2. Span of a set of vectors

From time to time, we will call $\mathbb{R}^n$ the Euclidean space. While $\mathbb{R}^n$ is the set of all vectors with n entries, it is not only a *set*. It has an algebraic structure on it which is the single most important fact we care about in linear algebra. You need to know three things:

1. Saying *two vectors are equal* is the same thing as saying *all entries of the two vectors are the same.*

2. You can add two vectors and this is done *coordinatewise.*

3. You can also multiply a vector by a scalar and this is also done *coordinatewise.*

Remark 2.2.1. We do not have a multiplicative structure on $\mathbb{R}^n$ in linear algebra. We only care about addition and scalar multiplication. Does this remind you of the first lecture?

Example 2.2.2. Let us put

$$v = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad , \quad w = \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix} .$$

Then,

$$v + w = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1+5 \\ 2+1 \\ 3+2 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \\ 5 \end{bmatrix} .$$

This is what we mean by coordinatewise addition.

Now, we will talk about a similar phenomenon which we will see in the future in more detail. The answer to this question is very easy. What I want you to focus on is: why did I ask you this, how is it related to our discussion?

Exercise 2.2.3. Consider two polynomials: $f = 2+3t+t^2$ and $g = 1+2t+4t^2$.

1. What is $2f$?
2. What is $f + g$?

The addition and scalar multiplication satisfies 8 rules that we care about:

1. For every three vectors u, v, w, we have $(u+v)+w = u+(v+w)$. This rule is called *associativity* and it basically tells us that we do not need parantheses when we do addition: if I give you $u+v+w$, you do not have to ask me "Should I do $u+v$ first or $v+w$ first?"
2. We have a *zero vector*. This is the vector that consists of all zeroes and we denote it by 0 as we have seen in the previous section. It has the property that $v + 0 = v$ for every vector v.
3. We have *additive inverses*. This works like *the negative of a number*.

   $$\text{if } v = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{then } \; -v = \begin{bmatrix} -1 \\ -2 \end{bmatrix} .$$

   For every vector v, there is a vector $-v$ such that $v + (-v) = 0$.
4. For every two vectors u, v, we have $u + v = v + u$. This is called *commutativity*.

Up to this point, we only dealt with rules about addition. Now, we will add rules about scalar multiplication and how it behaves with addition.

5. If you multiply any vector with the scalar 1, nothing changes. We have $1v = v$ for any vector v.
6. You can multiply two scalars with each other and then do scalar multiplication or you can do scalar multiplication twice and this does not change the result: for any two scalars a, b and any vector v, we have $(ab)v = a(bv)$.
7. We have a distrubition rule: if a, b are scalars and v is a vector, then $(a + b)v = av + bv$.
8. We have another distrubition rule: if a is a scalar and u, v are vectors, then $a(u + v) = au + av$.

Exercise 2.2.4. Choose some scalars and choose some vectors in $\mathbb{R}^3$. Verify these 8 rules.

Exercise 2.2.5. Appreciate associativity by remembering rock, paper, scissors. Define an operation on these by the winner of a match. (Example: rock*paper=paper).

1. Compute (rock*paper)*scissors.
2. Compute rock*(paper*scissors).

Is this a commutative operation?

Now, we will make some observations.

If we start with a vector v, using the two operations we have, which vectors can I create? Well, I have addition. So, I can do $v + v$. But

$$v + v = 1v + 1v = (1 + 1)v = 2v.$$

So, I ended up with a multiple of v, when I tried to add things. Let's try $v + 2v$. We have

$$v + 2v = 1v + 2v = (1 + 2)v = 3v.$$

Exercise 2.2.6. 1. Which of the 8 rules am I using when I do these?

2. Convince yourself that starting from a single vector, the allowed two operations give me multiples of v.

Next, let us start with a vector v and another vector w. I can create multiples of v and w as before and I can add them. So, starting from v and w, I can create $2v + 3w$ for example. I can multiply this with 2 to get $2(2v + 3w)$ but

$$2(2v + 3w) = 4v + 6w$$

(which rule did we use here?) and it is of the form $av + bw$ again. We can try to add two things of this form:

$$(2v + 3w) + (v + 2w) = 2v + 3w + v + 2w = 2v + v + 3w + 2w3v + 5w.$$

You see what is happening?

Exercise 2.2.7. Convince yourself that starting from two vectors v and w

1. you can create all vectors of the form $av + bw$ where a, b are scalars,
2. you can not create any other vectors.

Exercise 2.2.8. Convince yourself that starting from three vectors u, v and w

1. you can create all vectors of the form $au + bv + cw$ where a, b, c are scalars,
2. you can not create any other vectors.

We are now ready to make some definitions.

Definition 2.2.9. Let $v_1, \ldots, v_n$ be vectors.

1. A *linear combination* of $v_1, \ldots, v_n$ is a vector you can create using the two operations starting from these vectors: That is, a linear combination is a vector of the form $c_1v_1 + \ldots + c_nv_n$ where $c_1, \ldots, c_n$ are scalars.

2. The *span* of $v_1, \ldots, v_n$ is the set of all vectors you can create using the two operations starting from these vectors. In other words, it is the set of all linear combinations of these vectors.

We will now go back to our starting point. Recall from the previous section that a system of equations

$$\begin{array}{cccccc} a_{11}x_1 + \ldots + a_{1n}x_n = b_1 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = b_m \end{array}$$

can also be represented as a vector equation

$$x_1v_1 + \ldots + x_nv_n = b.$$

Then, the following sentences say the same thing in a different language.

1. The system

$$\begin{array}{c} a_{11}x_1 + \ldots + a_{1n}x_n = b_1 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = b_m \end{array}$$

   is consistent.

2. The system

$$\begin{array}{c} a_{11}x_1 + \ldots + a_{1n}x_n = b_1 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = b_m \end{array}$$

   has a solution.

3. There are numbers $x_1, \ldots, x_n$ such that the equations

$$\begin{array}{c} a_{11}x_1 + \ldots + a_{1n}x_n = b_1 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = b_m \end{array}$$

   hold true.

4. There are numbers $x_1, \ldots, x_n$ such that

$$x_1 v_1 + \ldots + x_n v_n = b.$$

5. The vector b is a linear combination of $v_1, \ldots, v_n$.

6. The vector b is in the span of $v_1, \ldots, v_n$.

There will be times where either of these six sentences will be more useful. In the next section, we will see another language.

2.3. Matrix-vector equations

We know that if we have a system of linear equations

$$\begin{array}{ccccccc} a_{11}x_1 & + & \ldots & + & a_{1n}x_n & = & b_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ a_{m1}x_1 & + & \ldots & + & a_{mn}x_n & = & b_m \end{array}$$

then its coefficient matrix is

$$\begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix}.$$

Let us call this matrix A. Now, in this we captured the coefficients. On the left hand side, we also have variables $x_1, \ldots, x_n$. Let us put them together as a column vector

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

and let us call this vector x. Note that I understand from the context that x is a vector. I do not like using different notation for vectors, as I previously mentioned, and unless it is not at all obvious from the context, I will not make an effort. Then, *the matrix-vector equation* representing our system of linear equations is simply denoted by

$$Ax = b.$$

Exercise 2.3.1. What do you think? You just used three letters and one equality sign to talk about the same thing!

Using only three letters is not the only benefit of this representation but I will not go into details at this point. Let us finish this section quickly by rewriting things in different language.

1. The system

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad &\vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

   is consistent.

2. There are $x_1, \ldots, x_n$ such that the equations

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad &\vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

   hold true.

3. There is a vector x such that the matrix-vector equation $Ax = b$ holds.
4. The equation $Ax = b$ has a solution.
5. The equation $x_1v_1 + \ldots + x_nv_n = b$ has a solution.
6. The vector b is in the span of $v_1, \ldots, v_n$.
7. The vector b is in the span of the columns of A.

2.4. Linear transformations

Before starting this section, make sure that you know what a function is. When I was a teaching assistant during my PhD for an engineering linear algebra course, we had weekly meetings with a *head TA* who helped us keep organized through different tutorial sections. I will never forget that this head TA gave us the following hint: “Oh, you can check if a linear transformation is onto like you do with functions” to which we responded “Umm, yes, because a linear transformation *is* a function?”. This shocked our head TA and their response was “but how do you check if it satisfies the vertical line test?”.

Exercise 2.4.1. Make sure that you know what a function is. I would be happy to chat with you about this and if I have more time, I will add an appendix to the end of these notes.

Now consider the system

$$\begin{aligned} 2x + 3y + 2z &= 14 \\ 3x + 2y + 4z &= 19. \end{aligned}$$

Exercise 2.4.2. Verify that $x = 1, y = 2, z = 3$ is a solution to this system.

If we want to write this system as a matrix-vector equation, we get

$$\begin{bmatrix} 2 & 3 & 2 \\ 3 & 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 14 \\ 19 \end{bmatrix}.$$

And the fact that $x = 1, y = 2, z = 3$ is a solution to this system means we have an equality

$$\begin{bmatrix} 2 & 3 & 2 \\ 3 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 14 \\ 19 \end{bmatrix}.$$

This, we see that this matrix *takes/transforms/sends/maps* the vector

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

to the vector

$$\begin{bmatrix} 14 \\ 19 \end{bmatrix}.$$

This is the first thing I want you to understand. Now, the second thing is something that looks trivial but is important. You know that two vectors are equal if they have the same entries. We talked about this in the first section of this chapter. So, combining the system of equations and the matrix-vector equation, we have that

$$\begin{bmatrix} 2 & 3 & 2 \\ 3 & 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 14 \\ 19 \end{bmatrix} = \begin{bmatrix} 2x + 3y + 2z \\ 3x + 2y + 4z \end{bmatrix}$$

and forgetting the vector in the middle, we get the equality

$$\begin{bmatrix} 2 & 3 & 2 \\ 3 & 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2x + 3y + 2z \\ 3x + 2y + 4z \end{bmatrix}.$$

You see what happened? Our matrix *transforms/sends/takes/maps* a vector

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

that lives in $\mathbb{R}^3$ to

$$\begin{bmatrix} 2x + 3y + 2z \\ 3x + 2y + 4z \end{bmatrix}$$

that lives in $\mathbb{R}^2$. In other words, this matrix defines a function from $\mathbb{R}^3$ to $\mathbb{R}^2$! Let us denote this function by f. Then, we say $f : \mathbb{R}^3 \to \mathbb{R}^2$ is a function from $\mathbb{R}^3$ to $\mathbb{R}^2$ and it is given by the rule

$$f \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2x + 3y + 2z \\ 3x + 2y + 4z \end{bmatrix}.$$

Exercise 2.4.3. Now, let us put

$$u = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \quad , \quad v = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$$

and let us take the function f from above.

1. Compute $f(u)$ and $f(v)$.
2. Compute $f(u) + f(v)$.
3. Compute $u + v$.
4. Compute $f(u + v)$.
5. Observe that $f(u) + f(v) = f(u + v)$. That is, it does not matter if you apply the function to u and v first (the first item) and then add (the second item) OR if you first add u and v (the third item) and then apply the function (the fourth item).

6. Compute $2f(u)$.
7. Compute $2u$.
8. Compute $f(2u)$.
9. Observe that $2f(u) = f(2u)$. That is, it does not matter if you first apply f to u and then multiply the result by 2 OR if you first multiply u by 2 and then apply f.
10. Now, choose different u, v and replace 2 with another number and repeat. You will see that the two observations you made are independent of your choice. It works for *all* choices of u, v and scalar.

Recall that we do have a structure on our sets of vectors. We can *add* and *multiply with scalars.* These operations are important to us and we want functions to *behave well* with respect to these structures.

Definition 2.4.4. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function.

1. If we have
$$f(u+v) = f(u) + f(v)$$
for every two vectors u and v, we say that f *respects* addition.
2. If we have
$$f(cv) = cf(v)$$
for every scalar c and every vector v, we say that f *respects* scalar multiplication.
3. We say that f is a *linear function* or *linear map* or *linear transformation* if it respects both structures.

You have already seen an example of a linear transformation. Now, I will give you a non-example.

Example 2.4.5. The function $f : \mathbb{R}^2 \to \mathbb{R}^3$ defined by the rule

$$f \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} xy \\ x \\ y \end{bmatrix}$$

is not a linear function. It does not respect scalar multiplication: indeed, we have

$$3f \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 6 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 18 \\ 6 \\ 9 \end{bmatrix} \neq \begin{bmatrix} 54 \\ 6 \\ 9 \end{bmatrix} = f \begin{bmatrix} 6 \\ 9 \end{bmatrix} =$$

Example 2.4.6. Let A be an $m \times n$ matrix. Then, the function $f : \mathbb{R}^m \to \mathbb{R}^n$ defined by the rule $f(v) = Av$ is a linear transformation. You already convinced yourself that this is true for

$$A = \begin{bmatrix} 2 & 3 & 2 \\ 3 & 2 & 4 \end{bmatrix}$$

in the exercise before the definition. The same arguments work in general to prove this more general statement.

Definition 2.4.7. Let A and B be two sets and $f : A \to B$ be a function. Then, an element $b \in B$ is said to be *in the image of* f if there is an $a \in A$ such that $f(a) = b$. The set of all elements in B consisting of all elements which are in the image of f is called the *image of* f. So,

$$\operatorname{im}(f) = \{b \in B \colon f(a) = b \text{ for some } a \in A\}.$$

Now, it is time for more observations. These are some crucial observations which will be very helpful in the future.

Exercise 2.4.8. Decide whether each of the following statements is true or false.

1. We can write

$$\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

2. We can write

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

3. If $e_i \in \mathbb{R}^n$ denotes the vector which has a 1 in the $i$the coordinate and 0 in other coordinates, we can write

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 e_1 + \ldots + x_n e_n$$

4. The vector space $\mathbb{R}^n$ is spanned by $\{e_1, \ldots, e_n\}$.

Now, in the definition of a linear transformation we said a linear function f respects addition and scalar multiplication. But then, it respects all linear combinations, too as one can easily see that

$$f(au + bv) = f(au) + f(bv) = af(u) + bf(v).$$

By the way, all four statements in the previous exercise are true. Then, using the fact that a linear mapping takes linear combinations to linear combinations and the item number 3 in the exercise, we get for all

$$v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

we can compute

$$f(v) = f(x_1 e_1 + \ldots + x_n e_n) = x_1 f(e_1) + \ldots + x_n f(e_n).$$

Therefore, if we have a linear transformation $f : \mathbb{R}^n \to \mathbb{R}^m$, then we have

$$\begin{aligned} w \in \text{im} f &\iff \text{there is a } v \in V \text{ such that } f(v) = w \\ &\iff \text{ there are } x_1, \dots, x_n \text{ such that } f(x_1 e_1 + \dots + x_n e_n) = w \\ &\iff \text{ there are } x_1, \dots, x_n \text{ such that } w = x_1 f(e_1) + \dots + x_n f(e_n) \\ &\iff w \in \text{span}\{f(e_1), \dots, f(e_n)\} \end{aligned}$$

Therefore,

$$\text{im} f = \text{span}\{f(e_1), \dots, f(e_n)\}.$$

Example 2.4.9. Consider the linear transformation $f : \mathbb{R}^2 \to \mathbb{R}^3$ defined by the rule

$$f \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2x + y \\ x + 2y \\ x - y \end{bmatrix}.$$

Then, our discussion above yields

$$\text{im} f = \text{span} \left\{ \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \right\}.$$

We are almost ready to tie things up. You have seen that the value of a linear function is determined by its action on the *standard basis vectors* $e_1, \dots, e_n$.

We will now see what an $m \times n$ matrix does to these vectors. Recall that if

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \quad \text{and} \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

then, we have

$$Ax = \begin{bmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n \end{bmatrix}$$

from our discussion in the previous section. The following exercise is crucial.

Exercise 2.4.10. Show that if A is an $m \times n$ matrix, then Ae_1 is equal to the first column of the matrix, Ae_2 is equal to the second column of the matrix and so on.

Ok, we are about to finish the discussion. Let us consider a linear transformation $f : \mathbb{R}^n \to \mathbb{R}^m$ and let us compute $f(e_1), \ldots, f(e_n)$. Note that these are vectors in R^m. Consider the matrix A whose columns are $f(e_1), \ldots, f(e_n)$.

Theorem 2.4.11. *For every $v \in V$, we have $f(v) = Av$.*

Exercise 2.4.12. By summarizing the discussion in this section, prove the theorem.

Definition 2.4.13. If f and A are as in the theorem, we say that A is *the matrix of f* (with respect to the standard basis).

Remark 2.4.14. At the beginning of this section, we said that multiplying by a matrix is a linear transformation. We now concluded that every linear transformation can be represented as multiplication by a matrix.

We will end this chapter by our favorite activity: writing things in a different language. Let us consider the system of linear equations

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots &\quad \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

and call it SYS. Let us put

$$A = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix}$$

as the coefficient matrix and consider the linear transformation $f : \mathbb{R}^n \to \mathbb{R}^m$ given by the rule $f(v) = Av$. Then, the following statement are equivalent.

1. SYS is consistent.
2. The matrix-vector equation $Ax = b$ has a solution.
3. The vector b is in the span of the columns of A.
4. The vector b is a linear combination of columns of A.
5. The vector b is in the image of f.
6. The vector b is spanned by $f(e_1), \ldots, f(e_n)$.

3. Homogeneous systems

Let us start quickly with a definition and leave the motivation for later.

Definition 3.0.1. A system of linear equations is called a *homogeneous system* if the corresponding matrix-vector equation is of the form $Ax = 0$. That is, if the right hand side of the system has $b_1 = \ldots, b_m = 0$.

The obvious nice thing about a homogeneous system is that it is always consistent. Indeed, we always have the *trivial solution* $x_1 = \ldots = x_n = 0$ to a homogeneous system.

Exercise 3.0.2. Check that $x = y = z = 0$ is a solution to the system

$$\begin{aligned} 3x + 4y + 2z &= 0 \\ 5x + 3y + 6z &= 0 \\ 2x + 5y + 4z &= 0. \end{aligned}$$

3.1. Why care about homogeneous systems

From now on, instead of writing systems of linear equations, I am going to use the corresponding matrix-vector equations and use the term system of linear equations vaguely.

Consider a homogeneous system of linear equations $Ax = b$. Let u, v be two solutions to this equation so that we have $Au = b$ and $Av = b$. Then, we see that

$$Au - Av = b - b = 0$$

or in other words, $u - v$ is a solution to the homogeneous equation $Ax = 0$ because $A(u - v) = Au - Av = 0$. On the other hand, if w is a solution to the homogeneous equation $Ax = 0$ and v is again a solution to the equation $Ax = b$, then we get that $v + w$ is also a solution to $Ax = b$ since

$$A(v + w) = Av + Aw = b + 0 = b.$$

I used too many "solution"s and "equations" in the above paragraph. Make sure that you read it again and actually understand what I said.

Theorem 3.1.1. *Consider an equation $Ax = b$ and suppose that you know one solution v to this equation. Then, every other solution to this equation is of the form $v + w$ where w is a solution to $Ax = 0$.*

Suppose that you go to an electronic store and they have two machines: the first machine can solve every system of linear equations (if consistent) and it is expensive; the second machine can solve only homogeneous systems of linear equations and it is cheaper. You want to solve a system of the form $Ax = b$. You can by hand find a *single solution*, make the second machine do the work to find the solution set to the homogeneous equation $Ax = 0$, and voilà, you have all the solutions to the $Ax = b$ without buying the expensive machine. So, systems of linear equations work like *keys*.

3.2. An algebraic structure

We are slowly getting used to linear algebra. In this section, we are going to see its essence.

From day 1, we said that what we care about is addition and scalar multiplication. This is the *structure* we care about. We will now see that the solution set of a homogeneous system inherits this structure. What do I mean by this?

Theorem 3.2.1. *Let $Ax = 0$ be a homogeneous system of equations.*

1. *The zero vector is a solution to this system.*
2. *If v, w are solutions to this system, then $v + w$ is also a solution to this system.*
3. *If v is a solution to this system and if c is a scalar, then cv is also a solution to this system.*

The proof of this theorem is just a summary of our previous discussions to together with an understanding of the definitions. I leave it as an exercise.

Exercise 3.2.2. Write a proof to theorem.

Remark 3.2.3. The theorem tells us that we can *do algebra* inside the solution set of a homogenous system of linear equations. We can add things, we can multiply things with scalars, we have a zero vector. Basically, every rule that we listed for addition and scalar multiplication in Section 2 of Chapter 2 work for the solution set of a homogeneous system.

Exercise 3.2.4. Take a moment, go outside, do some meditation or something and appreciate this remark.

3.3. Linear independence

We have seen that a homogeneous system is always consistent because $x = 0$ is always a solution to an equation $Ax = 0$. But we might have more solutions, maybe? From the first chapter, you know that you can check this by row reducing the coefficient matrix and seeing if you have any free variables.

Recall from the previous chapter that we have access to different languages to talk about the same thing. This time, I want to talk about things in terms of vector equations.

So, suppose that we have homogeneous system of linear equations and the corresponding vector equation is

$$c_1v_1 + \ldots + c_nv_n = 0.$$

Definition 3.3.1. We say that the vectors $v_1, \ldots, v_n$ are *linearly dependent* if there is a nonzero solution to this system. We say that they are *linearly independent* if there is no other solution than the trivial solution.

So, the game is simple. Again, the only possibly confusing part is to learn the new language, new terminology. You want to see if a set of vectors is linearly dependent or independent. What do you do? You consider the corresponding system of linear equations, better yet you consider the coefficient matrix of that system. Well, if you are comfortable with Chapter 2 and can translate between different languages easily, you immediately understand that what you do is to take these vectors and put them as columns in a matrix. Call this matrix A and reduce A to its echelon form: if you observe existence of free variables, that is if in the echelon form *not* every column has a leading entry, then your system has infinitely many solutions and therefore your vectors are linearly independent. Otherwise, that is if every column is a pivot column, then there is only one solution to this system and it is the trivial solution.

Let us next discuss the name a little bit. Where does the term *independence* come from?

Suppose that you have some vectors $v_1, \ldots, v_n$. Then,

$v_1, \ldots v_n$ are linearly dependent.

$$\Updownarrow$$

there is a nontrivial solution to $c_1v_1 + \ldots + c_nv_n = 0$

$$\Updownarrow$$

there are $c_1, \ldots, c_n$ (at least one of them is nonzero) such that $c_1v_1 + \ldots + c_nv_n = 0$

(Let's say that $c_1 \neq 0$. If it is zero, find the nonzero c_i and do the same thing to it).

$$\Updownarrow$$

there are $c_2, \ldots, c_n$ such that $v_1 + \frac{c_2}{c_1}v_2 + \ldots + \frac{c_n}{c_1}v_n = 0$. (Here I just divided by the scalar c_1. I am allowed to do this because we assumed $c_1 \neq 0$).)

$$\Updownarrow$$

there are $c_2, \ldots, c_n$ such that $v_1 = -\frac{c_2}{c_1}v_2 - \ldots - \frac{c_n}{c_1}v_n$.

$$\Updownarrow$$

v_1 is a linear combination of $v_2, \ldots, v_n$.

So, we can conclude with a theorem.

Theorem 3.3.2. *Vectors $v_1, \ldots, v_n$ are linearly dependent if and only if one of the vectors is in the span of the remaining ones. Equivalently, they are linearly independent if and only if none of the vectors can be written as a linear combination of the remaining ones.*

Exercise 3.3.3. Show that for any two vectors u, v, the set $\{u, v, u + v\}$ is linearly dependent.

Remark 3.3.4. Notice the language: sometimes I am saying that vectors are linearly independent and sometimes I am saying that the set of vectors is linearly independent. Both are correct uses.

3.4. Nullspace and kernel

Remember from the previous chapter that we have several different languages to express the same thing: systems of linear equations, vector equations, matrix-vector equations and linear functions. In this section, we will continue the theme of this chapter in the language of matrices and linear transformations. We start with a definition.

Definition 3.4.1. Let A be an $m \times n$ matrix and v be a vector in $\mathbb{R}^n$. Then, we say v *is in the nullspace of* A if $Av = 0$. The *nullspace* of A is the set of all vectors v in $\mathbb{R}^n$ such that $Av = 0$. We denote the nullspace of A by null$A$.

We have learned a new word. Let us see how it relates to what we already know. Suppose that we have a homogeneous system

$$\begin{array}{c} a_{11}x_1 + \ldots + a_{1n}x_n = 0 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = 0. \end{array}$$

We know that we can represent it as a vector equation

$$x_1v_1 + \ldots + x_nv_n = 0$$

or as a matrix vector equation

$$Ax = 0$$

where the vectors v_i's consists of the coefficients of x_i's in the system, the matrix A has v_i's as its columns and the vector x has entries x_i's.

Then, saying that x_i's form a solution to our homogeneous system is equivalent to saying that x is in the nullspace of A.

Exercise 3.4.2. Make sure that you understand the previous sentence.

If you understood the previous sentence, then you can prove the following theorem.

Theorem 3.4.3. *Let A be an $m \times n$ matrix and $v_1, \ldots, v_n$ be columns of A. Then, saying that $v_1, \ldots, v_n$ are linearly independent is equivalent to saying that the nullspace of A only contains the zero vector* ($\text{null}A = \{0\}$)*.*

Exercise 3.4.4. Prove the theorem. (Really, just write a couple paragraphs summarizing previous discussions to conclude the theorem).

Exercise 3.4.5. Read the paragraph after Definition 3.3.1, look at the above theorem, remember your knowledge about row-reducing and understand how these relate to each other.

In the previous chapter, you have seen that there is a correspondence between $m \times n$ matrices and linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. Now, let us say everything about homogeneous systems in terms of linear functions.

Definition 3.4.6. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then, we say that a vector $v \in \mathbb{R}^n$ is *in the kernel of f* if $f(v) = 0$. We call the set of all $v \in \mathbb{R}^n$ with the property $f(v) = 0$ the *kernel of f*. We denote the kernel of f by $\ker f$.

Now, you should be able to connect all the dots.

Exercise 3.4.7. Show that 0 is always in $\ker f$ for every linear function f by following the following questions.

1. What is $0 + 0$?

2. Is it true that $f(0+0) = f(0) + f(0)$?

If you need more hints, let me know.

Now, we know that the kernel of a linear transformation always contains the zero vector. What if this linear transformation is special and the kernel *only* contains the zero vector and nothing else?

Exercise 3.4.8. Let f be a linear transformation and A be its matrix. Show that the columns of v are linearly independent if and only if $\ker f = \{0\}$.

Definition 3.4.9. Let A, B be two sets (not necessarily vector spaces). We say that a function $g : A \to B$ is *one-to-one* or *injective* if g does not map two different elements of A to the same element of B.

While this definition is more intuitive, it is easier to check injectivity as follows: firstly, understand that the definition says g is injective if and only if $x \neq y$ implies $g(x) \neq g(y)$. So, *if the inputs are different, then the outputs are different.* We can rewrite this as *if the outputs are the same, then the inputs are the same.* So, in practice, we use the following alternative definition.

Definition 3.4.10 (Alternative definition)**.** We say that $g : A \to B$ is injective if $g(x) = g(y)$ implies $x = y$ for every $x, y \in A$.

Now, let us consider linear functions between two vector spaces. Suppose that I have a linear transformation $f : \mathbb{R}^n \to \mathbb{R}^m$. Then,

f is one-to-one

$\Updownarrow$

for every $u, v \in \mathbb{R}^n$, we have $f(u) = f(v)$ implies $u = v$.

$\Updownarrow$

(Since we can do addition and subtraction in vectorspaces)
for every $u, v \in \mathbb{R}^n$, we have $f(u) - f(v) = 0$ implies $u - v = 0$.

$\Updownarrow$

(Since f is linear)
for every $u, v \in \mathbb{R}^n$, we have $f(u - v) = 0$ implies $u - v = 0$.

$\Updownarrow$

(Putting $x = u - v$)

$$\text{for every } x \in \mathbb{R}^n, \text{ we have } f(x) = 0 \text{ implies } x = 0.$$
$$\Updownarrow$$
$$\ker f = \{0\}.$$

Theorem 3.4.11. *Saying that a linear transformation f is one-to-one is equivalent to saying that* $\ker f = \{0\}$.

Exercise 3.4.12. We already proved this theorem right above, the exercise is to make sure you understand each step.

The following exercise is the essence of this entire chapter.

Exercise 3.4.13. Write an essay which talks about the relations between the following concepts:

1. free variables in a system of linear equations,
2. linear independence of columns of a matrix,
3. one-to-oneness of a linear transformation.

4. Matrix Operations

At this point in the course, you are assumed to be more familiar with matrices. In this chapter, you are going to see some algebraic structures on the space of matrices. The first two structures are familiar: *addition* and *scalar multiplication*. You will notice that the rules of addition and scalar multiplication are also familiar. Then, you will see an additional structure: *multiplication*.

4.1. Addition, scalar multiplication

These familiar operations of addition and scalar multiplication are defined *coordinatewise*. You are going to explore these two operation through exercises.

Exercise 4.1.1. Do CTRL-F (or your equivalent of the search function) and find where we used the word *coordinatewise* before in the notes. Then, try to understand what I mean by the above definition.

Exercise 4.1.2. Pick some 2×3 matrices and add them. Pick some scalars and perform scalar multiplication with some matrices. Familiarize yourself with these operations.

Exercise 4.1.3. In Chapter 2, you have learned that the addition and scalar multiplication on $\mathbb{R}^n$ satisfy 8 basic rules.

1. Go back and read those 8 rules.
2. Show that the same 8 rules hold for addition and scalar multiplication of matrices!

4.2. Composition of linear functions and definition of matrix multiplication

Before you start with this section, you need to go back and refresh your knowledge from Chapter 2.

Exercise 4.2.1. Redo Exercise 2.4.10.

Exercise 2.4.10 and the discussion afterwards tell you that the columns of the matrix of a linear transformation given by the images of the standard basis vectors under the linear transformation.

Now, let us consider two linear transformations $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^n$. Then, we can consider the composition $g \circ f$.

Exercise 4.2.2.

1. Convince yourself that $g \circ f$ is a function from $\mathbb{R}^n$ to $\mathbb{R}^k$.
2. Convince yourself that $g \circ f$ is also linear.

Let B be the matrix of f. This means that for every $v \in \mathbb{R}^n$, we have $f(v) = Bv$. Note that B is an $m \times n$ matrix.

Similarly, let A be the matrix of g. This means that for every $w \in \mathbb{R}^m$, we have $g(w) = Aw$. Note that A is an $k \times m$ matrix.

Now, since $g \circ f$ is also a linear transformation, it should also have a corresponding matrix. We have

$$(g \circ f)(v) = g(f(v)) = g(Bv) = A(Bv)$$

for every $v \in \mathbb{R}^n$. Make sure that you understand this because it is the crucial step. Up to this point, we only wrote down what we already know.

Definition 4.2.3. Let A, B, g, f be as above. The matrix product AB is the matrix with the property $(AB)v = A(Bv) = (g \circ f)v$ for every $v \in \mathbb{R}^n$. That is, the product AB is the matrix of $g \circ f$.

Exercise 4.2.4. Which one of the following is the correct size for AB according to our discussion?

(a) $m \times n$

(b) $m \times k$

(c) $n \times m$

(d) $n \times k$

(e) $k \times m$

(f) $k \times n$

And why?

We made a conceptual definition. Now, let us investigate it from a more computational point of view. This is why you were asked to redo Exercise 2.4.10 at the beginning of this section. We will compute the product AB by computing each column of it.

According to Exercise 2.4.10, the first column of AB is equal to ABe_1 where e_1 is the first standard basis vector. According to our conceptual definition of matrix multiplication, we must have

$$ABe_1 = A(Be_1).$$

Again from Exercise 2.4.10, we do know that Be_1 is the first column of B. So,

$$ABe_1 = A(\text{first column of } B).$$

However, you know from the section on matrix-vector equations how to multiply a matrix with a vector. So, you are done! Okay, too fast? Let's see an example.

Example 4.2.5. Let's compute the first column of AB where

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 2 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 \\ 5 & 4 \\ 6 & 2 \end{bmatrix}.$$

According to what we said above, the first column should be

$$ABe_1 = A(Be_1) = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 2 & 3 \end{bmatrix} \left(\begin{bmatrix} 1 & 2 \\ 5 & 4 \\ 6 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \\ 6 \end{bmatrix}.$$

And you know from our discussion in Chapter 2 on matrix-vector equations that

$$\begin{bmatrix} 2 & 3 & 4 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \times 1 + 3 \times 5 + 4 \times 6 \\ 1 \times 1 + 2 \times 5 + 3 \times 6 \end{bmatrix} = \begin{bmatrix} 41 \\ 29 \end{bmatrix}.$$

If you do not remember the discussion from Chapter 2, please go back and read it. Do not just read this example without understanding, please.

We found the first column of the product by using the first standard basis vector, we will use the second standard basis vector to find the second column. That is, the second column of AB will be equal to ABe_2.

Exercise 4.2.6. Consider the two matrices from the previous example. Show that the second column of AB is

$$\begin{bmatrix} 2 & 3 & 4 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}$$

which is equal to

$$\begin{bmatrix} 24 \\ 16 \end{bmatrix}.$$

Conclude that

$$\begin{bmatrix} 2 & 3 & 4 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 5 & 4 \\ 6 & 2 \end{bmatrix} = \begin{bmatrix} 41 & 24 \\ 29 & 16 \end{bmatrix}.$$

Let us play with this example a little bit more. Again, consider 41 which is the first row first column entry of AB. How did we find it? If you look at our computations, essentially what we did was

take the first row of A and the first column of B

and multiply the entries one by one and add them up.

$$\begin{bmatrix} 2 & 3 & 4 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 5 & 4 \\ 6 & 2 \end{bmatrix} = \begin{bmatrix} 2 \times 1 + 3 \times 5 + 4 \times 6 & \\ & \end{bmatrix}.$$

Now, following the same ideas, we get the next theorem.

Theorem 4.2.7. *Let A be an $m \times n$ matrix and B be an $n \times k$ matrix. Then, AB is an $m \times k$ matrix whose ith row jth column entry is given by multiplying the entries of ith row of A with entries of jth column of B and adding them up.*

Exercise 4.2.8. Convince yourself that this is what we did in the above example and what we described in the above discussion.

Exercise 4.2.9. Illustrate the theorem in the previous example by filling up the remaining three entries using this method and comparing with the correct result.

Remark 4.2.10. Note that matrix multiplication takes two inputs: an $m \times n$ matrix and an $n \times k$ matrix and the result is an $m \times k$ matrix. So, if we take two $n \times n$ matrix, then their product is also an $n \times n$ matrix. So, matrix multiplication defines an operation *on* the space of $n \times n$ matrices.

4.3. Properties of matrix multiplication

Matrix multiplication behaves like you would expect it to behave in many ways and behaves like you would not expect it to behave in many other ways.

Firstly, matrix multiplication is associative. That is, for any three matrices A, B, C (of correct sizes so that we can actually perform the multiplication) we have

$$A(BC) = (AB)C.$$

There are at least three ways to prove this:

1. You can say that matrix multiplication corresponds to composition of linear functions and we know that composition is associative.

2. You can argue by finding the columns of both matrices by multiplying them with standard bases vectors.

3. You can argue by computing ijth entry of both matrices.

I am not going to ask you to prove this rigorously because I do not think that there is value in it. But I want you to think about all three ways until you can convince yourself that you *could* write a rigorous proof if your life depended on that.

Secondly, matrix multiplication distributes over addition. If A, B, C are three matrices (of correct sizes so that operations make sense), then we have

$$\begin{aligned} A(B+C) &= AB + AC \\ (A+B)C &= AC + BC. \end{aligned}$$

Example 4.3.1. Pick three randomly chosen matrices and verify the two properties we have discussed.

Unfortunately, matrix multiplication is not commutative. That is, usually we do not have $AB = BA$.

Exercise 4.3.2. Show that if A is a 2×3 matrix and B is a 3×4 matrix, we can not have $AB = BA$. (Hint: has a silly reason).

Exercise 4.3.3. Let

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 4 \\ 3 & 5 \end{bmatrix}.$$

1. Compute AB. What is the action of A on B in this case?
2. Compute BA. What is the action of A on B in this case?
3. Is AB equal to BA?

Remark 4.3.4. The next question is a very interesting question for me. Not that the content of the question is very interesting but my experience with student's responses is. The question is very simple *once* you see what you are supposed to be doing. But in order to see what you are supposed to be doing, you need to start playing with the question. For a second, forget that you are trying to solve the question. Just try to understand what is up with the question. Once you understand the question, the answer will be in front of you, anyways.

If you can not solve it, that's okay. Contact me and we will figure it out together.

Exercise 4.3.5. Find all matrices A such that $AB = BA$ where B is the matrix

$$B = \begin{bmatrix} 2 & 4 \\ 3 & 5 \end{bmatrix}.$$

Now, one thing you would expect from multiplication is to have an identity element. Indeed, in real numbers we do have an identity element, namely 1. For every real number x, we have $1x = x$. The question is: do we have this in matrix multiplication? Can we find a matrix I such that for every matrix A (of correct size) you have $AI = A$ (or for every matrix B of correct size, you have $IB = B$)?

You can approach this question in several different ways but I think the cleanest way is to realize that the identity transformation $\mathrm{id} : \mathbb{R}^{\mathrm{n}} \to \mathbb{R}^{\mathrm{n}}$ defined by $\mathrm{id}(\mathrm{v}) = \mathrm{v}$ is a linear transformation. Note also that if f is any function from $\mathbb{R}^m$ to $\mathbb{R}^n$, we would have $\mathrm{id} \circ f = f$. Indeed the rule of id is *do nothing*. You can apply f and stop OR you can apply f, then wait and do nothing and then stop. The result will not change. So, the matrix we are looking for is the matrix of the identity transformation with respect to the standard basis. And do you remember how we find the matrix of a linear map?

We should have the first column of I is $Ie_1 = \mathrm{id}e_1 = e_1$! The second column of I should be e_2 and so on. So, in the 3×3 case I should look like:

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

Exercise 4.3.6. Show that for any $m \times n$ matrix A and any $n \times k$ zero matrix, the product $A0$ is the zero matrix.

4.4. Interesting things

Now, we will see some interesting things about matrix multiplication. I will give them as exercises.

Exercise 4.4.1. Consider the matrices

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

1. Is A the zero matrix?
2. Is B the zero matrix?
3. Compute AB. What do you observe? Does this happen in multiplication of real numbers?

Remark 4.4.2. How did you do the matrix multiplication? Did you multiply the matrix A with the columns of B and used ideas from matrix-vector equation from Chapter 2, or did you use the "ijth entry of the product AB is given by the ith row of A and jth column of B" principle? I think using the first idea is much quicker here, it immediately tells me that the first column of AB is zero and the second column of AB is equal to 'the second column of A'.

Exercise 4.4.3. Consider the matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

1. Is A the zero matrix, is it the identity matrix?
2. Is B the zero matrix, is it the identity matrix?
3. Compute AB and compute BA. Write two observations.
4. Are B and C equal?
5. Compute AC and compare it with AB.

Exercise 4.4.4. Let A be the 2×2 matrix whose first column is zero and whose second column is e_1.

1. Is A the zero matrix?
2. Compute A^2.
3. Write a one sentence slogan for your observation.

Exercise 4.4.5. Let A be the 3×3 matrix whose first column is zero, second column is e_1 and third column is e_2.

1. Compute A^2.
2. Compute A^3.

Exercise 4.4.6. In at most 5 trials, try to find a 3×3 matrix such that $A^3 \neq 0$ but $A^4 = 0$. If after 5 trials you couldn't find it, then skip the question. We will return to this later.

Exercise 4.4.7. Show that if $AB = 0$, then the columns of B are in the nullspace of A.

4.5. Transpose

The next operation we will learn is the transpose operation which is a *unary* operation. While there are several reasons to learn about transpose of a matrix, I am not sure what the best way is to motivate this, the best way to start.

4.6. Invertible matrices

We have seen that if we take two nonzero real numbers x and y then their product xy is nonzero but we may take two nonzero matrices X, Y such that $XY = 0$. Remember this means that the columns of Y are in the nullspace of X. Let us rewrite the situation in real numbers in different words: If $xy = 0$ and $x \neq 0$, then we must have $y = 0$. Make sure that you understand this is the exact same thing as the first sentence. Let us look further into this:

Suppose that $xy = 0$ and $x \neq 0$. Since $x \neq 0$, we know that there is a number $1/x$. The property of this number is that when you multiply with x, you get 1. So, you can do:

$$xy = 0 \implies \frac{1}{x}(xy) = \frac{1}{x}0$$

by multiplying with $1/x$ on both sides and you would then get

$$\frac{1}{x}(xy) = \frac{1}{x}0 \implies \left(\frac{1}{x}x\right) y = 0$$

by using associativity of multiplication for the left hand side and the fact that when you multiply with zero, you get zero. Then, you get the next step:

$$\left(\frac{1}{x}x\right) y = 0 \implies 1y = 0 \implies y = 0$$

and this is because of the property of $1/x$ mentioned above and the fact that when you multiply with 1, nothing changes.

Now, in the space of matrices, we do have multiplication. But we also know that it is a little bit weird. Like you could get $AB = 0$ while $A \neq 0$ and $B \neq 0$. So apparently our chain of thoughts above do not work in the case of matrices. We do have associativity for matrix multiplication and we know that if you multiply with the zero matrix, you would get zero. We do have a matrix, namely the identity matrix, that acts like the number 1. So, looking at the discussion above there is only one explanation, we may not have a "$1/A$" for matrices. Later, I will warn you that this is a very bad notation and this is the reason I put this in quotation marks.

Let us rewrite the same thing: If we had two matrices A and B such that $AB = 0$ and if A had an inverse C such that $CA = I$, then we would have to have

$$AB = 0 \implies C(AB) = 0 \implies (CA)B = C0 \implies IB = 0 \implies B = 0$$

but as we have seen we may have a situation where $A \neq 0, B \neq 0$ and $AB = 0$. What is the conclusion? Not all matrices have inverses.

Definition 4.6.1. We say that a matrix A is *invertible* if there exists a matrix C such that $AC = CA = I$.

Now, this is a very careless definition to be honest. I did not say anything about the sizes of these matrices. But you can see that if $AC = CA$, then they both need to be square matrices. Let us actually write it as an exercise.

Exercise 4.6.2. Show that if A is an $m \times n$ matrix and C is a $k \times l$ matrix with the property (i) AC is defined (ii) CA is defined and (iii) $AC = CA$, then $m = n = k = l$.

Before figuring out what C looks like, let us give it a name.

Definition 4.6.3. If C is as in the previous definition, we call it *an inverse* of A.

Remark 4.6.4. Here is an important detail: I used *an inverse* instead of *the inverse*. This suggests that there may be more than one inverse. As you have seen, matrix multiplication is different than multiplication of numbers. So, who knows, maybe this is one of the weird things about it? Maybe there are multiple inverses for a matrix? This is something that we need to investigate.

Theorem 4.6.5. *If a matrix A is invertible, then its inverse is unique. That is, there is only one inverse. Therefore, we can call it* ***the*** *inverse and denote it by A^{-1}.*

Proof. Let us suppose that we have two matrices C and D such that

$$AC = CA = I \text{ and } AD = DA = I.$$

Then, answer the following questions:

1. **True or False:** $C = CI$.
2. **True or False:** $CI = C(AD)$.

3. **True or False:** $C(AD) = (CA)D$.
4. **True or False:** $(CA)D = ID$.
5. **True or False:** $ID = D$.

If you have answered true for all of them, reading all the equations from the top to the bottom, you will see that $C = D$. This shows that you can not find two *different* matrices which are inverses of A. □

Remark 4.6.6. We do not use the notation $1/A$ for matrices. Here is a reason for it: If you have two matrices A, B and if you know that A is invertible, then when you write B/A there is no way to understand whether you mean $B(1/A)$ or $(1/A)B$ and as we have seen earlier, matrix multiplication is not commutative and usually these two are different.

Now, let us figure out more about this inverse. For instance, how do you know if it exists and if it does exist, how do you find it?

Remember when we talked about matrix multiplication, we talked about it in detail. I did not give you a quick rule for matrix multiplication and told you to memorize it. Instead, we went through a process to understand how it works.

Exercise 4.6.7. Reread Section 4.2 before you continue.

Let's say we have a square matrix A and we want to see if it has in inverse and if it does we want to find it. In other words, we would like to solve the equation $AX = I$ where X is also a square matrix. Let us denote by $v_1, v_2, \ldots, v_n$ the columns of X. Then, if you understand Section 4.2 well enough, you should be comfortable with the following:

$$AX = I \iff Av_1 = e_1, Av_2 = e_2, \ldots, Av_n = e_n.$$

In other words, saying that A is invertible is equivalent to saying that for every $i = 1, \ldots, n$, the matrix-vector equation $Ax = e_1$ has a solution. How do we find this solution? Well, we make an augmented matrix $[A \mid e_i]$ if you remember the first week of classes. Then, we reduce A to a reduced row echelon form and we got it, whatever ends up on the right hand side of the augmentation line is the solution. Cool. So, in order to find the inverse, you should do this for all n equations above. That is, you should row reduce the augmentation matrices

$$[A \mid e_1] \quad , \quad [A \mid e_2] \quad , \quad \ldots \quad , \quad [A \mid e_n].$$

Now, while it looks like you will have to do n row reductions, if you actually try to do this, you will quickly realize that row reductions only depend on the matrix A and for all of these, you will do the same row reductions. Therefore, instead of doing n row reductions, put $e_1, \ldots, e_n$ on the right hand side of your augmented matrix and do all the row reductions at once.

To sum up, if you want to see if A is invertible and if you want to compute its inverse, then you should construct the augmented matrix $[A \mid I]$ and row reduce it until your augmentation matrix turns into $[I \mid X]$. Of course, you can not always do this (some systems are inconsistent) and then your matrix is not invertible. In case your matrix is invertible, the matrix X you get on the right hand side is your inverse.

Example 4.6.8. Let us illustrate these ideas on an example. Consider

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

We want to find a matrix X such that $AX = I$. So, we want to find a, b, c, d such that

$$\begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

By our definition/discussion of matrix multiplication, we should then try to solve the two equations

$$\begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
$$\begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

but I can do this simultaneously by constructing my augmentation matrix as

$$\begin{bmatrix} 3 & 2 & \mid & 1 & 0 \\ 1 & 1 & \mid & 0 & 1 \end{bmatrix}$$

and row reducing this to

$$\left[\begin{array}{cc|cc} 3 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 3 & 2 & 1 & 0 \\ 0 & 1 & -1 & 3 \end{array}\right]$$
$$\to \left[\begin{array}{cc|cc} 3 & 0 & 3 & -6 \\ 0 & 1 & -1 & 3 \end{array}\right]$$
$$\to \left[\begin{array}{cc|cc} 1 & 0 & 1 & -2 \\ 0 & 1 & -1 & 3 \end{array}\right]$$

So, we get that A is invertible and

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

Okay, we know when a matrix is invertible and how to find the inverse. Good. But what do we do this information?

Let's consider a matrix-vector equation $Ax = b$. Suppose that you know A is invertible. Then, multiplying both sides on the left by A^{-1} you can get

$$Ax = b \implies A^{-1}Ax = A^{-1}b \implies x = A^{-1}b.$$

You have a unique solution and the solution is given simply by $A^{-1}b$. You can also do the same thing in matrix equations. Suppose that there is a matrix equation $AB = C$ where A, B, C are square matrices. You know A^{-1} and you know C. Then, by a similar argument you would get $B = A^{-1}C$.

As it is our tradition, let us write down things in different words again.

Theorem 4.6.9. *The following are equivalent for an $n \times n$ matrix A.*

1. *A is invertible.*
2. *For every $b \in \mathbb{R}^n$, the equation $Ax = b$ has a unique solution.*
3. *A can be row-reduced to the identity matrix.*
4. *Standard basis vectors $e_1, \dots, e_n$ can be written as a linear combination of the columns of A.*
5. *Every vector in $\mathbb{R}^n$ can be written as a linear combination of the columns of A.*

6. *When you row reduce A to its echelon form, every column has a leading one (a pivot entry).*
7. *The linear transformation $x \mapsto Ax$ is one-to-one and onto.*
8. *Columns of A are linearly independent.*

Exercise 4.6.10. Convince yourself that all of the statements in the theorem are equivalent.

5. Determinant

5.1. Examples and computations

5.2. Geometric properties

6. Exercises

You finished the first part of the course, congratulations. So far, you hopefully realized all the chapters were connected to each other. So, it is a good time to stop and test your knowledge.

Exercise 6.0.1. Solve all the exercises within the sections so far.

Exercise 6.0.2. Construct a system of linear equations whose solution set is spanned by the vector

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Explain how you determine the minimum number of equations in a system like above.

Exercise 6.0.3.

1. Find a set of vectors which span $\mathbb{R}^3$.
2. Find another set of vectors which span $\mathbb{R}^3$.
3. Find a set of vectors which span $\mathbb{R}^4$ and linearly independent.
4. Find a set of vectors which span $\mathbb{R}^4$ and linearly dependent.

Exercise 6.0.4. Identify each of the following statements as True or False.

1. If a system of linear equations has more equations than variables, then the system must have a unique solution.

2. If a system has more variables than equations, then the system must have infinitely many solutions.
3. Homogeneous systems are always consistent.

Exercise 6.0.5. What is a relationship between one-to-one functions and linearly independent vectors?

Exercise 6.0.6. What is a relationship between onto functions and consistency of linear systems?

Exercise 6.0.7. For each of the following, either give a numerical example or write a paragraph explaining such a function does not exist.

1. a linear function from $\mathbb{R}^5$ to $\mathbb{R}^3$ which is onto.
2. a linear function from $\mathbb{R}^5$ to $\mathbb{R}^4$ which is one-to-one.
3. a linear function from $\mathbb{R}^5$ to $\mathbb{R}^3$ which is not onto.
4. a linear function from $\mathbb{R}^4$ to $\mathbb{R}^5$ which is onto.
5. a linear function from $\mathbb{R}^4$ to $\mathbb{R}^5$ which is one-to-one.
6. a linear function from $\mathbb{R}^3$ to $\mathbb{R}^5$ which is not one-to-one.

Exercise 6.0.8. What is a diagonal matrix and what can you say about the product of diagonal matrices?

Exercise 6.0.9. If $u, v, w \in \mathbb{R}^n$ are linearly independent vectors, then what can you say about $u + v, u - v, v + w$? Do they also have to be linearly independent or can you choose u, v, w so that these are linearly dependent?

Exercise 6.0.10. Consider the matrix

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

1. Compute $A^2, A^3, \ldots, A^9$.
2. Let

$$v = \begin{bmatrix} 2 \\ 3 \\ 1 \\ 6 \\ 2 \end{bmatrix}$$

   and compute $Av, A^2v, A^3v, A^4v, A^5v$.

3. Let f be the polynomial function given by the rule $f(x) = 2 + 3x + x^2 + 6x^3 + 2x^4$. Compute $f'(x), f''(x), f'''(x)$ and a couple more.
4. Compare your answers to the previous two parts. What do you observe?

Exercise 6.0.11. Show that if A is an invertible matrix such that $A^2 = A+I$, then $A^{-1} = A - I$.

Part II

Vector Spaces

7. Introduction to Vector Spaces

At the beginning of these notes, we have emphasized over and over again: we are considering the structures of addition and scalar multiplication here. We have seen that the space $\mathbb{R}^n$ carries this structure. We can add vectors and we can multiply vectors with scalars *in a meaningful way*. What is that meaningful way? Well, they satisfy some *rules* or *axioms*. Addition and scalar multiplication behave nicely on $\mathbb{R}^n$.

Exercise 7.0.1. Do you remember the 8 rules we have discussed in previous chapters? If yes, can you write them down? If not, go for a hunt and locate those paragraphs where we discuss these 8 rules. Do not continue reading before you do so.

The fact that we have this additional structure on $\mathbb{R}^n$ allows us to talk about linear combinations of vectors, we can talk about linear independence, we can talk about special functions which respect this extra structure. We have observed that the solution set of a homogeneous system of equations also inherits the structure. In this part, we will talk about abstract ideas related to addition and scalar multiplication which allows us to do the same thing on different landscapes.

You have actually seen one of these landscapes. Consider the space $\mathrm{Mat}_{m\times n}$ of $m \times n$ matrices. We know that we can add matrices and we can multiply matrices with scalars. And in Exercise 4.1.3, you have showed that the 8 rules of addition and scalar multiplication are satisfied. Then, I can talk about *linear combination of matrices* for example. Given two $m \times n$ matrices A and B, you can make sense of something like $2A + 3B$. You have also seen the *transpose* of a matrix which takes an $m \times n$ matrix A and gives an $n \times m$

matrix A^T. Therefore, transpose defines a function:

$$T : \mathrm{Mat}_{m\times n}(\mathbb{R}) \to \mathrm{Mat}_{n\times m}(\mathbb{R})$$

defined by the rule $T(A) = A^T$. And you know from properties of transpose that for any two $m\times n$ matrices A, B, you have $(A+B)^T = A^T+B^T$ which says that this function T we just defined *respects* the additive structure on these spaces. We have $T(A+B) = T(A)+T(B)$. Similarly, we have $T(cA) = cT(A)$ because you know that $(cA)^T = cA^T$. Therefore taking transpose is a *linear transformation* from $\mathrm{Mat}_{m\times n}(\mathbb{R})$ to $\mathrm{Mat}_{n\times m}(\mathbb{R})$.

Exercise 7.0.2.

1. Make sure you understand the above argument.
2. For $m = 2$ and $n = 3$, pick two $m \times n$ matrices A and B and verify that transpose respects addition.

So, we will keep the rules of the game we played in the first part of the course but we will change the players from now on.

7.1. Definition and Examples

In order to define a vector space, we need two operations. The first of these operations should be a *binary* operation meaning that it takes two inputs from our space and gives an output. Addition of vectors and addition of matrices is a binary operation. You add *two* vectors and get a vector as a result. You add *two* matrices and get a matrix as a result. We will call this operation *addition* but you will see in some examples there are weird operations which we call addition because it is a binary operation which satisfy the rules we want. This is just a name.

The second operation should take a scalar and one input from your space and should spit out an output from your space. We will call this operation *scalar multiplication*.

Definition 7.1.1. A *vector space* V is a set equipped with two operations which we call addition and scalar multiplication as above satisfying the following 8 axioms.

1. For every $u, v, w \in V$, we have

$$(u + v) + w = u + (v + w).$$

2. There exists a special element, which we call the zero vector and denote by 0, such that for every $v \in V$, we have
$$0 + v = v + 0 = v.$$

3. For every $v \in V$, there is a special element which we denote by $-v$ such that
$$v + (-v) = (-v) + v = 0$$

4. For every $u, v \in V$, we have
$$u + v = v + u.$$

5. For every $v \in V$, we have
$$1v = v.$$

6. For every scalar c and $u, v \in V$, we have
$$c(u + v) = cu + cv.$$

7. For every two scalars c, d and $v \in V$, we have
$$(c + d)v = cv + dv.$$

8. For every two scalars c, d and $v \in V$, we have
$$(cd)v = c(dv).$$

You can immediately observe two things: Firstly, these are very *natural* axioms. To be able to do some *good behaving algebra* these 8 rules are necessary (sort of, but we are not doing higher level mathematics now so we can believe this). These 8 rules are exactly copied from the properties of $\mathbb{R}^n$ with its usual addition and scalar multiplication. Thus, we *made* this definition so that we can do things which we did with vectors in $\mathbb{R}^n$. Let us see a couple familiar examples that you have seen before.

Example 7.1.2. $\mathbb{R}^n$ with usual addition and scalar multiplication is a vector space.

Example 7.1.3. The space of matrices $\mathrm{Mat}_{m\times n}(\mathbb{R})$ is a vector space.

Example 7.1.4. Consider the space of polynomials $P_2(\mathbb{R})$ with polynomials of degree at most 2. Elements of this set are *things* of the form

$$a_0 + a_1x + a_2x^2$$

where a_0, a_1, a_2 are real numbers and x is a symbol. We call these *things* polynomials. Note that we do not see these *things* as functions [1]. We say that two polynomials are equal if they have the same coefficients. We define addition as follows:

$$(a_0 + a_1x + a_2x^2) + (b_0 + b_1x + b_2x^2) = a_0 + b_0 + (a_1 + b_1)x + (a_2 + b_2)x^2$$

and we define scalar multiplication as follows:

$$c(a_0 + a_1x + a_2x^2) = ca_0 + ca_1x + ca_2x^2.$$

It is then easy to show that the 8 axioms are satisfied. It is maybe worth noting that the zero vector in this case is the *zero polynomial* $0 + 0x + 0x^2$ which we will simply denote by 0.

Exercise 7.1.5. Show that $P_2(\mathbb{R})$ is a vector space. That is, check all 8 axioms are satisfied.

Example 7.1.6. Let us denote by $C[0, 1]$ the set of all continuous functions from the closed interval $[0, 1]$ to $\mathbb{R}$. You have seen from your calculus courses that the sum of two continuous functions is also continuous and a scalar multiple of a continuous functions is also continuous. Therefore, we see that addition and scalar multiplication is defined on this set naturally and it is easy to check that this space is a vector space.

Remark 7.1.7. While we will see this in the upcoming chapters, some people are quick to see that the addition and scalar multiplication in $P_2(\mathbb{R})$ is *exactly* like the addition and scalar multiplication in $\mathbb{R}^3$. One can actually identify a polynomial $a_0 + a_1x + a_2x^2$ with the vector

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}.$$

However, the space of continuous functions is a little bit more interesting. We will see that this space is an *infinite dimensional vector space.*

Remark 7.1.8. Similar to the space of continuous functions, you can consider the space of all differentiable functions, the space of all integrable functions, the space of all functions which have a Taylor series expansion at 0 and so on. There is a lot of linear algebra structure inside calculus, too!

Example 7.1.9. Next, let us see a more *exotic* example. Consider the set

$$V = \{x \in \mathbb{R} \colon x > 0\}$$

of all *positive* real numbers. On V define *addition* as multiplication and *scalar multiplication* as taking powers. Since this is a little bit confusing, let us denote *addition* by $\oplus$ and *scalar multiplication* by $\otimes$. Then, $x \oplus y = xy$ and $c \otimes x = x^c$ for every $x, y \in V$ and $c \in \mathbb{R}$. For instance, $2 \oplus 3 = 6$ and $3 \otimes 5 = 3^5$. With these operations V becomes a vector space.

Exercise 7.1.10. Firstly, make sure that you understand this example. Show that the set of positive real numbers equipped with the operations in the previous example becomes a vector space. What is the *zero vector*?

Now, let us see a non-example.

Exercise 7.1.11. Consider the set $V = \{(x, y) \colon x, y \in \mathbb{R}\}$, the set of points in the plane. Define *addition* of two points as the midpoint of the line segment between them and *scalar multiplication* as usual coordinatewise scalar multiplication. That is,

$$(a,b) \oplus (x,y) = \left(\frac{a+x}{2}, \frac{b+y}{2}\right),$$
$$c(x,y) = (cx, cy).$$

Now, consider the following two sentences:

First Sentence. For every $P \in V$, there is a Q such that $P+Q = Q+P = P$.

Second Sentence. There is an element Q such that for every $P \in V$ we have $P+Q = Q+P = P$.

And solve the following questions.

1. Compute $(1,2) \oplus (3,4)$.
2. Is the first sentence true?
3. Is the second sentence true?
4. Does V have a zero vector?
5. Is V a vector space with these operations?

7.2. Linear combinations

Whenever we have addition and scalar multiplication, we can talk about linear combinations. Note that we do not even need to have a vector space structure to do this. A linear combination of vectors $v_1, \ldots, v_n$ is simply a vector you can create using these vectors and allowed operations: addition and scalar multiplication. We have already seen this notion when we were dealing with vectors in $\mathbb{R}^n$. Please go and read Definition 2.2.9. You will notice that I intentionally used a general language there. I did not say anything about $\mathbb{R}^n$. So, the same definition works for any vector space.

Therefore, the definition of *span of vectors* is also the same as before. The span of vectors $v_1, v_2, \ldots, v_n$ is the set of all linear combinations of these vectors. Hence, a vector belongs to the span of $v_1, \ldots, v_n$ if and only if you can create that vector using addition and scalar multiplication from the original vectors $v_1, \ldots, v_n$.

Example 7.2.1. Consider the two polynomials $p = 1 + x + x^2$ and $q = 2 + 3x + 4x^2$. Suppose somebody asks you "Is $3 + 2x^2$ in the span of p and q?" What do you do?

Well, as we have always said, this is a language problem. The difficulty here (if there is any difficulty) is to translate this question into a language that we can already speak with. And luckily, we can do this by just answering questions of the form "What does this mean"?

- Is $3 + 2x^2$ in the span of p and q? What does this mean?
- It means "can we write $3 + 2x^2$ as a linear combination of p and q?" What does this mean?
- It means "can we find scalars a, b such that $3 + 2x^2 = ap + bq$?" At this point, let us put what p and q are.
- The question is now "can we find scalars a, b such that

  $$a(1 + x + x^2) + b(2 + 3x + 4x^2) = 3 + 2x^2 \text{ ''}$$

  And what does this mean?
- This means "can we find scalars a, b such that

  $$(a + 2b) + (a + 3b)x + (a + 4b)x^2 = 3 + 2x^2 \text{ ''}$$

  and again what does this mean?
- It means "can we find a, b such that

  $$\begin{aligned} a + 2b &= 3 \\ a + 3b &= 0 \\ a + 4b &= 2 \quad \text{''} \end{aligned}$$

- I think you already have seen what happened here. But let us write it down: Is the system

  $$\begin{aligned} a + 2b &= 3 \\ a + 3b &= 0 \\ a + 4b &= 2 \end{aligned}$$

  consistent?

So, we did ask a question about polynomials and we have realized that we can translate this problem to a linear algebra problem.

Remark 7.2.2. It is important to note that you should always go step by step. What does this mean? What is this question trying to ask me? Can I turn this problem into something that I already know? As Ali Nesin - the founder of the Nesin Math Village - says: Do not try to *solve* the problem. Try to *understand* the problem. Once you understand the problem, the answer jumps out of the page. And we have an example of this philosophy in the previous example. At this point, you know how to check if a system is consistent or not, you know this from the first week of this course. And the key point here is to understand that the question is asking you this. Once you figure that out, solving the question is easy. As you see, I did not even finish solving the problem.

Exercise 7.2.3. Consider the three functions f, g, h in the space $C[0, 1]$ given by the rules: $f(x) = 1$ (this is a constant function), $g(x) = \sin^2 x$ and $h(x) = \cos^2 x$. Is f in the span of f, g, h? (The definition for $C[0, 1]$ was given in the previous section).

Remark 7.2.4. In a vector space, we know how to add two things. As a result, we can add three things by first adding the first two and then adding the third one. So, $v_1 + v_2 + v_3 = (v_1 + v_2) + v_3$. (Remember associativity rule). Then, I know that I can add four things similarly. In a similar way, I can add one hundred million vectors together. But.. I do not know how to add "infinitely many vectors".

So, while I know that the exponential function has a power series expansion

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

this does not tell me that

$$e^x \in \operatorname{span}\{1, x, x^2, \dots\}$$

inside the space of continuous functions. Because this is an infinite sum. This will be important for those of you who will deal with infinite dimensional spaces in the future.

7.3. Subspaces

For me a *space* is a set with some extra structure on it. In our case, the extra structure is addition and scalar multiplication. This is what we consider and what linear algebra is all about. If we had other goals, we could have considered only an additive structure. Or maybe only a multiplicative structure, who knows.

When you have a space, you don't only think about elements. You also care about the extra structure that is carried. Therefore, when you think about *things* related to this space, you want to make sure that the structure is preserved. You have seen examples of it: a linear transformation was just a special type of function which respects addition and scalar multiplication, you have seen that you can still do addition and scalar multiplication on the solution set of a homogeneous system.

I think this is enough motivation to make the definition.

Definition 7.3.1. Let V be a vector space. A subset $W \subset V$ is called a *subvectorspace* or a *subspace* if W is also a vector space with the addition and scalar multiplication defined for V.

Quickly, we make some remarks.

Remark 7.3.2. If you look at the definition, it looks like you have to check all the eight axioms of a vector space for W. However, most of these axioms are satisfied because elements of W already live in the vector space V (because W is a subset of V). For example, you do not need to check if the associativity axiom $w_1 + (w_2 + w_3) = (w_1 + w_2) + w_3$ for every three vectors in W because these vectors live in V and you *know* that this property holds for these vectors in V.

Motivated by this remark, we have the following *subspace test*.

Proposition 7.3.3. *Let V be a vector space and W be a subset of V. In order to check if W is a subspace or not, it is enough (and necessary) to check only the following three:*

1. *The zero vector of V is inside W.*
2. *For every $w_1, w_2 \in W$, we have $w_1 + w_2 \in W$.*
3. *For every scalar c and vector $w \in W$, we have $cw \in W$.*

The first one here guarantees (sort of) that the space is not empty. The last two say that you can do addition and scalar multiplication in W *without leaving* it.

Example 7.3.4. You have seen that the space all functions from real numbers to real numbers form a vector space under usual addition and scalar multiplication. The zero vector in this space is the (constant) zero function - the function which takes every input to zero. Now, we can say that the subset of this space consisting of all continuous functions is a subspace: indeed, you learn from calculus 1 that

1. The constant zero function is continuous,
2. The sum of two continuous functions is again continuous,
3. If you multiply a continuous function with a scalar, you will get another continuous function.

Example 7.3.5. The same arguments from the previous example work for differentiable functions. So, the set of differentiable functions is a subspace of the set of all functions from real numbers to real numbers. It is also a subspace of the set of all continuous functions from real numbers to real numbers because you know that differentiable functions are continuous.

The previous two examples were your first interactions with linear algebra back in your calculus days. The next example was your first interaction with subspaces in this course.

Example 7.3.6. The solution set of a homogeneous linear system with n variables is a subspace of $\mathbb{R}^n$.

Example 7.3.7. Given a vector space V and vectors $v_1, \ldots, v_n$ in V, the span of $v_1, \ldots, v_n$ is a subspace of V. Indeed,

1. The zero vector belongs to the span because it can be written as a linear combination of $v_1, \ldots, v_n$ as

$$0 = 0v_1 + \ldots + 0v_n,$$

2. The sum of two linear combinations is again a linear combination:

$$a_1v_1 + \ldots + a_nv_n + b_1v_1 + \ldots + b_nv_n = (a_1 + b_1)v_1 + \ldots + (a_n + b_n)v_n.$$

3. For a scalar c and a linear combination $a_1v_1 + \ldots + a_nv_n$, the product

$$c(a_1v_1 + \ldots + a_nv_n) = ca_1v_1 + \ldots + ca_nv_n$$

is again a linear combination.

The next exercise is very standard.

Exercise 7.3.8. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Show that

1. The kernel of f is a subspace of $\mathbb{R}^n$,
2. The image of f is a subspace of $\mathbb{R}^m$.

Exercise 7.3.9. Consider the space $P_3(\mathbb{R})$ of all polynomials with real coefficients with degree at most three. Which one of the following sets are subspaces?

1. the set of polynomials with degree at most 1,
2. the set of polynomials with degree exactly 1,
3. the set of polynomials which do not have an x^2 term (the coefficient of x^2 is zero).

Considering the same space $P_3(\mathbb{R})$, give two subspace examples.

Exercise 7.3.10. Consider the following subsets of $\mathrm{Mat}_{2\times 2}(\mathbb{R})$ and determine which ones are subspaces:

1. The set of matrices who are equal to their own transpose.
2. The set of matrices who are equal to the negative of their transpose.
3. The set of diagonal matrices.
4. The set of upper triangular matrices.

5. The set of matrices whose diagonal entries sum up to zero.
6. The set of matrices whose top right entry is zero.
7. The set of invertible matrices.
8. The set of noninvertible matrices.

8. Basis and dimension

In the previous chapter, we have learned about vector spaces and we have seen that once we have the definition of vector spaces, we can talk about linear combinations.

Here is a very natural question, then: if I give you bunch of vectors, what other vectors can you *create* from them? This was the notion of span. You have seen that the span of a set of vectors is a subspace.

The next natural question is: given a subspace (or a vector sapce) can you find a set of vectors which span that subspace? This is the topic of the first section of this chapter.

We will then continue with the question: how can you make sure that you can choose as few vectors as possible to span the entire space? This is the second section.

This will bring us to the third section: the notion of basis and dimension.

8.1. Spanning sets

Suppose that you give a subspace to me: a subspace S of a vector space V. Then, I can write any element $x \in S$ as $x = 1x$. Hence, x can be written as a linear combination of elements in S.

Well, that was too easy, but we just proved that every subspace has a spanning set. So, if somebody asks you the following question: *I give you a subspace, can you give me a set of vectors from which you can create the entire subspace?*, you can tell them *Yes, I can give you the entire thing back.* Well, of course, this solves the problem but as you can imagine it is overkill and not what we are after. We want to make our lives easy by just considering a small amount of vectors and create every other vector by using them.

8.2. Linear independence

8.3. Basis and dimension

9. Linear transformations

10. Matrix representation of linear transformations

Part III

Jordan Canonical Form

11. Eigenthings

12. Invariant Subspaces

13. Nilpotent Operators

14. Jordan Canonical Form

Part IV

Inner Product Spaces