LECTURE NOTES - MATH 113 (SPRING 2021)

UMUT VAROLGUNES

Contents

1. Lecture 1: Some remarks on mathematical rigor, sets, operations, fields 2
2. Lecture 2: Isomorphisms of fields, equivalence relations, Examples of fields: $\mathbb{F}_2$, $\mathbb{F}_p$, $\mathbb{C}$ 5
3. Lecture 3: Geometric interpretation of complex numbers and its operations, first examples of vector spaces 9
4. Lecture 4: Vector spaces, subspaces 12
5. Lecture 5: Sums and direct sums of subspaces, linear combinations, span of a list of vectors, finite dimensionality 14
6. Lecture 6: More on the span of a list of vectors, linear dependence 16
7. Lecture 7: Our first theorem 18
8. Lecture 8: Basis and dimension for finite dimensional vector spaces 20
9. Lecture 9: Dimension of a subspace, characterizations of bases, dimensions of sums of subspaces 21
10. Lecture 10: Linear maps 24
11. Lecture 11: Column vectors and matrices, invertible maps, composition of linear maps 26
12. Lecture 12: Rank-nullity Theorem 28
13. Lecture 13: Determinant of a matrix 30
14. Lecture 14: Polynomials 33
15. Lecture 15: Matrix of a linear map with respect to bases, operators 36
16. Lecture 16: Eigenvalues/vectors, invariant subspaces 38
17. Lecture 17: Existence of eigenvalues over complex numbers 39
18. Lecture 18: Splittings and block matrices, triangular matrices for operators over complex numbers 42
19. Lecture 19: Eigenspaces, Jordan blocks, generalized eigenvectors/spaces 44
20. Lecture 20: Proof of Jordan normal form theorem I 46
21. Lecture 21: Proof of Jordan normal form theorem II: Nilpotent operators 48
22. Lecture 22: Inner product spaces, orthonormal bases 50
23. Lecture 23: Projection formula, self-adjoint operators, isometries 53
24. Lecture 24: Spectral theorem 55
25. Lecture 25: Hermitian inner products on complex vector spaces, eigenvalues of self-adjoint operators, isometries as distance preserving maps 57
26. Lecture 26: Multiplicativity of determinants, orientations, isometries of $(\mathbb{R}^n, \cdot)$ 60
27. Lecture 27: Proof of the normal form theorem for isometries, polar decomposition, SVD 62

28. Some problems 64
28.1. Constructing the rational numbers 64
28.2. Finite fields 65
28.3. Subfields 65
28.4. Solving an equation over different fields 65
28.5. A weird vector space 66
28.6. Subspaces of $\mathbb{F}_2^5$ 66
28.7. When do we consider two vector spaces to be the same? 66
28.8. Polynomial and functional vector spaces 66
28.9. Matrix multiplication 67
28.10. Finite fields revisited 67
28.11. Matrix represenations of linear maps 67

1. Lecture 1: Some remarks on mathematical rigor, sets, operations, fields

Mathematically proving a statement means reducing it to axioms/definitions and (very importantly) previously proven statements by way of logic. In mathematics, you simply cannot know that a statement is true without proving it.

On the other hand, when you cannot prove something or give an invalid proof it is very rarely the case that this is because you do not know the mechanics of proofs. It is either laziness (just making a statement which feels true but not properly thinking about why) or quite possibly you do not have a sufficiently good understanding of the actual content of what you are thinking about. The latter might be not knowing the definition of something, but you might also have all the definitions of everything and still not be able to prove a statement that actually is true.

You have given arguments that come pretty close to proofs your entire life.

Question 1. Find all rational numbers $x \in \mathbb{Q}$ satisfying the equation

$$x^2 = 4.$$

□

When you write that x is either 2 of -2 on your answer sheet, what you are really saying is the following.

Claim 1. *$x \in \mathbb{Q}$ satisfies the equation $x^2 = 4$ if and only if $x = 2$ or $x = -2$.*

Here is something that is not a proof of Claim 1: Clearly, $x = \pm 2$ solve the equation and I cannot find any other solutions – I hope we will all agree that this is not a proof.

Here is the outline of a proof:

$$x^2 = 4 \iff x^2 - 4 = 0 \iff (x-2)(x+2) = 0 \iff x = 2 \text{ or } x = -2$$

This is an outline because you would need to justify each if and only if sign. Since we are very used to doing arithmetic with rational numbers each of these steps might look obvious (and they are not that difficult). Do not worry about this now. We will come back to this at the end of this lecture.

In order to erase the effects of this sort of ”looking obvious“, we will start the course with introducing abstract fields, learn how to do arithmetic with them and develop the fundamental notions of linear algebra in this very general setting. Once the linear algebra part starts having some real content, we will switch to working with real and complex numbers.

Remark 1. There is something slightly awkward that I want to warn you about. We will not give a rigorous construction of real numbers. This is non-trivial business and the ideas involved are not really that relevant in linear algebra. We are just going to assume that the real numbers are somehow defined and they form a field with the operations we are accustomed to. This awkwardness will not occur today. □

Here is another notion that we will be vague about. For us a set is something that has elements. Some examples are: $\{1, 2, 3, 4, 5\}$, $\mathbb{Z}$, $\{\text{apple}, \text{orange}\}$, $\mathbb{R}$, $\mathbb{C}$. A map of sets $S \to S'$ is something that assigns an element of S' to each element of S.

Let S be a set. An operation on S is a machine that takes in two elements of S in an order (which matters in general) and produces an element of S. Let us say this in more proper language.

If S and S' are sets, then we can define the product set

$$S \times S' : \{(s, s') \mid s \in S, s' \in S'\}.$$

Definition 1. Let S be a set. An operation on S is a map

$$S \times S \to S.$$

□

When talk about an operation we generally choose a symbol like $\odot$ and denote the result of applying the operation to element $(s_1, s_2) \in S \times S$ by $s_1 \odot s_2$.

Example 1. $S = \{1, 2\}$ with $1 \odot 1 = 1$, $1 \odot 2 = 1$, $2 \odot 1 = 2$, $2 \odot 2 = 2$. □

Exercise 1. Concisely describe this operation in words. □

Exercise 2. The set of integers $\mathbb{Z}$ has operations addition $+$ and multiplication $\cdot$. Are subtraction and/or division operations on $\mathbb{Z}$? □

We are ready for the main definition of the day, but it is a long one.

Definition 2. A field $\mathbb{F}$ is a set equipped with two operations $\oplus$ (referred to as addition) and $\odot$ (referred to as multiplication) which satisfy the following axioms:

- commutativity of both operations:
$$a \oplus b = b \oplus a \text{ and } a \odot b = b \odot a,$$
for every $a, b \in \mathbb{F}$
- associativity of both operations:
$$(a \oplus b) \oplus c = a \oplus (b \oplus c) \text{ and } (a \odot b) \odot c = a \odot (b \odot c),$$
for every $a, b, c \in \mathbb{F}$
- existence of identity elements for operations: there exists elements $0 \neq 1 \in \mathbb{F}$ such that
$$a \oplus 0 = a \text{ and } a \odot 1 = a,$$
for every $a \in \mathbb{F}$

- existence of additive inverses: for every $a \in \mathbb{F}$, there exists $a' \in \mathbb{F}$ such that $a \oplus a'$ is an additive identity
- existence of multiplicative inverses for non-zero elements: for every $a \in \mathbb{F}$ which is not an additive identity, there exists $a' \in \mathbb{F}$ such that $a \odot a'$ is a multiplicative identity
- distributivity of multiplication over addition:
  $$a \odot (b \oplus c) = (a \odot b) \oplus (a \odot c),$$
  for every $a, b, c \in \mathbb{F}$

□

Exercise 3. Is $(\mathbb{Z}, +, \cdot)$ a field? □

Conveniently, it follows from the axioms that the identity elements and inverse are in fact unique.

Lemma 1. *Let $\mathbb{F}$ be a field.*

(1) There is only one additive identity.
(2) There is only one multiplicative identity.
(3) For every $a \in \mathbb{F}$, there exists only one additive inverse.
(4) For every $a \in \mathbb{F} - \{0\}$, there exists only one multiplicative inverse.

Proof. I will do only one, the first one. Assume that both 0 and $0'$ are additive identity elements. Then, we have $0 \oplus 0' = 0$, but also $0 \oplus 0' = 0' \oplus 0 = 0'$. Therefore, $0 = 0'$. □

Exercise 4. Carefully prove the other three statements. □

We will denote the additive identity by 0, and the multiplicative identity by 1 as above. Let us also denote the additive inverse of a by $-a$ and, if $a \neq 0$, we denote the multiplicative inverse by a^{-1}. It is also customary to write

$$a - b \text{ for } a \oplus (-b)$$

In your first homework, you will go through the construction of rational numbers and prove the following important result.

Lemma 2. $(\mathbb{Q}, +, \cdot)$ *is a field. The identities and inverses are what you have been saying they are all your life.*

Exercise 5. Let's assume that rationals are what you had been saying they are so far and assume this lemma. Turn the proof outline

$$x^2 = 4 \iff x^2 - 4 = 0 \iff (x-2)(x+2) = 0 \iff x = 2 \text{ or } x = -2$$

of Claim 1 to an actual proof. □

This is probably still slightly confusing. Let us now start doing arithmetic over fields. We will now remove the circles from the operations of an abstract field. The $\cdot$ will slowly simply disappear, and you will be expected to understand that it is there from the context.

We can simply repeat the discussion from the beginning.

Question 2. Let $\mathbb{F}$ be a field. Find all elements $x \in \mathbb{F}$ satisfying the equation

$$x^2 = 4.$$

□

Exercise 6. What is 4 here? How about x^2? Define them carefully in the only possible way. □

Remark 2. Part of doing mathematics is also to make good sensible definitions. Even though you will not be doing that a lot in this course, I gave you one chance here. □

The following claim which is a full answer of the question is still true.

Claim 2. *$x \in \mathbb{F}$ satisfies the equation $x^2 = 4$ if and only if $x = 2$ or $x = -2$.*

The same outline of the proof can still be turned into a full proof:

$$x^2 = 4 \iff x^2 - 4 = 0 \iff (x-2)(x+2) = 0 \iff x = 2 \text{ or } x = -2$$

If you solved Exercise 5 in the right way the same proof works here as well. If you were confused there maybe just do this one, which implies Exercise 5 by Lemma 2.

We finish with a trick question.

Question 3. Is it true that for every $\mathbb{F}$, $x^2 = 4$ has two solutions? □

2. Lecture 2: Isomorphisms of fields, equivalence relations, Examples of fields: $\mathbb{F}_2$, $\mathbb{F}_p$, $\mathbb{C}$

There are many many different kinds of fields. Today we will give examples and non-examples of fields.

Let us start with a definition. Assume that we have two fields $\mathbb{F}$ and $\mathbb{F}'$, and a bijective map $\phi : \mathbb{F} \to \mathbb{F}'$ such that

$$\phi(a+b) = \phi(a) + \phi(b) \text{ and } \phi(a \cdot b) = \phi(a) \cdot \phi(b),$$

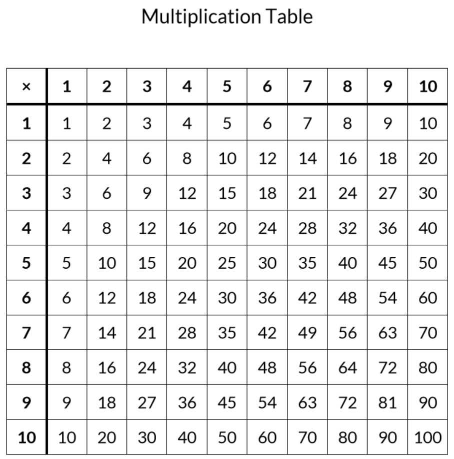
for every $a, b \in \mathbb{F}$.

Exercise 7. Prove that it also follows that $\phi(0) = 0$ and $\phi(1) = 1$. □

We call ϕ an isomorphism, and say that $\mathbb{F}$ and $\mathbb{F}'$ are isomorphic fields. This means that if I consider the addition and multiplication tables (see the figure in the next page for an example of how these look like) of $\mathbb{F}$ and replace all the elements in the tables with elements of $\mathbb{F}'$ using ϕ, what I obtain is precisely the full addition and multiplication tables of $\mathbb{F}'$.

Remark 3. In mathematics, generally, two sets equipped with the same kind of "extra structure" being isomorphic means that they are the same up to a relabeling of the elements as above. We will consider this notion for vector spaces as well later. □

Remark 4. Even though in this lecture this will not come up, it is an important point that two fields can be isomorphic in different ways. For example, there can be an isomorphism $\phi : \mathbb{F} \to \mathbb{F}$, which is not the identity map. These would be called symmetries of $\mathbb{F}$. We will come back to this notion later in the course for vector spaces and normed vector spaces. □



Multiplication Table

×	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10
2	2	4	6	8	10	12	14	16	18	20
3	3	6	9	12	15	18	21	24	27	30
4	4	8	12	16	20	24	28	32	36	40
5	5	10	15	20	25	30	35	40	45	50
6	6	12	18	24	30	36	42	48	54	60
7	7	14	21	28	35	42	49	56	63	70
8	8	16	24	32	40	48	56	64	72	80
9	9	18	27	36	45	54	63	72	81	90
10	10	20	30	40	50	60	70	80	90	100

FIGURE 1. A respectable looking multiplication table for numbers $1 - 10$ to explain you the form of an addition or multiplication table.

Let us start with fields with finitely many elements. Since we assumed $0 \neq 1$, a field needs to have at least two elements.

Proposition 1. *On the set $\{a, m\}$ there exists exactly one way to define addition and multiplication operations so that these form a field with the property that a is the additive identity and m is the multiplicative identity.*

Proof. Let us try to determine the operations assuming they give a field. Using what it means to be an additive or multiplicative identity (along with commutativity), we automatically know what the operations should be immediately except $m + m$ and $a \cdot a$.

We know that m needs to have an additive inverse and a cannot be that inverse. Therefore, we need $m + m = a$.

We also have that

$$a \cdot a = a \cdot (m + m) = a \cdot m + a \cdot m = a + a = a.$$

Hence, we know what the addition and multiplication operations need to be if they are to define a field. What we now need to do is to check that these operations indeed do define a field by checking the axioms. I leave this to you - note that we could really fail here. □

Remark 5. When writing this proof, I could have denoted a by 0 and m by 1. It would be sort of like giving nicknames to a and m, and refer to them by their nicknames. Note that sometimes nicknames describe a person better than their name. □

Exercise 8. Prove that there exists a unique field with two elements up to isomorphism. □

Exercise 9. Show that for an arbitrary field $\mathbb{F}$, we have $0 \cdot 0 = 0$. □

Another one of your homework problems will be showing that for every prime number p there exists a unique field $\mathbb{F}_p$ with p elements up to isomorphism. Just as a reminder: prime numbers are $2, 3, 5, \dots$ - positive integers who are divisible only by 1 and themselves.

Remark 6. We have just constructed $\mathbb{F}_2$ and showed, as you will understand hopefully while solving the exercise, that it is unique up to isomorphism. □

Today I will get you started on that problem and use this chance to introduce another abstract notion.

An equivalence relation on a set S is a subset $E \subset S \times S$ with some properties that we will momentarily state. First we introduce a notation. If $(s_1, s_2) \in E$, we write $s_1 \sim s_2$ and say s_1 is equivalent to s_2. The properties/axioms that E is supposed to satisfy to be an equivalence relation are

- $s \sim s$ for every $s \in S$.
- $s_1 \sim s_2$ if and only if $s_2 \sim s_1$ for every $s_1, s_2 \in S$.
- If $s_1 \sim s_2$ and $s_2 \sim s_3$, we have $s_1 \sim s_3$ for every $s_1, s_2, s_3 \in S$.

It is customary to simply say $\sim$ is an equivalence relation on S and never give a name to the subset. This definition albeit being abstract is a simple one. Let's say you have some students in a class and you want to divide them into groups, but you want to do this by writing down which ordered pairs of students are going to be in the same group. If you do not want want to lose the respect of your students, you better choose these ordered pairs in such a way that they satisfy the three properties listed.

As this analogy suggests, if you have a set S with an equivalence relation $\sim$, you can talk about the set of equivalence classes denoted by $S/\sim$. An equivalence class is a subset of S such that all of its elements are equivalent to each other and none of its elements are equivalent to an element in its complement. These are the groups of students from above.

Example 2. Let $n > 0$ be an integer. We define an equivalence relation on $\mathbb{Z}$ as follows. We declare $a \sim b$ if $a - b$ is divisible by n, i.e.

$$a \equiv b \mod n$$

□

Exercise 10. Prove that that this is indeed an equivalence relation by quickly checking the axioms. □

Let us call the set of equivalence classes $\mathbb{Z}/n\mathbb{Z}$ and the denote the equivalence class of $a \in \mathbb{Z}$ by $[a] \in \mathbb{Z}/n\mathbb{Z}$. By definition $[a] = [a + n]$. Notice that:

$$\mathbb{Z}/n\mathbb{Z} = \{[0], [1], \dots [n-1]\}$$

Finally, we define addition and multiplication on $\mathbb{Z}/n\mathbb{Z}$:

- $[a] + [b] = [a + b]$, for $0 \leq a, b \leq n - 1$
- $[a] \cdot [b] = [a \cdot b]$, for $0 \leq a, b \leq n - 1$

where on the RHS we are using the addition and multiplication operations on the set of integers.

In the homework you will show that if n is a prime number, then these operations turn $\mathbb{Z}/n\mathbb{Z}$ into a field.

Exercise 11. Show that $\mathbb{Z}/4\mathbb{Z}$ with these two operations is not a field. □

Remark 7. There are other finite fields. We will prove in this course that the number of elements of a finite field has to be q^k for some prime q and $k \geq 1$. In fact, up to isomorphism such a field is unique, but we will not show this in this course.

□

Enough about finite fields! Let's go to infinite fields. As I mentioned last time, that $\mathbb{Q}$ is a field is in your homework. You should also be able to convince yourself that $\mathbb{R}$ is a field as well (with the caveat I mentioned in the previous class, shh..)

The next example is complex numbers. As a set $\mathbb{C} := \mathbb{R} \times \mathbb{R}$, but we denote an element $(a, b) \in \mathbb{C}$ by

$$a + ib.$$

In the next class, I will explain to you how to think about complex numbers geometrically, but for today let us simply define the addition and multiplication on them:

- $(a + ib) + (c + id) = (a + c) + i(b + d)$
- $(a + i0) \cdot (c + id) = ac + iad$
- $(0 + i1)^2 = -1 + i0$

As you may have realized the last two look quite weird unless we introduce the following shorthand notations:

- $a + i0$ is denoted by a
- $0 + ib$ is denoted by ib
- $i1$ is denoted by i

Exercise 12. We put ourselves in danger here since $a + ib$ can be interpreted in two ways. Show that these two interpretations are in fact the same. □

The last two rules with these conventions looks like:

- $a \cdot (c + id) = ac + iad$
- $i^2 = -1$

Exercise 13. Deduce the general multiplication rule for two complex numbers from these two rules assuming that the addition and multiplication operations turn $\mathbb{C}$ into a field. □

Next class we will also prove that these operations (addition as defined above and the multiplication you just derived) indeed turn $\mathbb{C}$ into a field.

Let's finish with an example of a set with two operations which come very close to being a field called quaternions (we will never see them again). As a set $\mathbb{H} = \mathbb{R}^4$ with elements denoted by

$$a + ib + jc + kd.$$

Addition of two quaternions and multiplication of a "purely real" quaternion with an arbitrary quaternion is defined exactly as in complex numbers. In addition:

- $i^2 = j^2 = k^2 = -1$
- $ij = -k$

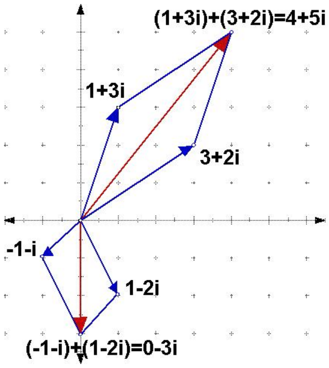


Figure 2. Adding complex numbers geometrically

Exercise 14 (Bonus). Show that multiplication of two quaternions can be defined in only one way so that the two rules above are satisfied and along with the addition operation all the axioms of a field except commutativity of the multiplication are satisfied. After defining the multiplication operation this way, show that commutativity of multiplication is not satisfied. □

3. Lecture 3: Geometric interpretation of complex numbers and its operations, first examples of vector spaces

Recall from last time that complex numbers $\mathbb{C}$ are defined to be the set $\mathbb{R} \times \mathbb{R}$ with the operations

- $(a + ib) + (c + id) = (a + c) + i(b + d)$
- $(a + ib) \cdot (c + id) = (ac - bd) + i(bc + ad)$

Note that on the right hand side of these definitions, we are using addition and multiplication operation of real numbers.

Let's start checking that these operations make $\mathbb{C}$ into a field.

- Commutativity of addition and multiplication follows from the formulas and commutativity of addition and multiplication for real numbers
- Associativity of addition follows again from associativity of addition for real numbers. Associativity of multiplication can be checked directly but let's leave it for later.
- Easy to see that $0 + i0$ is an additive identity and $1 + 0i$ is a multiplicative identity.
- Again it is clear that $-a - ib = -a + i(-b)$ is an additive inverse for $a + ib$. We leave the existence of multiplicative inverses for later.
- Let us leave the distributivity axiom for later as well.

Before we check these three remaining axioms, we make an excursion into the geometry of complex numbers.

First of all, we think of complex numbers as points in the Cartesian plane where $a + ib$ is the point with coordinates (a, b).

We can also of course think of points in the Cartesian plane (and therefore complex numbers) as vectors with starting points at the origin. The addition of complex numbers correspond precisely to the addition of vectors.

The more challenging task is to understand how to represent multiplication of complex numbers geometrically. We can represent every non-zero vector inside the Cartesian plane in a unique way by its polar coordinates (r, θ):

- r is its magnitude - a positive real number
- θ its phase - an element of $\mathbb{R}/2\pi\mathbb{Z} := \mathbb{R}/\sim$, where $\phi \sim \phi'$ if $\phi - \phi' = 2\pi k$ for some $k \in \mathbb{Z}$

Definition 3. We define the magnitude of a complex number $a + ib$ as

$$|a + ib| := \sqrt{a^2 + b^2}.$$

We define the phase of a non-zero complex number as the unique $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ which satisfies

$$\cos(\theta) = \frac{a}{\sqrt{a^2 + b^2}} \text{ and } \sin(\theta) = \frac{b}{\sqrt{a^2 + b^2}}$$

□

The magnitude and phase of a complex number is of course nothing but the magnitude and phase of the vector that corresponds to it in the Cartesian plane.

Definition 4. For any $\theta \in \mathbb{R}/2\pi\mathbb{Z}$, we define the complex number

$$e^{i\theta} := \cos(\theta) + i\sin(\theta)$$

□

Lemma 3. *Let r and θ be the magnitude and phase of a non-zero complex number $a + ib$. Then,*

$$a + ib = re^{i\theta}.$$

Proof. The right hand side is equal to $\sqrt{a^2 + b^2}(\frac{a}{\sqrt{a^2+b^2}} + i\frac{b}{\sqrt{a^2+b^2}})$ by definition. □

Proposition 2. *For any $\theta, \theta' \in \mathbb{R}/2\pi\mathbb{Z}$, we have*

$$e^{i\theta} \cdot e^{i\theta'} = e^{i(\theta+\theta')}.$$

On the left we used the multiplication of two complex numbers as defined above.

Proof. Direct computation of the left hand side gives

$$(\cos(\theta)\cos(\theta') - \sin(\theta)\sin(\theta')) + i(\cos(\theta)\sin(\theta') + \sin(\theta)\cos(\theta')).$$

Using trigonometric identities, this is equal to

$$\cos(\theta + \theta') + i\sin(\theta + \theta'),$$

which is the right hand side. □

With this proposition, it is straightforward to multiply complex numbers written in terms of their magnitude and phase

$$re^{i\theta} \cdot r'e^{i\theta'} = rr'e^{i(\theta+\theta')}.$$

In words: when you multiply two complex numbers you multiply their magnitudes as real numbers and add their phases (modulo 2π). If you think about one of the complex numbers as a vector in the Cartesian plane and the other one in $re^{i\theta}$ form then the result of multiplication rotates the vector by θ and then scales it by r.

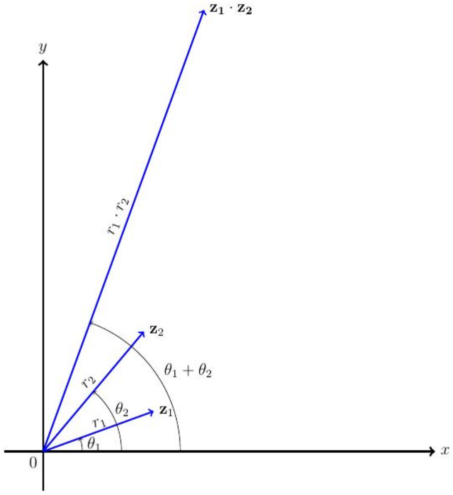


FIGURE 3. Multiplying complex numbers geometrically

The complex number 0 does not have a well defined phase, but multiplying any complex number with 0 clearly results in 0.

Let us first check the associativity of complex multiplication. Multiplication of real numbers is associative and addition of phases modulo 2π is likewise associative. Hence, we have our associativity.

Secondly, we can easily see that any non-zero complex number has a multiplicative inverse using that the magnitude and phase of 1 are 1 and 0 modulo 2π respectively. Therefore if the magnitude and phase of $a + ib$ is r and θ, it is clear that the complex number with magnitude r^{-1} and phase $-\theta$ is a multiplicative inverse.

In standard representation the inverse complex number is

$$\frac{a}{\sqrt{a^2+b^2}} - i\frac{b}{\sqrt{a^2+b^2}}.$$

Exercise 15. Check directly that this is the inverse. □

Finally let us check the distributive law. We need to show that

$$z \cdot (z_1 + z_2) = z \cdot z_1 + z \cdot z_2.$$

This is clearly true for $z = 0$. We write z as $re^{i\theta}$ and think of $z_1, z_2, z_1 + z_2$ as vectors in the Cartesian plane. Multiplication with $re^{i\theta}$ is a θ radian rotation of the plane followed by a scaling by r. These operations both send the send the sum of any two vectors in the plane to the sum of their images. If one thinks about the usual parallelogram picture for adding vectors (as in Figure 2), this becomes visible by rotating or scaling the entire parallelogram.

This finishes our discussion of complex numbers for now. We now give some examples of what will be vector spaces next week.

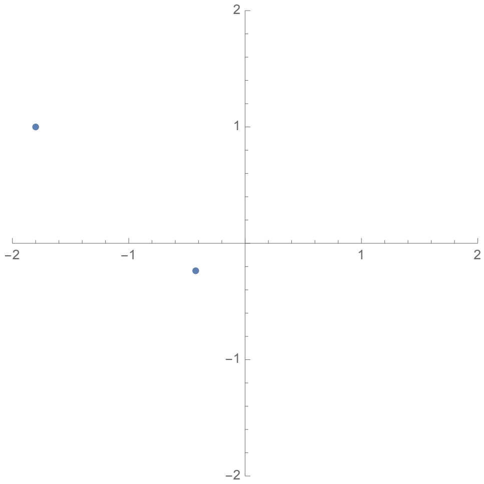


Figure 4. The complex number $-1.8+i$ and its inverse

Let $\mathbb{F}$ be a field. Let $\mathbb{F}^n$ be the set of n-tuples of elements of $\mathbb{F}$:

$$\mathbb{F}^n = \underbrace{\mathbb{F} \times \mathbb{F} \cdots \times \mathbb{F}}_{n} = \{(a_1, \ldots, a_n) \mid a_1, \ldots, a_n \in \mathbb{F}\}$$

If we call an arbitrary element of $\mathbb{F}$ a scalar, we can define scalar multiplication for elements of $\mathbb{F}^n$:

$$c \cdot (a_1, \ldots, a_n) = (ca_1, \ldots, ca_n).$$

We can also add two elements of $\mathbb{F}^n$:

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n).$$

These two operations turn $\mathbb{F}^n$ into a vector space over $\mathbb{F}$ as we will define on Monday.

Here is a slightly more complicated example of a to-be vector space. Fix

$$\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n \in \mathbb{F}.$$

Consider the subset $V \subset \mathbb{F}^n$ of elements $(a_1, \ldots, a_n)$ that satisfy the equations

$$\alpha_1 a_1 + \ldots + \alpha_n a_n = 0 \text{ and } \beta_1 a_1 + \ldots + \beta_n a_n = 0.$$

Exercise 16. Prove that V is closed under multiplication by any scalar and addition. □

As a result we will have a vector space structure on V as well. Notice that V does not come with a "basis". We will see that one can be chosen but even without choosing one, we will be able to say that V is a vector space over $\mathbb{F}$ and work with it.

4. Lecture 4: Vector spaces, subspaces

Definition 5. Let $\mathbb{F}$ be a field. A vector space V over $\mathbb{F}$ is a set equipped with

- scalar multiplication: a map $\mathbb{F} \times V \to V$ denoted by $(c, v) \mapsto c \cdot v$
- vector addition: a map $V \times V \to V$ denoted by $(v_1, v_2) \mapsto v_1 + v_2$

satisfying the following axioms:

(1) vector addition is commutative, associative, admits additive identity and additive inverses
(2) scalar multiplication satisfies
    - $1 \cdot v = v$, for every $v \in V$
    - $(a \cdot b) \cdot v = a \cdot (b \cdot v)$ for every $a, b \in \mathbb{F}$, $v \in V$
(3) the two together satisfy
    - $a \cdot (v + w) = a \cdot v + a \cdot w$ for every $a \in \mathbb{F}$, $v, w \in V$
    - $(a + b) \cdot v = a \cdot v + b \cdot v$ for every $a, b \in \mathbb{F}$, $v \in V$

□

Some remarks:

- Vector addition is an operation in the sense that we defined before. The axioms satisfied by vector addition alone are the same axioms that are satisfied by the addition operation alone in a field.
- In the second axiom for scalar multiplication one out of the four multiplication symbols denote the multiplication in $\mathbb{F}$, whereas the other three are the scalar multiplication. I hope you can see which one is which.
- There are two distributivity axioms: one for vector addition, one for addition in $\mathbb{F}$.

We will use the same symbols for addition in the field and vector addition, and also for multiplication in the field and scalar multiplication. Context will take care of the ambiguity.

The most important example of a vector space is $\mathbb{F}^n$ for $n \geq 0$ with the scalar multiplication and vector addition as we defined in the previous lecture. For $\mathbb{F} = \mathbb{R}$, $n = 1, 2, 3$, we know how to geometrically think about scalar multiplication and vector addition. This will provide important intution but you have to always use intution along with rigor so that it does not lead to mistakes.

Exercise 17. Check the axioms of a vector space for $\mathbb{F}^n$. □

Here is a generalization of $\mathbb{F}^n$. Let S be a set, we define $\mathbb{F}^S$ as the set of all maps $S \to \mathbb{F}$. On $\mathbb{F}^S$ we can define scalar multiplication and vector addition so that it becomes a vector space over $\mathbb{F}$.

- Given $c \in \mathbb{F}$ and $f \in \mathbb{F}^S$, we define $c \cdot f$ to be the map $S \to \mathbb{F}$ that sends s to $c \cdot f(s)$
- Given $f_1, f_2 \in \mathbb{F}^S$, we define $f_1 + f_2$ to be the map $S \to \mathbb{F}$ that sends s to $f_1(s) + f_2(s)$

Exercise 18. Check the axioms of a vector space for $\mathbb{F}^S$. □

Question 4. Why did I say that $\mathbb{F}^S$ is a generalization of $\mathbb{F}^n$? □

Let me also give a weird example of a vector space: $\mathbb{R}$ is a vector space over $\mathbb{Q}$. We use the usual multiplication of a rational number with a real number as scalar multiplication and usual addition of real numbers as the vector addition.

We now note some direct implications of the axioms of a vector space. You should make sure that you can parse them first.

Proposition 3.
- $0 \cdot v = 0$ *for all* $v \in V$.

- $a \cdot 0 = 0$ *for all* $a \in \mathbb{F}$.
- $(-1) \cdot v = -v$ *for all* $v \in V$.

Proof. I will only do the first one.

$$0 \cdot v = (0+0) \cdot v = 0 \cdot v + 0 \cdot v.$$

Adding $-0 \cdot v$ to both sides we obtain $0 = 0 \cdot v$ as desired. □

Remark 8. Noting that $\mathbb{F}$ is a vector space over itself using the field operations (i.e. $\mathbb{F}^n$ for $n = 1$), you can see that these implications hold for field operations as well. □

Here is another important definition.

Definition 6. A non-empty subset W of a vector space V over $\mathbb{F}$ is called a subspace if it is closed under scalar multiplication with any element of $\mathbb{F}$ and vector addition. □

Exercise 19. A subspace $W \subset V$ is automatically a vector space with the scalar multiplication and vector addition operations taken from V. □

Last lecture we had considered the subset $W \subset \mathbb{F}^n$ of elements $(a_1, \ldots, a_n)$ that satisfy the equations

$$\alpha_1 a_1 + \ldots + \alpha_n a_n = 0 \text{ and } \beta_1 a_1 + \ldots + \beta_n a_n = 0,$$

where $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n \in \mathbb{F}$. You showed (hopefully) that W is a subspace, without knowing the word, in an exercise from last class.

One can show that all subspaces of $\mathbb{F}^n$ are the set of solutions of a finite number of linear equations as in this example. Eventually this will become an "obvious" statement for us. For now let us accept it for the simple case $\mathbb{R}^3$ and think about the following question.

Question 5. What are the subspaces of $\mathbb{R}^3$ geometrically? □

Of course, we could also think about the subspaces of for example $\mathbb{F}_2^5$. It is much harder to use visual intuition in this case. There will be a problem on this in your next homework.

Exercise 20. What should be the definition of an isomorphism of vector spaces over the same field $\mathbb{F}$? The answer will be on your next homework as well, along with some questions about it. □

5. Lecture 5: Sums and direct sums of subspaces, linear combinations, span of a list of vectors, finite dimensionality

Let's begin with a definition. Throughout this lecture $\mathbb{F}$ denotes an arbitrary field.

Definition 7. Let V be a vector space over $\mathbb{F}$ and $U_1, \ldots, U_n \subset V$ be subspaces. We define the subset

$$U_1 + \ldots + U_n := \{v \in V \mid \text{there exists } u_1 \in U_1, \ldots u_n \in U_n \\ \text{such that } v = u_1 + \ldots + u_n\}.$$

□

In words $U_1 + \ldots + U_n$ consists of the vectors in V which can be written as a sum of one vector from each U_i.

Lemma 4. *The sum of subspaces $U_1 + \ldots + U_n$ is a subspace.*

Proof. Clearly, $0 = 0 + \ldots + 0$ is an element of $U_1 + \ldots + U_n$. We need to check that it is closed under scalar multiplication and vector addition in V.

- Let $c \in \mathbb{F}$, and $u_1 \in U_1, \ldots u_n \in U_n$, we need to show that
$$c \cdot (u_1 + \ldots + u_n) \in U_1 + \ldots + U_n.$$
By distributivity:
$$c \cdot (u_1 + \ldots + u_n) = c \cdot u_1 + \ldots + c \cdot u_n.$$
Since each U_i is a subspace, we have $c \cdot u_i \in U_i$. This proves the desired claim.
- Let $u_1, u_1' \in U_1, \ldots u_n, u_n' \in U_n$, we need to show that
$$(u_1 + \ldots + u_n) + (u_1' + \ldots + u_n') \in U_1 + \ldots + U_n.$$
By commutativity and associativity the sum is equal to:
$$(u_1 + u_1') + \ldots + (u_n + u_n').$$
Since each U_i is a subspace, we have $u_i + u_i' \in U_i$. This proves the desired claim.

□

Definition 8. Let V be a vector space over $\mathbb{F}$ and $U_1, \ldots, U_n \subset V$ be subspaces. We call $U_1 + \ldots + U_n$ a direct sum if for $u_1 \in U_1, \ldots u_n \in U_n$,
$$u_1 + \ldots + u_n = 0 \text{ implies } u_1 = \ldots = u_n = 0.$$
If $U_1 + \ldots + U_n$ is a direct sum, then we denote it by
$$U_1 \oplus \ldots \oplus U_n.$$

□

The following lemma explains the meaning of a sum being a direct sum.

Lemma 5. *$U_1 + \ldots + U_n$ is a direct sum if and only if for every $v \in U_1 + \ldots + U_n$, there are unique elements $u_1 \in U_1, \ldots u_n \in U_n$ such that*
$$v = u_1 + \ldots + u_n.$$

Proof. Let us first prove the if direction, which means that we will assume that for every $v \in U_1 + \ldots + U_n$, there are unique elements $u_1 \in U_1, \ldots u_n \in U_n$ such that $v = u_1 + \ldots + u_n$ and prove that $U_1 + \ldots + U_n$ is a direct sum. We use the definition: assume that $u_1 + \ldots + u_n = 0$. Since $0 + \ldots + 0 = 0$ and because of the uniqueness that we assumed, this means $u_1 = \ldots = u_n = 0$ as desired.

Now we prove the converse, so now we assume that $U_1 + \ldots + U_n$ is a direct sum. Let us assume that for some $v \in U_1 + \ldots + U_n$ and $u_1, u_1' \in U_1, \ldots u_n, u_n' \in U_n$ we have
$$v = u_1 + \ldots + u_n = u_1' + \ldots + u_n'.$$
Let us subtract $u_1 + \ldots + u_n$ from the two sides of the last equality and use associativity to obtain:
$$(u_1' - u_1) + \ldots + (u_n' - u_n) = 0.$$

By our assumption we get $u_1 = u'_1, \ldots, u_n = u'_n$, proving the desired uniqueness. □

Below are some examples of subspaces of $\mathbb{F}^3$. I will use the notation to ask you exercises today and next class.Let us denote the set

- $U_1 = \{(x, 0, 0) \mid x \in \mathbb{F}\}$
- $U_2 = \{(x, x, 0) \mid x \in \mathbb{F}\}$
- $U_3 = \{(x, y, 0) \mid x, y \in \mathbb{F}\}$
- $U_4 = \{(x, y, z) \mid x, y, z \in \mathbb{F} \text{ with } x + y + z = 0\}$
- $U_5 = \{(x, y, z) \mid x, y, z \in \mathbb{F} \text{ with } x + y + z = 0, x = y\}$
- $U_6 = \{(x, y, z) \mid x, y, z \in \mathbb{F} \text{ with } z = x + y\}$

Exercise 21. Check that these are indeed subspaces. For $\mathbb{F} = \mathbb{R}$, imagine (if you want draw) them geometrically inside the three dimensional Euclidean space. □

Let us denote the set $\{1, \ldots, 6\}$ by $[6]$.

Exercise 22. Can you find $i, j, k \in [6]$ such that

(1) $U_i \oplus U_j = U_k$?
(2) $U_i + U_j = U_k$ but $U_i + U_j$ is not a direct sum?
(3) $U_i \oplus U_j \oplus U_k = \mathbb{F}^3$?
(4) $U_i + U_j = \mathbb{F}^3$ but $U_i + U_j$ is not a direct sum?
(5) $U_i \oplus U_j = \mathbb{F}^3$?

□

Definition 9. V is a vector space over $\mathbb{F}$. A linear combination of a list of vectors $v_1, \ldots, v_n \in V$ is any vector in V of the form

$$a_1 v_1 + \ldots + a_n v_n,$$

for some $a_1, \ldots, a_n \in \mathbb{F}$. □

The set of all linear combinations of a list of vectors $v_1, \ldots, v_n \in V$ is called the span of $v_1, \ldots, v_n$:

$$span(v_1, \ldots, v_n) = \{a_1 v_1 + \ldots + a_n v_n \mid a_1, \ldots, a_n \in \mathbb{F}\} \subset V.$$

Next time we will prove that $span(v_1, \ldots, v_n)$ is the smallest subspace of V which contains the vectors $v_1, \ldots, v_n$.

Here is our final definition.

Definition 10. V is a vector space over $\mathbb{F}$. If there is a finite list of vectors $v_1, \ldots, v_n \in V$ whose span equals V, then V is called finite dimensional. Otherwise, we call it infinite dimensional. □

Question 6. For S a set, when is $\mathbb{F}^S$ finite dimensional? □

6. Lecture 6: More on the span of a list of vectors, linear dependence

Throughout this lecture, let $\mathbb{F}$ be a field and V a vector space over it.

Last time, we ended with defining the span of a list of vectors. We start with a simple claim.

Proposition 4. *Let $v_1, \ldots, v_n \in V$. Then $span(v_1, \ldots, v_n) \subset V$ is a subspace that contains each vector v_i.*

Exercise 23. Carefully write a proof of this. □

Proposition 5. *Let $v_1, \ldots, v_n \in V$. Let $W \subset V$ be a subspace that contains the vectors $v_1, \ldots, v_n$. Then $span(v_1, \ldots, v_n) \subset W$.*

Proof. First of all, since W is closed under scalar multiplication for any scalar $a_i \in \mathbb{F}$, $a_i \cdot v_i$ is in W. Moreover, since W is closed under vector addition (using a simple induction) we can also show that for $a_1, \ldots, a_n \in \mathbb{F}$,

$$a_1 v_1 + \ldots + a_n v_n \in W.$$

This finishes the proof. □

Colloquially, we can say that $span(v_1, \ldots, v_n)$ is the smallest subspace containing $v_1, \ldots, v_n$.

Exercise 24. Let $U_1, \ldots, U_n$ be subspaces of V. Prove that $U_1 + \ldots + U_n$ is the smallest subspace of V containing $U_1, \ldots, U_n$ in the same sense. □

Definition 11. Let $v_1, \ldots, v_n \in V$. If $span(v_1, \ldots, v_n) = V$, we say that $v_1, \ldots, v_n$ span V. □

We can restate a definition we made last time as: if there are $v_1, \ldots, v_n \in V$ that span V, then we call V finite dimensional. Our next order of business is to show that if V is finite dimensional, then there is a well-defined notion of minimum number of vectors that span V. This will lead us to the definition of the dimension of V.

Definition 12. Let $v_1, \ldots, v_n \in V$. If $a_1 v_1 + \ldots + a_n v_n = 0$ for scalars $a_1, \ldots, a_n \in \mathbb{F}$ implies $a_1 = \ldots = a_n = 0$, we say that $v_1, \ldots, v_n$ are linearly independent.

Otherwise, i.e. if there exists $a_1, \ldots, a_n \in \mathbb{F}$ which are not all equal to 0 such that $a_1 v_1 + \ldots + a_n v_n = 0$, then we call them linearly dependent. □

Exercise 25. Prove that the non-zero vectors $v_1, v_2 \in V$ are linearly dependent if and only if $v_2 = a v_1$ for some non-zero $a \in \mathbb{F}$. Visualize what this statement means in $\mathbb{R}^3$. □

If we have a list of linearly dependent vectors, there is a systematic way to remove vectors from the list until the remaining vectors are linearly independent without changing the span. To state the precise statement let us introduce a notation. If $v_1, \ldots, v_n$ is a list of vectors, then $v_1, \ldots, \widehat{v_j}, \ldots, v_n$ be the list of vectors that remains after removing v_j. Here j can be 1 or n as well. Let us also define the span of an empty list of vectors to be $\{0\}$.

Lemma 6. *Let $v_1, \ldots, v_n \in V$ be linearly dependent. Then, there exists $j \in \{1, \ldots, n\}$ such that*

- *$span(v_1, \ldots, \widehat{v_j}, \ldots, v_n) = span(v_1, \ldots, v_n)$*
- *$v_j \in span(v_1, \ldots, v_{j-1})$*

Proof. By the assumption of linear dependence, there exists $a_1, \ldots, a_n \in \mathbb{F}$, not all of them zero, such that

$$a_1 v_1 + \ldots + a_n v_n = 0.$$

Let us choose $j \in \{1, \ldots, n\}$ be the largest such that $a_j \neq 0$. This means that we actually have:

$$a_1 v_1 + \ldots + a_j v_j = 0$$

with $a_j \neq 0$. Subtracting $a_1 v_1 + \ldots + a_{j-1} v_{j-1}$ from both sides and dividing by a_j (which by definition means multiplying by a_j^{-1} using that $a_j \neq 0$), we obtain

$$v_j = -a_j^{-1} a_1 v_1 - \ldots - a_j^{-1} a_{j-1} v_{j-1}. \tag{1}$$

To see the first bullet point, we take any $v \in V$ that is equal to a linear combination of $v_1, \ldots, v_n$ and use the Equation (1) to show that v is also a linear combination of $v_1, \ldots, \widehat{v_j}, \ldots, v_n$. The second bullet point follows by definition from Equation (1). □

Exercise 26. Recall our subspaces $U_1, \ldots, U_6 \subset \mathbb{F}^3$ from the previous lecture.

(1) Which of these subspaces are the span of one vector in $\mathbb{F}^3$?
(2) Are all of these vector spaces finite dimensional? Prove your result.
(3) Find the minimum number of vectors that span each of these subspaces.

□

7. Lecture 7: Our first theorem

In this lecture we will prove a result that is extremely important. You should notice that it is the first result that we are calling a theorem.

Here is a convention for what follows. We have used it in the previous lecture but I will make it more visible here. If we have a list of vectors $v_1, \ldots, v_n$, we will sometimes consider the list $v_i, \ldots, v_k$ for $i \in \{1, \ldots, n, n+1\}$ and $k \in \{0, 1, \ldots, n\}$. If $i < k$, this just means what you think: we take all the vectors in the list between and including v_i and v_k in the list in the same order. If $i = k$, it will mean the list with one vector v_i, and finally if $i > k$, it is simply the empty list.

Here is another piece of notation. Given a list of vectors $v_1, \ldots, v_n$ and a possibly empty subset $\sigma \subset \{1, \ldots, n\}$, we can construct a new list of vectors $v_{\sigma(1)}, \ldots, v_{\sigma(k)}$. This is the empty list if σ is empty, and otherwise $\sigma(1) < \ldots < \sigma(k)$ are all the elements of σ written in increasing order.

Theorem 1. *Let $\mathbb{F}$ be a field and V be a vector space over $\mathbb{F}$. Assume that $v_1, \ldots, v_n$ is a list of vectors that span V and $w_1, \ldots, w_m$ be one that is linearly independent. Then, $m \leq n$.*

Proof. Assume the contrary, i.e. that $m > n$.

We will prove the following claim using induction on j:

- For $0 \leq j \leq n$, there is an $n - j$ element subset $\sigma \subset \{1, \ldots, n\}$ such that the list of vectors
$$w_1, \ldots, w_j, v_{\sigma(1)}, \ldots, v_{\sigma(n-j)}$$
span V.

Let's start with $j = 0$. All we are claiming then is that $v_1, \ldots, v_n$ span V, which we are given.

We go on to $j = 1$ to illustrate the procedure even though we could directly go on to the induction step. This time we are claiming that there is an $i \in \{1, \ldots, n\}$ such that

$$w_1, v_1, \ldots, \widehat{v_i}, \ldots, v_n$$

span V.

Clearly, $w_1, v_1, \ldots, v_n$ span V, since even without w_1 we know that it does. Moreover, this list of vectors is linearly dependent as w_1 is in the span of $v_1, \ldots, v_n$. Now we use Lemma 6, which will be our main tool in the proof. It says that we should be able to remove a vector from $w_1, v_1, \ldots, v_n$ which is in the span of the vectors that came before it in the list, but without changing the span as a result of this removal. The key point is that this vector cannot be w_1, since no vector comes before it. Therefore, the vector that we remove has to be one of $v_1, \ldots, v_n$ and this gives exactly what we were trying to prove.

Let us go on with our induction step. We know that the result is true for $j = 0$. Let us assume that it is true for $0 \leq j \leq n-1$ and prove it for $j+1$.

Our induction hypothesis says that there is an $n-j$ element subset $\sigma \subset \{1, \ldots, n\}$ such that the list of vectors

$$w_1, \ldots, w_j, v_{\sigma(1)}, \ldots, v_{\sigma(n-j)}$$

span V. Now, we consider the list

$$w_1, \ldots, w_{j+1}, v_{\sigma(1)}, \ldots, v_{\sigma(n-j)}.$$

This list clearly spans V and is linearly dependent. The latter is because w_{j+1} is a linear combination of $w_1, \ldots, w_j, v_{\sigma(1)}, \ldots, v_{\sigma(n-j)}$.

We then use our Lemma 6 to deduce that we can remove a vector from the list $w_1, \ldots, w_{j+1}, v_{\sigma(1)}, \ldots, v_{\sigma(n-j)}$, which is in the span of the vectors that come before it and without changing the span. Can this vector be one of $w_1, \ldots, w_{j+1}$? No, because we are given that these vectors are linearly independent! Therefore, the removed vector has to be one of $v_{\sigma(1)}, \ldots, v_{\sigma(n-j)}$. Defining the new σ by removing the corresponding integer from the σ for j, we finish our induction.

For $j = n$, therefore we obtain that $w_1, \ldots, w_n$ span V. In particular, w_{n+1} is a linear combination of $w_1, \ldots, w_n$. This is a contradiction to $w_1, \ldots, w_m$ being linearly independent. Therefore, we cannot have $m > n$, or equivalently $m \leq n$. □

Remark 9. Make sure that you have a good sense of the direction of the inequality. If you need to look at the statement to remember what it says that means you did not yet get the point. □

Let us finish this lecture with a corollary.

Corollary 1. *Let $\mathbb{F}$ be a field and V be a finite dimensional vector space over $\mathbb{F}$. Then any subspace $U \subset V$ is also finite dimensional.*

Proof. Let us assume that there is a list of vectors of length d in V that span V. Our previous Theorem implies that a list of linearly independent vectors in U can have length at most d. This is because if a list of vectors is linearly independent in U, then they are also linearly independent in V.

Therefore, there is a list of vectors $u_1, \ldots, u_n$ in U that is linearly independent and such that no matter what vector from U we add to the list it becomes linearly dependent (otherwise, we could keep adding and contradict the previous paragraph). Now, any $u \in U$ has to be in the span of $u_1, \dots, u_n$ because otherwise $u_1, \dots, u_n, u$ would also be linearly independent, contradicting the choice of $u_1, \dots, u_n$ as in the previous sentence (see Lemma 7 in the next lecture). This finishes the proof since $u_1, \dots, u_n$ then is a list of vectors that span U certifying its finite dimensionality. □

Please go back and solve the exercises that you did not have time to solve.

8. Lecture 8: Basis and dimension for finite dimensional vector spaces

I want to start by turning an argument I used before into a lemma. We will use it in the future as well.

Lemma 7. *Let $\mathbb{F}$ be a field and V be a vector space over $\mathbb{F}$. Assume that we have a linearly independent list of vectors $v_1, \dots, v_n$. Then, if $v \in V$ is not in the span of $v_1, \dots, v_n$, then $v_1, \dots, v_n, v$ is also a linearly independent list.*

Proof. Take any $v \in V$ that is not in the span of $v_1, \dots, v_n$. If we have scalars $a_1, \dots, a_{n+1} \in \mathbb{F}$ such that

$$a_1 v_1 + \ldots + a_{n+1} v = 0,$$

then there are two options: $a_{n+1} = 0$ or $a_{n+1} \neq 0$. In the former case, using that $v_1, \dots, v_n$ is linearly independent, we conclude that all $a_i = 0$, which gives what we want. In the latter case (i.e. $a_{n+1} \neq 0$), subtracting $a_1 v_1 + \ldots + a_n v_n$ from both sides and dividing by a_{n+1} (which by definition means multiplying by a_{n+1}^{-1} using that $a_{n+1} \neq 0$), we obtain

$$v = -a_{n+1}^{-1} a_1 v_1 - \ldots - a_{n+1}^{-1} a_n v_n. \tag{1}$$

This contradicts that v is not in the span of $v_1, \dots, v_n$, showing that the latter option is not actually possible. □

We make one of the most important definitions in this class.

Definition 13. Let $\mathbb{F}$ be a field and V be a vector space over $\mathbb{F}$. A list of vectors $v_1, \dots, v_n$ is called a basis if they are linearly independent and they span V. □

The following proposition explains the usefulness of bases.

Proposition 6. *Let $\mathbb{F}$ be a field and V be a vector space over $\mathbb{F}$. A list of vectors $v_1, \dots, v_n$ is a basis if and only if every $v \in V$ can be written uniquely as a linear combination of $v_1, \dots, v_n$.*

Proof. If $v_1, \dots, v_n$ is a basis, then every $v \in V$ is a linear combination of $v_1, \dots, v_n$ since $v_1, \dots, v_n$ span V. Assuming

$$v = a_1 v_1 + \ldots + a_n v_n = a_1' v_1 + \ldots + a_n' v_n$$

we obtain

$$(a_1 - a_1') v_1 + \ldots + (a_n - a_n') v_n = 0.$$

By the linear independence of $v_1, \dots, v_n$, we see that $a_i = a_i'$, proving the uniqueness.

Conversely, if every $v \in V$ can be written uniquely as a linear combination of $v_1, \dots, v_n$ then we immediately get that $v_1, \dots, v_n$ span V. Moreover, since 0 can be written as a linear combination of $v_1, \dots, v_n$ by choosing all coefficients to be 0, the uniqueness part of our assumption implies the linear independence of $v_1, \dots, v_n$. □

Exercise 27. Assume that $v_1, \dots, v_n$ is a basis for V. Prove that V is isomorphic to $\mathbb{F}^n$! □

As soon as we know that our vector space is finite dimensional, we can conclude that it has a basis using the following proposition.

Proposition 7. *Let $\mathbb{F}$ be a field and V be a finite dimensional vector space over $\mathbb{F}$. Then, V admits a basis.*

Proof. By definition, there is a finite list of vectors that span V. Choose the minimum $n \geq 0$ such that there is a list of vectors $v_1, \dots, v_n$ that span V. If $v_1, \dots, v_n$ were to be linearly dependent, then using Lemma 6, we could produce a shorter list of vectors that span V. Therefore, $v_1, \dots, v_n$ is linearly independent. Since, we already knew that they span, we found a basis. □

We had set out to show that for a finite dimensional vector space V, there is a meaningful notion of a minimum number of vectors that span V. Now we know what it means for a spanning list of vectors to be devoid of redundancies - they should be linearly independent, and hence be a basis. We now prove that the number of elements in any basis has to be the same, finishing the task.

Proposition 8. *Let $\mathbb{F}$ be a field and V be a finite dimensional vector space over $\mathbb{F}$. Then, the number of elements in any basis of V is the same.*

Proof. This immediately follows from Theorem 1. Let $v_1, \dots, v_n$ and $w_1, \dots, w_m$ be two bases. Since $v_1, \dots, v_n$ is linearly independent and $w_1, \dots, w_m$ span V, we get $m \geq n$. Reversing the roles of the two bases in this argument, we also get $n \geq m$. This implies $m = n$ as desired. □

Definition 14. Let $\mathbb{F}$ be a field and V be a finite dimensional vector space over $\mathbb{F}$. Then, the number of elements in any basis of V is called the dimension of V. □

Exercise 28. Compute the dimension of $\mathbb{F}^n$, its subspace given by two equations that we considered first in Lecture 3 (there are cases here, be careful), $U_1, \dots, U_6$. Think about the case $\mathbb{F} = \mathbb{R}$ to make a connection with the intuitive understanding of dimension that you have from before. □

9. Lecture 9: Dimension of a subspace, characterizations of bases, dimensions of sums of subspaces

We continue our development of the notion of a basis of a vector space. Throughout this lecture assume that $\mathbb{F}$ is a field and V is a finite dimensional vector space over $\mathbb{F}$.

Lemma 8. *Every linearly independent list of vectors can be extended to a basis by adding finitely many vectors to the list on the right*[1]*.*

Proof. Let $v_1, \dots, v_n$ be linearly independent. If $v_1, \dots, v_n$ span V, we are done. If not, there exists a $v_{n+1} \in V$ which is not a linear combination of $v_1, \dots, v_n$. Then, by Lemma 7, we have that $v_1, \dots, v_n, v_{n+1}$ is still linearly independent

If $v_1, \dots, v_{n+1}$ span V, we are done. Otherwise, we can add a v_{n+2} to the list, keeping it linearly independent. We keep going with this procedure knowing that it has to stop after finitely many iterations since the length of a linearly independent

[1] Not adding any new vectors is allowed.

list of vectors can be at most the dimension of V by Theorem 1. This produces the desired type of basis. □

Exercise 29. We could structure this proof in a way that is similar to the proof of Proposition 7. Do this by yourself. □

Recall that in Lecture 7, we showed that a subspace of a finite dimensional vector space has to be finite dimensional.

Proposition 9. *Let $U \subset V$ be a subspace. Then the dimension of U is at most the dimension of V.*

Proof. Let $u_1, \dots, u_n$ be a basis for U. Then, $u_1, \dots, u_n$ is a linearly independent list of vectors in V, which can be extended to a basis of V by the previous lemma. This finishes the proof since dimension is defined as the number of elements in a basis. □

Here is an important characterization of bases.

Proposition 10. *Let n be the dimension of V. Then,*

(1) A linearly independent list of n vectors $v_1, \dots, v_n$ is a basis.
(2) A list of n vectors $v_1, \dots, v_n$ which span V is a basis.

Proof. Let's start with (1). We know that $v_1, \dots, v_n$ can be extended to a basis by Lemma 8. Since n is the dimension, this shows that $v_1, \dots, v_n$ already has to be a basis.

We move on to (2). We know that if $v_1, \dots, v_n$ is not linearly independent, then we can remove a vector from the list using Lemma 6 without changing the span. This contradicts Theorem 1 since a basis in V gives a list of linearly independent vectors in V of length n. □

Let us finish by now considering multiple subspaces in V. For the following two exercises, we do not make the assumption that V is finite dimensional.

Exercise 30. Let $U_1, U_2 \subset V$ be subspaces. Prove that $U_1 \cap U_2 \subset V$ is also a subspace. □

Exercise 31. Let $U_1, U_2 \subset V$ be subspaces. Prove that $U_1 + U_2$ is a direct sum if and only if $U_1 \cap U_2 = \{0\}$. □

Let V be finite dimensional again.

Proposition 11. *Let $U_1, U_2 \subset V$ be subspaces. We have the following dimension formula*

$$dim(U_1 + U_2) = dim(U_1) + dim(U_2) - dim(U_1 \cap U_2).$$

Proof. Let $w_1, \dots, w_n$ be a basis for $U_1 \cap U_2$. We find vectors $u_1, \dots, u_l \in U_1$ and $v_1, \dots v_k \in U_2$ such that

- $w_1, \dots, w_n, u_1, \dots, u_l$ is a basis of U_1
- $w_1, \dots, w_n, v_1, \dots v_k$ is a basis of U_2

using Lemma 8.

We claim that we need to have

$$span(u_1, \dots, u_l) \cap (U_1 \cap U_2) = \{0\}.$$

Take a vector v in this intersection. There must be scalars such that

$$v = a_1 u_1 + \ldots + a_l u_l = b_1 w_1 + \ldots + b_n w_n,$$

which implies

$$a_1 u_1 + \ldots + a_l u_l - b_1 w_1 - \ldots - b_n w_n = 0.$$

Since $w_1, \ldots, w_n, u_1, \ldots, u_l$ is linearly independent, this means that all the scalars has to be zero. This shows that $v = 0$. The same argument shows that

$$span(v_1, \ldots v_k) \cap (U_1 \cap U_2) = \{0\}$$

as well.

Now, we claim that

$$w_1, \ldots, w_n, u_1, \ldots, u_l, v_1, \ldots v_k$$

is a basis of $U_1 + U_2$. Let's first show linear independence. We take a linear dependence relation:

$$\underbrace{a_1 w_1 + \ldots + a_n w_n}_{w} + \underbrace{b_1 u_1 + \ldots + b_l u_l}_{u} + \underbrace{c_1 v_1 + \ldots + c_k v_k}_{v} = 0.$$

It immediately follows that u is an element of both U_1 and U_2. Therefore u is in both $U_1 \cap U_2$ and $span(u_1, \ldots, u_l)$. From our discussion above, this means $u = 0$. Similarly, $v = 0$ and therefore $w = 0$. This implies that all the scalars in the linear dependence relation needs to be 0, since $w_1, \ldots, w_n$ and $u_1, \ldots, u_l$ and $v_1, \ldots v_k$ are all linearly independent among themselves.

Second we show that this list spans $U_1 + U_2$. By definition any $u \in U_1 + U_2$ can be written as $\tilde{u_1} + \tilde{u_2}$ with $\tilde{u_1} \in U_1$ and $\tilde{u_2} \in U_2$[2]. Now write $\tilde{u_1}$ as a linear combination of $w_1, \ldots, w_n, u_1, \ldots, u_l$ and $\tilde{u_2}$ as a linear combination of $w_1, \ldots, w_n, v_1, \ldots v_k$. Adding them up and collecting the terms that have w_i's together using distributivity, we obtain a linear combination of $w_1, \ldots, w_n, u_1, \ldots, u_l, v_1, \ldots v_k$ that equals u. The proof that we have a basis of $U_1 + U_2$ is complete.

By definition, $dim(U_1 \cap U_2) = n$, $dim(U_1) = n + l$, $dim(U_2) = n + k$, and $dim(U_1 + U_2) = n + l + k$. This proves the claim. □

Exercise 32. Let $U_1, \ldots, U_n \subset V$ be subspaces such that $U_1 + \ldots + U_n$ is a direct sum. Prove that

$$dim(U_1 + \ldots + U_n) = dim(U_1) + \ldots + dim(U_n).$$

□

Remark 10. Notice that reordering the vectors in a finite list does not change their span, whether they are linearly independent or not, or whether they form a basis or not. □

Next week we continue with linear maps between vector spaces.

[2] I put tilde's to deal with the notation clash with basis elements

10. Lecture 10: Linear maps

The importance of the following definition cannot be overstated.

Definition 15. Let V and W be vector spaces over $\mathbb{F}$. A linear map $T : V \to W$ is a map satisfying the following properties:

- additivity: $T(u + v) = Tu + Tv$ for every $u, v \in V$
- homogeneity: $T(c \cdot v) = c \cdot Tv$ for every $c \in \mathbb{F}$ and $v \in V$

□

You have already seen these conditions in HW2 in the definition of an isomorphism of vector spaces. Indeed an isomorphism of vector spaces is nothing but a bijective linear map.

Exercise 33. Prove that a linear map has to send 0 to 0 □

Exercise 34. Prove that $T : V \to W$ is linear if and only if

$$T(c_1 \cdot u + c_2 \cdot v) = c_1 \cdot Tu + c_2 \cdot Tv$$

for every $c_1, c_2 \in \mathbb{F}$ and $u, v \in V$. □

Here are some examples of linear maps:

(1) A linear map $T : \mathbb{F}^1 \to \mathbb{F}^1$ is nothing but multiplication by $T(1) \in \mathbb{F}$. We will see pretty soon that a linear map $T : \mathbb{F}^m \to \mathbb{F}^n$ is nothing but multiplication of column vectors with m components by a matrix of size $n \times m$ with entries in $\mathbb{F}$.

(2) To get a sense of the definition let us take $\mathbb{F} = \mathbb{R}$ and analyze the image of a linear map $T : \mathbb{R}^2 \to \mathbb{R}^3$.

Consider the vectors $(1, 0)$ and $(0, 1)$ in the plane $\mathbb{R}^2$. Note that $(1, 0)$ and $(0, 1)$ clearly are linearly independent. They also span $\mathbb{R}^2$ since

$$(a, b) = a \cdot (1, 0) + b \cdot (0, 1).$$

Succintly, they are a basis. Let $T(1, 0) = v$ and $T(0, 1) = w$, which are vectors in $\mathbb{R}^3$. By Exercise 34,

$$T(a, b) = av + bw,$$

the image of T are all vectors in $\mathbb{R}^3$ that is a linear combination of v and w. We want to understand what this is geometrically.

It is possible that $v = w = 0$, in which case the entire $\mathbb{R}^2$ has to be sent to 0. A less trivial case is when v and w are contained in a unique line l in $\mathbb{R}^3$. This means that $a_1 v = a_2 w$, for some real numbers a_1, a_2 that are not both equal to zero. Any vector on l can be written as a scalar multiple of v or w (one of them is non-zero) and moreover any vector $av + bw$ has to lie on l. Therefore, the image of T is precisely l.

The last case is the most common one, namely that there is not a line that contains both v and w. In this case there is a unique plane $P \subset \mathbb{R}^3$ that contains both v and w. Geometrically it is clear that the sum of two vectors that are contained in a plane is also contained in the same plane via the parallelogram picture of vector addition. Moreover the scaling of a vector that is contained in a plane is also contained in the plane. These mean that the image of T has to be be contained in P. You had to essentially come up with this argument for the last problem of HW2.

Last but not least, we want to show that every vector u in P is in the image of T, or equivalently that every vector on P is $av + bw$ for some real numbers a and b. To see this draw the parallelogram contained in P with edges parallel to v and w that admits u as a diagonal. The picture shows that some (possibly negative) multiples of v and w adds up to u as desired. Hence, the image is the plane P.

(3) Make sure you understand that rotations and scalings of the plane give linear maps $\mathbb{R}^2 \to \mathbb{R}^2$. Multiplication with the matrix

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

i.e. $(a, b) \mapsto (a + b, b)$ is another example of a linear map $\mathbb{R}^2 \to \mathbb{R}^2$. This is a little harder to think about visually - it is called a shear.

The following lemma is sometimes called the linear map construction lemma. It makes it clear what the data of a linear map is.

Lemma 9. *Let V and W be vector spaces over $\mathbb{F}$. Let $v_1, \ldots, v_n$ be a basis for V and $w_1, \ldots, w_n \in W$ be arbitrary vectors. Then, there exists a unique linear map $T : V \to W$ such that $T(v_i) = w_i$ for every $i = 1, \ldots, n$.*

Proof. Let us start by proving the uniqueness. Assume that

$$v = a_1 v_1 + \ldots + a_n v_n.$$

Linearity of T implies that we need to have

$$\begin{aligned} Tv &= T(a_1 v_1 + \ldots + a_n v_n) \\ &= T(a_1 v_1) + \ldots + T(a_n v_n) \\ &= a_1 T(v_1) + \ldots + a_n T(v_n) \\ &= a_1 w_1 + \ldots + a_n w_n. \end{aligned}$$

By Proposition 6, we know that every $v \in V$ can be written uniquely as a linear combination of $v_1, \ldots, v_n$. As a result, we see that $T(v_i) = w_i$ for every $i = 1, \ldots, n$ determine what Tv has to be for every $v \in V$.

Now, we move on to the existence. From the uniqueness part we know that we have no other option but to define

$$T(v) = a_1 w_1 + \ldots + a_n w_n.$$

for $v = a_1 v_1 + \ldots + a_n v_n$. We need to check that T is linear and sends v_i to w_i for every $i = 1, \ldots, n$. The latter is clear. For the former let us start by proving the additivity of T.

If $v = a_1 v_1 + \ldots + a_n v_n$ and $u = b_1 v_1 + \ldots + b_n v_n$, we need to show that

$$T(v + u) = a_1 w_1 + \ldots + a_n w_n + b_1 w_1 + \ldots + b_n w_n.$$

Note that we have

$$v + u = (a_1 + b_1) v_1 + \ldots + (a_n + b_n) v_n.$$

This has to be the unique representation of $v + u$ as a linear combination of $v_1, \ldots, v_n$. By construction therefore

$$T(v + u) = (a_1 + b_1) w_1 + \ldots + (a_n + b_n) w_n.$$

This finishes the proof of additivity.

We move on to showing the homogeneity of T. If $v = a_1v_1 + \ldots + a_nv_n$ and $c \in \mathbb{F}$, we need to show that

$$T(cv) = c(a_1w_1 + \ldots + a_nw_n).$$

Note that we have

$$cv = ca_1v_1 + \ldots + ca_nv_n.$$

This has to be the unique representation of cv as a linear combination of $v_1, \ldots, v_n$. By construction therefore

$$T(cv) = ca_1w_1 + \ldots + ca_nw_n.$$

This finishes the proof of homogeneity and the entire proof. □

We end with a corollary of Lemma 9.

Corollary 2. *Let V and W be vector spaces over $\mathbb{F}$. Let $v_1, \ldots, v_n$ be a basis for V and $u_1, \ldots, u_m$ be one for W. Let $a_{ji} \in \mathbb{F}$, for $1 \leq j \leq m$ and $1 \leq i \leq n$, be arbitrary scalars. Then, there exists a unique linear map $T : V \to W$ such that*

$$T(v_i) = a_{1i}u_1 + \ldots + a_{mi}u_m$$

for every $i = 1, \ldots, n$.

Proof. This immediately follows from Proposition 6 and Lemma 9. □

This means that once bases are chosen a linear map between two finite dimensional vector spaces is nothing but a choice of a table of scalars. We will explore this further in the next class.

11. Lecture 11: Column vectors and matrices, invertible maps, composition of linear maps

Today we start with a look back at linear algebra as you probably knew it before this class; using column vectors and matrices.

A column vector with $n \geq 1$ entries in a field $\mathbb{F}$ is the same thing as an element of $\mathbb{F}^n$. The only difference is that instead of representing them as n-tuples $(a_1, \ldots, a_n)$ we represent them as columns of entries. This is entirely a convention. Sometimes we will think of elements of $\mathbb{F}^n$ as column vectors in this way.

An $m \times n$ matrix A with entries in $\mathbb{F}$ is simply an $m \times n$ table of scalars $A_{ji} \in \mathbb{F}$, for $1 \leq j \leq m$ and $1 \leq i \leq n$.

Lemma 10. *Let A be an $m \times n$ matrix and let us think of $\mathbb{F}^n$ and $\mathbb{F}^m$ as column vectors. Then matrix multiplication of column vectors define a linear map*

$$T_A : \mathbb{F}^n \to \mathbb{F}^m.$$

We will use the notation T_A in what follows so this lemma is in part a definition as well.

Exercise 35. Prove this important lemma. □

Let us continue exploring $\mathbb{F}^n$. Let us call

$$e_i = (0, \ldots, 0, \underset{\substack{\uparrow \\ i^{th}\text{-entry}}}{1}, 0, \ldots, 0)$$

the i^{th} standard basis vector $\mathbb{F}^n$.

Exercise 36. Prove that $e_1, \ldots, e_n$ is basis of $\mathbb{F}^n$. □

We call $e_1, \ldots, e_n$ the standard basis of $\mathbb{F}^n$. Let us rewrite Corollary 2 in a special case.

Corollary 3. *Let $e_1, \ldots, e_n$ be the standard basis for $\mathbb{F}^n$ and $f_1, \ldots, f_m$ be the one for $\mathbb{F}^m$. Let $a_{ji} \in \mathbb{F}$, for $1 \leq j \leq m$ and $1 \leq i \leq n$, be arbitrary scalars. Then, there exists a unique linear map $T : \mathbb{F}^n \to \mathbb{F}^m$ such that*

$$T(e_i) = a_{1i}f_1 + \ldots + a_{mi}f_m$$

for every $i = 1, \ldots, n$.

Defining a $m \times n$ matrix A by setting $A_{ji} = a_{ji}$, the map T in this statement is precisely the linear map $T_A : \mathbb{F}^n \to \mathbb{F}^m$ as in Lemma 10. To see this note that the linear map T_A sends e_i to the column vector

$$(a_{1i}, \ldots, a_{mi})^T,$$

which is nothing but $a_{1i}f_1 + \ldots + a_{mi}f_m$. Let us record this for future use.

Corollary 4. *Let $e_1, \ldots, e_n$ be the standard basis for $\mathbb{F}^n$ and $f_1, \ldots, f_m$ be the one for $\mathbb{F}^m$. Let $a_{ji} \in \mathbb{F}$, for $1 \leq j \leq m$ and $1 \leq i \leq n$, be arbitrary scalars and define the $m \times n$ matrix A by setting $A_{ji} = a_{ji}$. Then, $T_A : \mathbb{F}^n \to \mathbb{F}^m$ as in Lemma 10 is the unique linear map such that*

$$T(e_i) = a_{1i}f_1 + \ldots + a_{mi}f_m$$

for every $i = 1, \ldots, n$.

Here is a corollary of the corollary.

Corollary 5. *For every linear map $T : \mathbb{F}^n \to \mathbb{F}^m$, there exists a unique $m \times n$ matrix A with entries in $\mathbb{F}$ such that $T = T_A$.*

We will come back to matrices in this lecture to explain the origin of the matrix multiplication operation.

Recall that we had defined an isomorphism of vector spaces to be a bijective linear map. In your HW2 you showed the following:

Lemma 11. *Let V and W be vector spaces over $\mathbb{F}$ and $\phi : V \to W$ an isomorphism. Then, the inverse map $\phi^{-1} : W \to V$ is also linear and hence an isomorphism.*

Because of this lemma, isomorphisms of vector spaces are also invertible linear maps.

Definition 16. Let V and W be vector spaces over $\mathbb{F}$. We denote by

$$\mathcal{L}(V, W)$$

the set of all linear maps $V \to W$. □

We know that maps between sets $S \to S'$ and $S' \to S''$ can be composed. Now we show that composition of linear maps are linear.

Lemma 12. *Let U, V and W be vector spaces over $\mathbb{F}$; and $f \in \mathcal{L}(V, W)$ and $g \in \mathcal{L}(U, V)$, then the composition $f \circ g$ is a linear map as well:*

$$f \circ g \in \mathcal{L}(U, W).$$

Proof. We need to show that $f \circ g$ is additive and homogeneous.

- Additive: $f \circ g(u_1 + u_2) = f(g(u_1 + u_2)) = f(g(u_1) + g(u_2)) = f(g(u_1)) + f(g(u_2)) = f \circ g(u_1) + f \circ g(u_2)$
- Homogeneous: $f \circ g(cu) = f(g(cu)) = f(cg(u)) = cf(g(u) = cf \circ g(u)$

In the first item we used the additivity of f and g, and in the second one we used the homogeneity of f and g. □

We know that there is a one to one correspondence between linear maps $\mathbb{F}^n \to \mathbb{F}^m$ and $m \times n$ matrices as in Corolllary 5. Let A be an $m \times n$ matrix and B be a $k \times m$ matrix. We know that we can compose linear maps $T_A : \mathbb{F}^n \to \mathbb{F}^m$ and $T_B : \mathbb{F}^m \to \mathbb{F}^k$ to obtain a linear map $F^n \to F^k$, which has to be T_C for some $k \times n$ matrix C. This C nothing but the matrix multiplication of B and A:

$$C = BA.$$

This is best checked at home since it involves some book keeping of indices so I will make you do this in your HW4. Let us record it for the future:

Proposition 12. *Let A be an $m \times n$ matrix and B be a $k \times m$ matrix. Consider the linear maps $T_A : \mathbb{F}^n \to \mathbb{F}^m$ and $T_B : \mathbb{F}^m \to \mathbb{F}^k$ as in Lemma 10. Then,*

$$T_B \circ T_A = T_{BA}$$

as linear maps $\mathbb{F}^n \to \mathbb{F}^k$, where we BA is the matrix multiplication of B and A.

12. Lecture 12: Rank-nullity Theorem

Today we will prove the second result that I deemed to call a theorem. Before stating it we need a couple of definitions.

Definition 17. Let V and W be vector spaces over $\mathbb{F}$ and $T : V \to W$ a linear map. Then, the null-space of T are the subset of vectors in V that are mapped to 0 under T:

$$null(T) := \{v \in V \mid Tv = 0\}.$$

□

Lemma 13. *In the notation of the previous definition, $null(T)$ is a subspace.*

Exercise 37. Prove this lemma. □

Lemma 14. *In the notation of the previous definition, T is injective if and only if $null(T) = \{0\}$.*

Proof. If T is injective, there can be only one element $v \in V$ such that $Tv = 0$. We know that $v = 0$ is such an element, so $null(T) = \{0\}$.

The converse has more content. Assume $null(T) = \{0\}$ and $Tv_1 = Tv_2$. By linearity,

$$0 = Tv_1 - Tv_2 = T(v_1 - v_2).$$

Therefore $v_1 - v_2$ is in the null-space, which implies $v_1 = v_2$ as desired. □

Definition 18. Let V and W be vector spaces over $\mathbb{F}$ and $T : V \to W$ a linear map. Then, the image (or range) of T is the subset of vectors in W that are equal to Tv for some $v \in V$:

$$im(T) := \{w \in W \mid Tv = w \text{ for some } v \in V\}.$$

□

Lemma 15. *In the notation of the previous definition, $im(T)$ is a subspace.*

Exercise 38. Prove this lemma. □

We are ready to state out theorem, which is commonly called the rank-nullity theorem. Here rank refers to the dimension of $im(T)$ and nullity to the dimension of $null(T)$.

Theorem 2. *Let V be a finite dimensional vector space over $\mathbb{F}$ and W be a not-necessarily finite dimensional one. Let $T : V \to W$ be a linear map. Then*

(1) $im(T)$ is finite dimensional.
(2) $dim(V) = dim(null(T)) + dim(im(T))$.

Let us prove a preliminary lemma that will be useful in the proof first.

Remark 11. From now on, we will use that every finite dimensional vector space admits a basis (Proposition 7) and that every vector is a unique linear combination of a basis (Proposition 6) without mention. □

Lemma 16. *Let V be a finite dimensional vector space over $\mathbb{F}$ and $U \subset V$ be a subspace. Then, there exists another subspace $U' \subset V$ such that*

$$U \oplus U' = V.$$

Proof. Let $u_1, \ldots, u_m$ be a basis for U, which we know exists by Corollary 1. We can extend it to a basis of V using Lemma 8:

$$u_1, \ldots, u_m, u'_1, \ldots, u'_n.$$

Now choose

$$U' = span(u'_1, \ldots, u'_n).$$

Using that $u_1, \ldots, u_m, u'_1, \ldots, u'_n$ is a basis of V you quickly show that $U + U'$ is a direct sum and that every $v \in V$ can be written as $u + u'$. □

Exercise 39. Rigorously prove the last sentence of the proof. □

Proof of Theorem 2. Let us start with a special case. Namely assume that T is injective. Let $v_1, \ldots, v_n$ be a basis of V. Then, we claim that $Tv_1, \ldots, Tv_n$ is a basis of $im(T)$.

Let's first show that $Tv_1, \ldots, Tv_n$ span $im(T)$. Let $w \in im(T)$, which means that there exists $v \in V$ such that $Tv = w$. We can write v as a linear combination $a_1v_1 + \ldots + a_nv_n$ of $v_1, \ldots, v_n$. Then we have

$$\begin{aligned} Tv &= T(a_1v_1 + \ldots + a_nv_n) \\ &= T(a_1v_1) + \ldots + T(a_nv_n) \\ &= a_1Tv_1 + \ldots + a_nTv_n, \end{aligned}$$

which finishes the proof.

Next we show that $Tv_1, \ldots, Tv_n$ is linearly independent. Assume that $a_1Tv_1 + \ldots + a_nTv_n = 0$. This implies

$$T(a_1v_1 + \ldots + a_nv_n) = 0$$

by reading the equalities above backwards. By injectivity, $a_1v_1 + \ldots + a_nv_n = 0$ and by the linear independence of $v_1, \ldots, v_n$, we obtain $a_i = 0$, as desired.

$Tv_1, \dots, Tv_n$ being a basis of $im(T)$ implies (1) by the definition of finite dimensionality and (2) by the definition of dimension along with $null(T) = \{0\}$ (which is 0 dimensional).

With this special case at hand, we move on to the general case. We have that $null(T)$ is a subspace of V, so by Lemma 16, we can find a subspace $U \subset V$ such that $null(T) \oplus U = V$. Note that in this case we have

$$dim(V) = dim(U) + dim(null(T)), \tag{1}$$

by Proposition 11.

Let us consider the linear map $T|_U : U \to W$, which is obtained by restricting T to the subspace U. We claim two statements:

- $null(T|_U) = \{0\}$: this is true by the fact that $null(T) \cap U = \{0\}$.
- $im(T|_U) = im(T)$: this is true because for $v = n + u$ with $n \in null(T)$ and $u \in U$, $Tv = Tn + Tu = Tu$.

By Lemma 14, the first bullet point implies that $T|_U$ is injective. Hence, we can apply what we proved in the first part of the proof to conclude that $im(T|_U)$ is finite dimensional and $dim(im(T|_U)) = dim(U)$. Using the second bullet point we get that $im(T)$ is finite dimensional and

$$dim(im(T)) = dim(U).$$

Combining this with Equation (1), we get the desired equality.

□

Let us discuss some corollaries.

Corollary 6. *Let V and W be finite dimensional vector spaces over $\mathbb{F}$. Let $T : V \to W$ be an injective linear map, then $dim(V) \leq dim(W)$.*

Proof. We have $dim(V) = dim(im(T)) \leq dim(W)$. □

Exercise 40. Prove the following corollary of this corollary. Let V and W be finite dimensional vector spaces over $\mathbb{F}$. If $T : V \to W$ is an isomorphism, then $dim(V) = dim(W)$. □

In fact we can prove a stronger statement.

Proposition 13. *Let V and W be finite dimensional vector spaces over $\mathbb{F}$. Then V is isomorphic to W if and only if $dim(V) = dim(W)$.*

Proof. All that is left to prove is that if $dim(V) = dim(W)$, then there is an isomorphism $\phi : V \to W$. Let $v_1, \dots, v_n$ be a basis for V and $w_1, \dots, w_n$ be one for W. By Lemma 9, we know that we can choose a linear map $\phi : V \to W$ such that $\phi(v_i) = w_i$ for all $i = 1, \dots n$. One can prove in many ways that ϕ is an isomorphism. □

Exercise 41. Finish the proof. □

13. Lecture 13: Determinant of a matrix

Let $\mathbb{F}$ be a field as usual. In this lecture we will introduce the determinant of an $n \times n$ matrix with entries in $\mathbb{F}$. As you may have noticed the author of your textbook has certain feelings towards the determinant[3]. I agree with most of his opinions and we will keep following his approach.

On the other hand, determinants are magical. I think it is a shame to not talk about them at all in a linear algebra course. We will briefly cover determinants today and then mostly avoid them in the rest of the course.

From now on we will use the result that a linear map is injective if and only if its null space is $\{0\}$ without mentioning anything more about it.

We start with a lemma that will be used later in the course.

Lemma 17. *Let V and W be finite dimensional vector spaces over $\mathbb{F}$ with the same dimension. Let $T : V \to W$ be a linear map. Then,*

- *If T is injective, then it is an isomorphism.*
- *If T is surjective, then then it is an isomorphism.*

Exercise 42. These are direct consequences of the rank-nullity theorem. Prove them. □

We denote the set of $n \times n$ matrices with entries in $\mathbb{F}$ by $Mat(n, \mathbb{F})$. Determinant is a map

$$\det : Mat(n, \mathbb{F}) \to \mathbb{F},$$

defined for every $n \geq 1$.

Let us start by defining the determinant for small values of n. For $n = 1$, it is very easy to define. If our matrix is (a) then its determinant is a.

For $n = 2$, we define it with the following formula

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} := a_{11}a_{22} - a_{12}a_{21}$$

Let us note the following lemma, which will generalize to all $n \geq 1$.

Lemma 18. *Let n be 1 or 2. For any $n \times n$ matrix A, the map $T_A : \mathbb{F}^n \to \mathbb{F}^n$ is an isomorphism if and and only if $\det A \neq 0$.*

Proof. For $n = 1$, let $A = (a)$. We have $\det A = a$. If $a = 0$, T_A sends all elements of $\mathbb{F}$ to 0, and therefore it is not an isomorphism. If $a \neq 0$, then since $T_A(a^{-1}b) = b$, T_A is surjective. If $T_A(b) = ab = 0$, then we need to have $b = 0$, so the null-space is trivial and T_A is injective as well. We could also use Lemma 17 instead of one of the last two sentences but it would be an overkill.

For $n = 2$, let

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

First, let's assume $\det A \neq 0$. We will show that T_A is injective, which suffices by Lemma 17. Assume that $(v_1, v_2) \in \mathbb{F}^2$ is in $null(T_A)$, which concretely means

$$a_{11}v_1 + a_{12}v_2 = 0 \text{ and } a_{21}v_1 + a_{22}v_2 = 0$$

We multiply the equality on the left by a_{22}, the one on the right by a_{12}, and then subtract the second one from the first one. The result is

$$(a_{11}a_{22} - a_{12}a_{21})v_1 = 0.$$

[3] see `https://www.axler.net/DwD.html` for a clear picture

Since $\det A \neq 0$, $v_1 = 0$. By a similar argument, we can also show that $v_2 = 0$. Therefore, we have $null(T_A) = \{0\}$, which proves the injectivity.

Conversely, assume that T_A is an isomorphism. Then at least one of the entries of A has to be non-zero. Assume that one of a_{11} or a_{12} is not zero. We know that $T_A(-a_{12}, a_{11}) \neq 0$. Computing the LHS, we see that it is equal to $(0, \det A)$. Hence, we have $\det A \neq 0$ as desired. If one of a_{21} or a_{22} is not zero, then a similar argument proves the claim. □

Exercise 43. Explicitly find the "similar arguments" above. □

Let us move on to $n = 3$,

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

Now the determinant is defined as

$$\det A = a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} \dots$$

with four more terms that I did not write. Instead of writing them at this point we should make some definitions that will allow us to define the determinant in general.

Definition 19. Let $n \geq 1$. A permutation of $[n] := \{1, \dots, n\}$ is a bijection

$$\sigma : [n] \to [n].$$

The set of all permutations of $[n]$ is denoted by Σ_n. □

A common way to represent a permutation $\sigma \in \Sigma_n$, is to write

$$(\sigma(1)\sigma(2)\dots\sigma(n))$$

for it. For example for $n = 3$, the permutation defined by

$$\sigma(1) = 3, \sigma(2) = 2, \sigma(3) = 1 \text{ is represented by } (321).$$

Exercise 44. Make sure you understand why this is a valid representation of a permutation. This should in particular show you that your previous conception of a permutation is the same as what we are talking about here. □

Definition 20. To each $\sigma \in \Sigma_n$, we can associate its sign, which is equal to either -1 or 1. We define it by

$$sign(\sigma) = (-1)^{\#A},$$

where $\#A$ is the number of pairs $1 \leq i < j \leq n$ such that $\sigma(i) > \sigma(j)$. □

As an example $sign(312) = 1$, but $sign(21) = -1$.

Let us now go back to matrices. Fix an $n \times n$ matrix A with entries A_{ji}, $i, j \in [n]$. Let $\sigma \in \Sigma(n)$. We define

$$\sigma(A) := \text{sign}(\sigma) A_{\sigma(1)1} \dots A_{\sigma(n)n}.$$

Here is what this is in words: from the i^{th} column you choose the $\sigma(i)^{th}$ entry from the top for all columns, multiply each of these entries and finally multiply the whole thing with the sign of the permutation.

Definition 21. The determinant of an $n \times n$ matrix A is defined as:

$$\det A = \sum_{\sigma \in \Sigma_n} \sigma(A),$$

where $\sigma(A) := \operatorname{sign}(\sigma) A_{\sigma(1)1} \dots A_{\sigma(n)n}$. □

Exercise 45. Check that this agrees with our previous definitions for $n = 1, 2$. Write the full expression for $n = 3$. □

The main use of determinants in linear algebra stems from the following lemma that we already mentioned above.

Lemma 19. *Let $n \geq 1$. For any $n \times n$ matrix A, the map $T_A : \mathbb{F}^n \to \mathbb{F}^n$ is an isomorphism if and and only if* $\det A \neq 0$.

We will not use this lemma but we can still appreciate its beauty. Such a complicated condition as T_A being an isomorphism can be tested by checking whether a single explicit function of the entries of A vanishes or not.

The determinant behaves very nicely under row operations. Swapping two rows negates the determinant, multiplying a row with a scalar multiplies the determinant by that same scalar and adding a row to another one does not change the determinant. The proof of the lemma can be given using these properties but we omit it.

In fact determinant is the unique map $\det : Mat(n, \mathbb{F}) \to \mathbb{F}$ that satisfies these three properties in regards to row operations plus $\det I = 1$.

I want to finish by pointing out the geometric meaning of the determinant when $\mathbb{F} = \mathbb{R}$. Take an $n \times n$ matrix A and consider the n vectors in $\mathbb{R}^n$ given by the columns of A. Now take a parallelepiped $P(A)$ with edges parallel to these vectors. Then, we have that the volume of $P(A)$ is given by the absolute value of the determinant:

$$vol(P(A)) = |\det A|$$

Exercise 46. Explicitly check this for $n = 2$. □

Remark 12. Here is a purely geometric proof of Lemma 19 in this case. Note that the i^{th} column vector of A is nothing but $T_A e_i$. We know that T_A is not an isomorphism if and only if it is not surjective. The latter is equivalent to the image of T_A being contained in an $n-1$ dimensional subspace, or equivalently $P(A)$ being contained in an $n-1$ dimensional subspace, or even better $P(A)$ having zero volume. By the above formula, this in turn is the same as $\det A = 0$. □

14. Lecture 14: Polynomials

Today we will cover some basic properties of polynomials over fields, giving special attention to $\mathbb{F} = \mathbb{C}$.

A polynomial over a field $\mathbb{F}$ is a formal expression

$$a_d z^d + \ldots + a_1 z + a_0$$

with d a non-negative integer and $a_0, \ldots, a_d \in \mathbb{F}$. We will use the convention that $a_d \neq 0$ when we write a non-zero polynomial and call d the degree of the polynomial. The degree of the zero polynomial is defined to be $-\infty$.

For your homework, you had to think about the vector space structure on $P(\mathbb{F})$, the set of polynomials over $\mathbb{F}$.

We can also multiply polynomials

$$(a_nz^n+\ldots+a_1z+a_0)(b_mz^m+\ldots+b_1z+b_0)=$$
$$a_nb_mz^{m+n}+(a_nb_{m-1}+a_{n-1}b_m)z^{m+n-1}+\ldots+(a_1b_0+a_0b_1)z+a_0b_0.$$

You have seen this before except that now the coefficients are from an arbitrary field. Polynomial multiplication is commutative, associative, and it distributes over the addition of polynomials. It does not admit inverses for positive degree polynomials!

One can also do polynomial division with remainder. Formally this is similar to what one can do with integers. Let us write it as a lemma.

Lemma 20. *Let $P(z)=a_mz^m+\ldots+a_1z+a_0$ and $p(z)=b_nz^n+\ldots+b_1z+b_0$ be polynomials over $\mathbb{F}$. Assume that $p(z)$ is non-zero. Then there are unique polynomials $q(z)=c_lz^l+\ldots+c_1z+c_0$ and $r(z)=d_kz^k+\ldots+d_1z+d_0$ such that*

- $k<n$
- $P(z)=p(z)q(z)+r(z)$

Proof. Let us first prove the existence part. Assume the contrary, that there are $P(z)$ and $p(z)$ such that there is no such $q(z)$ and $r(z)$. Take such $P(z)=a_mz^m+\ldots+a_1z+a_0$ and $p(z)=b_nz^n+\ldots+b_1z+b_0$ with the property that the degree of $P(z)$ is minimal among such pairs of polynomials. First of all $m\geq n$ because otherwise we could choose $q(z)=0$ and $r(z)=P(z)$, which would be a contradiction to the non-existence that we assumed. Then, we can define

$$\tilde{P}(z):=P(z)-\frac{a_n}{b_m}z^{m-n}p(z).$$

Notice that the degree of $\tilde{P}(z)$ is strictly less than the degree of $P(z)$. Therefore, we can find $\tilde{q}(z)$ and $\tilde{r}(z)$ such that $deg(\tilde{r}(z))<n$ and

$$\tilde{P}(z)=p(z)\tilde{q}(z)+\tilde{r}(z).$$

Combining the two inequalities, we find

$$P(z)=p(z)\left(\frac{a_n}{b_m}z^{m-n}+\tilde{q}(z)\right)+\tilde{r}(z),$$

which is a contradiction. We have proved the existence.

Let us now move on to uniqueness. By a simple argument using distributivity, we see that it suffices to show that if $p(z)q(z)+r(z)=0$ with $deg(r(z))<n$, then $q(z)$ and $r(z)$ must be both the zero polynomial. It suffices to show that $q(z)$ is the zero polynomial. Assuming otherwise, we write $q(z)=c_lz^l+\ldots+c_1z+c_0$ with $c_l\neq 0$. Expanding the equation:

$$(b_nz^n+\ldots+b_1z+b_0)(c_lz^l+\ldots+c_1z+c_0)+r(z)=0,$$

and using $deg(r(z))<n$, we find that $b_nc_l=0$, which implies $c_l=0$. This is contradiction. The proof is complete. □

What makes polynomials interesting is the fact that they give rise to maps $\mathbb{F}\to\mathbb{F}$. Namely, given $p(z)=a_mz^m+\ldots+a_1z+a_0$ over $\mathbb{F}$, we can define a map $\mathbb{F}\to\mathbb{F}$ by

$$c\mapsto p(c):=a_m\cdot c^m+\ldots+a_1\cdot c+a_0\in\mathbb{F}$$

for every $c\in\mathbb{F}$. Here to make the point I have explicitly indicated the multiplication in $\mathbb{F}$ for once.

For example if $\mathbb{F} = \mathbb{R}$, and we are given a polynomial $p(z)$ with real coefficients, we obtain a function $\mathbb{R} \to \mathbb{R}$. We can draw its graph etc.

We also talked briefly about solving polynomial equations like

$$z^2 + 2z + 2 = 0$$

over $\mathbb{F}$. This of means nothing but finding the elements of $\mathbb{F}$ which map to zero under the map $\mathbb{F} \to \mathbb{F}$ defined by the polynomial $z^2 + 2z + 2$.

In general for any polynomial $p(z) = a_m z^m + \ldots + a_1 z + a_0$, if $c \in F$ is sent to 0 under the corresponding map $\mathbb{F} \to \mathbb{F}$, we say that c is a root of $p(z)$.

We will heavily rely on the following statement regarding roots of polynomials over complex numbers in the next weeks. It is called the fundamental theorem of algebra.

Theorem 3. *Every complex polynomial with positive degree has a root.*

I will give a sketch proof of this theorem. Let us consider $p(z) = a_m z^m + \ldots + a_1 z + a_0$ and the map that it defines $f : \mathbb{C} \to \mathbb{C}$. If $a_0 = 0$, we are done since 0 is a root. So assume $a_0 \neq 0$.

For any $r > 0$, let $C_r := \{x \in \mathbb{C} \mid |x| = r\} \subset \mathbb{C}$ be the circle of radius r centered at $0 \in \mathbb{C}$.

For $x \in \mathbb{C}$ with $|x|$ very small, the terms $a_i x^i$ for $i > 0$ have very small absolute value compared to a_0. Therefore, for r sufficiently small the image of C_r under f stays very close to a_0.

For $x \in \mathbb{C}$ with $|x|$ very large, the terms $a_i x^i$ for $i < m$ have very small absolute value compared to $a_m x^m$. We can parametrize C_r as $re^{i\theta}$ with $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. Plugging this in $a_m x^m$, we obtain

$$|a_i| r^m e^{i(m\theta + \arg a_i)}.$$

The means that the image of C_r under f for r very large winds around the origin $m \geq 1$ times.

Now we consider images of C_r under f for r changing from very large to very small. The image varies continuously with r. Assume that there is no root of $p(z)$. This means that the image of no C_r passes through the origin.

Think of the images of C_r as a rope on the floor and assume that there is a stick at a point. For large r we have that the rope goes around the stick m times. For small r we have that the rope is not going around stick at all, it is bundled up at some other point on the floor. We also know that there is a continuous movement of the rope that takes it from the first position to the second without ever crossing the stick. This is impossible! The polynomial needs to have a root.

This is possibly my favorite proof in all of mathematics. Here is a corollary.

Corollary 7. *For every complex polynomial $p(z) = a_m z^m + \ldots + a_1 z + a_0$ there exists $a, \lambda_1, \ldots, \lambda_m \in \mathbb{C}$ such that*

$$p(z) = a_m z^m + \ldots + a_1 z + a_0 = a(z - \lambda_1) \ldots (z - \lambda_m).$$

Before I prove this lemma, I have to mention something that is a bit confusing. Let $p(z) = b_m z^m + \ldots + b_1 z + b_0$and $q(z) = c_l z^l + \ldots + c_1 z + c_0$ be polynomials over $\mathbb{F}$. We can substitute $c \in \mathbb{F}$ in any polynomial over $\mathbb{F}$ as we did above when we defined the function $\mathbb{F} \to \mathbb{F}$ obtained from a polynomial. Again as above we denote the resulting element of $\mathbb{F}$ by putting c in the argument of the polynomial. Then, we have

$$p(c)+q(c)=(p+q)(c) \text{ and } p(c)q(c)=pq(c),$$

where on the RHS we have the sum and product polynomials.

Remark 13. The purely formal polynomial addition and multiplication operations are defined the way they are so that these equalities are true. This is why these statements look like they don't need proving - they do, but the proofs are confusingly easy. These statements will look less trivial when we consider substituting operators into polynomials in Lecture 17. □

Proof of Corollary 7. By fundamental theorem of algebra, we have that there is a root λ_1. Now let us apply Lemma 20, where we divide $p(z)$ by the polynomial $z-\lambda_1$. We have $p(z)=(z-\lambda)q(z)+r(z)$, where $r(z)=r_0$ has degree 0. Now substituting λ in this equation and using the rules above, we see that $r_0=0$. Therefore, we obtain

$$p(z)=(z-\lambda)q(z).$$

If $q(z)$ has degree 0, then we are done. Otherwise we apply the same procedure to $q(z)$ and so on. Since we reduce the degree by 1 every time, eventually we get what we want. □

Of course $\lambda_1,\dots,\lambda_m\in\mathbb{C}$ in the statement are precisely the roots of $p(z)$. Note that some λ_i's can be the same complex number. We will come back to this later.

15. Lecture 15: Matrix of a linear map with respect to bases, operators

Let $\mathbb{F}$ be a field and V,W be finite dimensional vector spaces over $\mathbb{F}$ throughout this lecture.

Let $v_1,\dots,v_n$ be a basis for V and $u_1,\dots,u_m$ be one for W. We define the matrix $M(T,v_1,\dots,v_n,u_1,\dots,u_m)$ of a linear map $T:V\to W$ with respect to these bases as follows. For every $1\le i\le n$, we obtain unique scalars $a_{1i},\dots,a_{mi}\in\mathbb{F}$ such that

$$T(v_i)=a_{1i}u_1+\dots+a_{mi}u_m.$$

We define the $(j,i)^{th}$ entry of the $m\times n$ matrix $M(T,v_1,\dots,v_n,u_1,\dots,u_m)$ as a_{ji}.

Let us denote the set of $m\times n$ matrices by $Mat(m,n,\mathbb{F})$. By sending each linear map $T:V\to W$ to $M(T,v_1,\dots,v_n,u_1,\dots,u_m)$, we define a map

$$\mathcal{M}(v_1,\dots,v_n,u_1,\dots,u_m):\mathcal{L}(V,W)\to Mat(m,n,\mathbb{F}).$$

Proposition 14. *$\mathcal{M}(v_1,\dots,v_n,u_1,\dots,u_m)$ is a bijection.*

Proof. Let us denote $\mathcal{M}(v_1,\dots,v_n,u_1,\dots,u_m)$ by $\mathcal{M}$ during the course of this proof.

We recall Corollary 4 from Lecture 10. It said that if $a_{ji}\in\mathbb{F}$, for $1\le j\le m$ and $1\le i\le n$, are arbitrary scalars, then there exists a unique linear map $T:V\to W$ such that

$$T(v_i)=a_{1i}u_1+\dots+a_{mi}u_m$$

for every $i=1,\dots,n$. This defines a map

$$\mathcal{T}:Mat(m,n,\mathbb{F})\to\mathcal{L}(V,W).$$

It is also immediate from definitions that

$$\mathcal{M} \circ \mathcal{T} : Mat(m, n, \mathbb{F}) \to Mat(m, n, \mathbb{F})$$

and

$$\mathcal{T} \circ \mathcal{M} : \mathcal{L}(V, W) \to \mathcal{L}(V, W)$$

are the identity maps. This finishes the proof. □

Exercise 47. Via entry-wise addition and entry-wise scalar multiplication we can equip $Mat(m, n, \mathbb{F})$ with a vector space structure over $\mathbb{F}$. Prove that $Mat(m, n, \mathbb{F})$ is isomorphic to $\mathbb{F}^{mn}$. □

Exercise 48. Equip $\mathcal{L}(V, W)$ with its natural vector space structure over $\mathbb{F}$ - there is only one natural way of doing this, I will leave it to you but if you are not sure you can look at your book. Finite dimensionality of V and W does not play a role here.

Then, show that $\mathcal{M}(v_1, \dots, v_n, u_1, \dots, u_m)$ is a linear map, and therefore an isomorphism of vector spaces. □

It is extremely important to understand that we can talk about the matrix corresponding to a linear map between finite dimensional vector spaces only after we choose bases. For the same linear map different choices of bases lead to different matrices.

Here is a question: given a linear map $T : V \to W$, can we choose the bases $v_1, \dots, v_n$ of V and $u_1, \dots, u_m$ of W such that the matrix of T becomes very simple? The answer is yes.

Proposition 15. *Let V be n dimensional and W be m dimensional. Take a linear map $T : V \to W$ whose image is k dimensional. Then, one can choose bases for V and W such that the matrix of T is*

$$\left(\begin{array}{c|c} \mathbf{Id}_k & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right).$$

In case $n = k$ and/or $m = k$ some of the 0-blocks are actually not there.

Proof. We follow the strategy in the proof of rank-nullity theorem. We have that $null(T)$ is a subspace of V, so by Lemma 16, we can find a subspace $U \subset V$ such that $null(T) \oplus U = V$. As we had shown in that proof, $T|_U$ is injective and $im(T|_U) = im(T)$. Also note that U is k-dimensional, which follows from the rank-nullity theorem.

Take bases $v_1, \dots, v_k$ of U and $v_{k+1}, \dots, v_n$ of $null(T)$. Let

$$u_i = Tv_i, \text{ for } i = 1, \dots k.$$

Then, $u_1, \dots, u_k$ is a basis for $im(T)$. See the proof of rank-nullity theorem if you can't figure out why. Finally, extend $u_1, \dots, u_k$ to a basis of W by adding the vectors $u_{k+1}, \dots, u_m$ using Lemma 8.

One easily sees that the bases $v_1, \dots, v_n$ of U and $u_1, \dots, u_m$ of W give the desired matrix for T, since

$$Tv_1 = u_1, \dots, Tv_k = u_k, Tv_{k+1} = 0, \dots, Tv_n = 0.$$

□

Definition 22. Let U be a vector space (not necessarily finite dimensional) over $\mathbb{F}$. We call a linear map $T : U \to U$ an operator on U. The set of operators on U is denoted by $\mathcal{L}(U)$. □

Given an operator T on the finite dimensional V, using Proposition 15, we can find two different bases of V which make the matrix of T very simple. But now we can ask for something more. What if we are allowed to choose only one basis of V and write the matrix of T using that basis in both roles?

More succintly, for a basis $v_1, \dots, v_n$ of V, and an operator $T : V \to V$, we define the $n \times n$ matrix:

$$M(T, v_1, \dots, v_n) := M(T, v_1, \dots, v_n, v_1, \dots, v_n).$$

The question of how can we choose $v_1, \dots, v_n$ so that $M(T, v_1, \dots, v_n)$ looks as simple as possible is a very important one that we will be dealing with for a while. It is significantly more difficult than Proposition 15. In fact it does not have a uniform answer for all fields $\mathbb{F}$ and we will soon start specifying $\mathbb{F}$ to be $\mathbb{C}$ or $\mathbb{R}$. Even though perhaps it is not clear now, because of the fundamental theorem of algebra $\mathbb{C}$ gives much better results in regards to this question.

16. Lecture 16: Eigenvalues/vectors, invariant subspaces

Recall the question we raised last time: given $T \in \mathcal{L}(V)$, where V is a finite dimensional vector space, how can we choose $v_1, \dots, v_n$ so that $M(T, v_1, \dots, v_n)$ is as simple as possible?

The simplest matrices without quesion are diagonal matrices. Unfortunately, we cannot always choose a basis so that the matrix is diagonal. But, we will see how close we can get. The following definition starts our investigation.

Definition 23. Let V be a vector space over an arbitrary field $\mathbb{F}$ and T be a linear operator on V.

- A scalar $\lambda \in \mathbb{F}$ is called an eigenvalue of T if there exists a non-zero vector $v \in V$ such that $Tv = \lambda v$.
- A vector $v \in V$ is called an eigenvector if $Tv = \lambda v$, for some $\lambda \in \mathbb{F}$.

□

Every non-zero eigenvector has a unique eigenvalue. On the other hand, for a given eigenvalue λ there are always multiple $v \in V$ which satisfy $Tv = \lambda v$. For example if v is one, then so is cv for all $c \in \mathbb{F}$.

Exercise 49. Rigorously prove all these statements. □

Eigenvalues do not always exist. Let us show this on an example. Consider $\mathbb{F} = \mathbb{R}$ and $T_A \in \mathcal{L}(\mathbb{R}^2)$ with

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

We will show that T_A has no eigenvalue. Assume that there is one. Namely that for some non-zero vector $(a, b) \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$, we have

$$T_A(a, b) = (\lambda a, \lambda b).$$

Computing the left hand side gives

$$(-b, a) = (\lambda a, \lambda b).$$

It follows that both a and b has to non-zero, which means $ab \neq 0$. Combining the equality of the two components we find

$$\lambda^2 ab = -ab.$$

Cancelling ab, we see that λ has to satisfy

$$\lambda^2 = -1.$$

This is a contradiction since we know that the square of a real number is non-negative.

On the other hand, if we consider the same matrix as defining an operator on $\mathbb{C}^2$, which is a vector space over $\mathbb{C}$, then we would find two eigenvalues $\pm i$. In fact, next class we will prove that every operator on a finite dimensional vector space over $\mathbb{C}$ has an eigenvalue!

Exercise 50. We can think of A as a matrix with entries in $\mathbb{F}_p$, p a prime number as well. For what values of p does T_A have an eigenvalue? Hint: recall your first homework. □

Lemma 21. *Let V be a vector space over an arbitrary field $\mathbb{F}$ and T be a linear operator on V. Then, $\lambda \in \mathbb{F}$ is an eigenvalue of T if and only if $T - \lambda Id : V \to V$ is not injective.*

Exercise 51. This is a good exercise in definitions, write down the proof. □

Using Lemma 17, we see that if V is finite dimensional, then $T \in \mathcal{L}(V)$ is an isomorphism if and only if it is injective. Hence, we get as a corollary.

Lemma 22. *Let V be a finite dimensional vector space over an arbitrary field $\mathbb{F}$ and T be a linear operator on V. Then, $\lambda \in \mathbb{F}$ is an eigenvalue of T if and only if $T - \lambda Id : V \to V$ is not an isomorphism.*

Let us end this lecture with another important notion.

Definition 24. Let V be a vector space over an arbitrary field $\mathbb{F}$ and T be a linear operator on V. We call a subspace $U \subset V$ an invariant subspace of T is it is closed under T, that is if

$$Tu \in U, \text{ for every } u \in U.$$

□

Exercise 52. Prove that the span of an eigenvector is an invariant subspace. Also prove a converse: if U is an invariant subspace of dimension 1, all the vectors in U are eigenvectors with the same eigenvalue. □

17. Lecture 17: Existence of eigenvalues over complex numbers

We start with a preliminary notion. We discussed in Lecture 14 what it means to substitute scalars into polynomials. We will now substitute operators.

Let V be a vector space over an arbitrary field $\mathbb{F}$ and T_1, T_2 be linear operators on V. Then, we can compose T_1 and T_2 to obtain an operator on V: $T_2 \circ T_1 : V \to V$.

Remark 14. Operator composition (and more specifically multiplication of square matrices of the same size) is not commutative. □

Recall that $\mathcal{L}(V)$ is also a vector space. In case you did not solve Exercise 48 from Lecture 15:

- For $c \in \mathbb{F}$ and $T \in \mathcal{L}(V)$, we define $cT \in \mathcal{L}(V)$ by
$$(cT)v = c(Tv), \text{ for every } v \in V.$$
- For $T_1, T_2 \in \mathcal{L}(V)$, we define $T_1 + T_2 \in \mathcal{L}(V)$ by
$$(T_1 + T_2)v = T_1v + T_2v, \text{ for every } v \in V.$$

Exercise 53. Prove that operator composition distributes over operator addition on either side. □

Definition 25. Let V be a vector space over an arbitrary field $\mathbb{F}$, T be a linear operator on V and $p(z) = a_dz^d + \ldots + a_1z + a_0$ be a polynomial over $\mathbb{F}$.

We define $p(T) \in \mathcal{L}(V)$ as
$$p(T) := a_dT^d + \ldots + a_1T + a_0\mathrm{Id}.$$
Here T^n means $\underbrace{T \circ \ldots \circ T}_{n\ T's}$. □

Please take a look at Remark 13, before you proceed. With notation from the definition, we have the trivial equality
$$(cp)(T) = c(p(T)), \text{ for every } c \in \mathbb{F}.$$

If $q(z)$ is another polynomial over $\mathbb{F}$, we have another trivial equality
$$(p + q)(T) = p(T) + q(T).$$
The next one is less trivial:

Lemma 23. *We have*
$$(pq)(T) = p(T) \circ q(T),$$
where on the left hand side we used polynomial multiplication.

Proof. First, by the distributivity of operator composition we have
$$\begin{aligned}(a_dT^d + \ldots + a_1T + a_0\mathrm{Id}) \circ (bT^k) &= a_dbT^d \circ T^k + \ldots + a_1bT \circ T^k + a_0\mathrm{Id} \circ T^k \\ &= a_dbT^{d+k} + \ldots + a_1bT^{k+1} + a_0T^k\end{aligned}$$
In other words we proved the result when $q(z) = bz^k$ for some $b \in \mathbb{F}$ and $k \geq 0$.

Next, assume that we have proven the result for $p(z), q(z)$ and also for $p(z), q'(z)$. We show that the result for $p(z), q(z) + q'(z)$ follows. Note that polynomial multiplication also distributes over polynomial addition, We have
$$\begin{aligned}(p(q + q'))(T) &= (pq + pq')(T) \\ &= (pq)(T) + (pq')(T) \\ &= p(T) \circ q(T) + p(T) \circ q'(T)) \\ &= p(T) \circ (q(T) + q'(T)) \\ &= p(T) \circ (q + q')(T)\end{aligned}$$

Since every polynomial is a sum of polynomials of the form bz^k, the result follows. □

Recall that in Lecture 14 we proved that for every complex polynomial $p(z) = a_m z^m + \ldots + a_1 z + a_0$ there exists $a, \lambda_1, \ldots, \lambda_m \in \mathbb{C}$ such that

$$p(z) = a_m z^m + \ldots + a_1 z + a_0 = a(z - \lambda_1) \ldots (z - \lambda_m).$$

This is Corollary 7, which was an immediate consequence of the fundamental theorem of algebra. In fact it also follows that $a = a_m$.

In particular by Lemma 23 and the discussion preceeding it, for any $T \in \mathcal{L}(V)$, we have

$$a_m T^m + \ldots + a_1 T + a_0 \mathrm{Id} = a_m (T - \lambda_1 \mathrm{Id}) \circ \ldots \circ (z - \lambda_m \mathrm{Id}).$$

Here is what we have been building towards.

Theorem 4. *Let V be a finite dimensional vector space over $\mathbb{C}$ and T be an operator on V. Then T has an eigenvalue.*

Proof. Let V be n-dimensional. Choose a non-zero $v \in V$ and consider the list of vectors

$$T^n v, T^{n-1} v, \ldots, Tv, v.$$

Since this list has more elements than the dimension, it has to be linearly dependent. Therefore, there exists $n \geq m \geq 1$ and complex numbers $a_m, \ldots, a_1, a_0$ such that

$$a_m T^m v + a_{m-1} T^{m-1} v + \ldots + a_1 T v + a_0 v = 0$$

and $a_m \neq 0$. This means that the operator

$$a_m T^m + a_{m-1} T^{m-1} + \ldots + a_1 T v + a_0 \mathrm{Id}$$

is not an isomorphism.

As we discussed above, there are $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$ such that

$$a_m T^m + \ldots + a_1 T + a_0 \mathrm{Id} = a_m (T - \lambda_1 \mathrm{Id}) \circ \ldots \circ (T - \lambda_m \mathrm{Id}).$$

We know that the left hand side is not an isomorphism, so neither is the right hand side. This means for at least one $1 \leq i \leq m$, $T - \lambda_i \mathrm{Id}$ is not an isomorphism, as compositions of isomorphisms are isomorphisms.

By the simple Lemma 22 from the previous lecture, this means that that λ_i is an eigenvalue, proving the desired claim. □

This is a beautiful proof. It is not the one that is most common. The most common proof uses determinants and in particular Lemma 19.

Exercise 54. Prove Theorem 4 using Lemma 19 and the fundamental theorem of algebra (no need to consider the factorization to linear polynomials, just that there is a root). □

What's bad about this proof? I really do not think it is bad. It is perhaps less elegant but it is definitely not bad. It's worth knowing.

18. Lecture 18: Splittings and block matrices, triangular matrices for operators over complex numbers

Today, we will show that the matrix of an operator on a finite dimensional vector space over $\mathbb{C}$ can be made upper triangular by choosing the right basis.

Remark 15. Let me first give you a heads up that if you are not quick about how to read what Tv_i is from the matrix $M(T, v_1, \ldots, v_n, u_1, \ldots, u_m)$, you will find it difficult follow the arguments below. This is straightforward, you look at the i^{th} column and take the linear combination of $u_1, \ldots, u_m$ with the entries from that column. After you get this straight, you should also be able to apply T to a linear combination of $v_1, \ldots, v_n$ and figure out the result using linearity of T. Note that when T is an operator, we generally take $v_1, \ldots, v_n$ and $u_1, \ldots, u_m$ to be the same basis.

All of this can be interpreted using the column vector representations of the vectors as explained in Homework 5, but that is not necessary. I am not even sure if it is helpful. □

We start with some general observations about block matrices. Most of the routine proofs below are left to you as exercises and these will be part of the homework. First, we give a definition that is not as standard as the other ones we made in this class, but I find it helpful.

Definition 26. Let V be a vector space over $\mathbb{F}$. We call a pair of subspaces $U_1, U_2 \subset V$ such that $U_1 \oplus U_2 = V$ a splitting of V. It will be convenient to say "Let $U_1 \oplus U_2 = V$... be a splitting" etc. as a shortcut in what follows.

If $U \subset V$ is a subspace, then another subspace $U' \subset V$ is called a complement to U if U, U' is a splitting, that is $U \oplus U' = V$. □

We have proved in Lemma 16 that if V is finite dimensional then any subspace $U \subset V$ admits a complement. We have already used this in a couple of proofs.

Given a splitting $U_1 \oplus U_2 = V$, we obtain linear maps $p_1 : V \to U_1$ and $p_2 : V \to U_2$. The first one is obtained by writing $v = u_1 + u_2$ with $u_1 \in U_1$ and $u_2 \in U_2$ and defining $p_1(v) = u_1$.

Exercise 55. Check that we indeed have a well-defined linear map $p_1 : V \to U_1$. Define the map $p_2 : V \to U_2$. □

Now let's take an operator $T : V \to V$. For every $i, j \in \{1, 2\}$, we can define a linear map

$$T_{ij} := p_j \circ T|_{U_i} : U_i \to U_j.$$

Recall that $T|_{U_i}$ is the map obtained by restricting the domain of T to U_i.

Lemma 24. *Let V be a finite dimensional vector space over $\mathbb{F}$, T an operator on V and $U_1 \oplus U_2 = V$ a splitting. Define $T_{ij} : U_i \to U_j$ as above for every $i, j \in \{1, 2\}$.*

Choose a basis $u^1_1, \ldots, u^1_k$ for U_1 and $u^2_1, \ldots, u^2_l$ for U_2, and define

$$M_{ji} := M(T_{ij}, (u^i_1, \ldots), (u^j_1, \ldots)).$$

The matrix of T with respect to the basis $u^1_1, \ldots, u^1_k, u^2_1, \ldots, u^2_l$ has the following block form:

$$\left(\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{21} & M_{22} \end{array}\right).$$

Exercise 56. This is a long statement and the proof is an exercise in definitions. Write down everything carefully and the give the proof. □

Note that T_{11} and T_{22} are operators and M_{11} and M_{22} are defined using the same bases for both roles as in T and $M(T, u_1^1, \ldots, u_k^1, u_1^2, \ldots, u_l^2)$.

We stress a special case.

Corollary 8. *With notation of Lemma 24,*

- *If U_1 from the statement is an invariant subspace of T, then T_{12} sends everything to zero. This means that the block matrix is of the form*
$$\left(\begin{array}{c|c} M_{11} & M_{12} \\ \hline 0 & M_{22} \end{array}\right).$$
- *If U_2 is also an invariant subspace of T, then the block matrix is even simpler*
$$\left(\begin{array}{c|c} M_{11} & 0 \\ \hline 0 & M_{22} \end{array}\right).$$

Exercise 57. Again, the proof is left to you as an exercise. □

Now let us finally do what we promised in the beginning.

Theorem 5. *Let V be a finite dimensional vector space over $\mathbb{C}$ and T be an operator on V. Then we can find a basis $v_1, \ldots, v_n$ of V such that*
$$M(T, v_1, \ldots, v_n)$$
is upper triangular.

Proof. We will prove this by induction on the dimension on V. If $dim(V) = 1$, the statement is trivial (all matrices are upper triangular). Let us now assume that the statement is true when V is $n-1$ dimensional and prove it when it is n dimensional.

By Theorem 4, we know that there T has an eigenvalue. Let v_1 be an eigenvector for T with eigenvalue λ. Let $U := span(v_1)$, which is an invariant subspace for T. We choose an arbitrary complement U' to U in V and obtain the splitting $U \oplus U' = V$.

Now consider the map $T' : U' \to U'$ which is defined as the composition of $T|_{U'}$ and $p' : V \to U'$ (same as how we defined T_{ij} above). By the induction hypothesis, we can choose a basis $v_2, \ldots, v_n$ of U' such that the matrix of T' is upper triangular. Let us call this matrix M'.

We claim that $v_1, v_2, \ldots, v_n$ is a basis for which the matrix of T is upper-triangular. By Corollary 8's first bullet point, the matrix looks like
$$\left(\begin{array}{c|c} \lambda & \star \\ \hline 0 & M' \end{array}\right).$$
Here $\star$ means that we do not care at all what those entries are. Since M' is upper triangular, this finishes the proof. □

Remark 16. Note that there is no reason we should be able to choose U' also be an invariant subspace. This is why we end up with an upper triangular matrix rather than a diagonal matrix. □

Upper triangular matrix representations of operators reveal important information about the operator in their diagonal entries.

Lemma 25. *Let V be a finite dimensional vector space over $\mathbb{F}$, T be an operator on V and $v_1, \dots, v_n$ be a basis of V such that $M(T, v_1, \dots, v_n)$ is upper triangular.*

1. *T is an isomorphism if and only if no diagonal entry of $M(T, v_1, \dots, v_n)$ is zero.*
2. *The set of eigenvalues of T is equal to the set of diagonal entries of the matrix $M(T, v_1, \dots, v_n)$.*

Proof. We only prove (1). Let's first prove that if no diagonal entry is 0, then T is an isomorphism. It suffices to show that it is injective. Let $v = a_1v_1 + \dots + a_nv_n$ be an arbitrary element of V that is in the nullspace of T. Because the matrix is upper triangular

$$Tv = v' + \lambda_n a_n v_n,$$

where $v' \in span(v_1, \dots, v_{n-1})$ and λ_n is the $(n, n)^{th}$ entry of the matrix. Since $Tv = 0$, $\lambda_n a_n = 0$. This implies $a_n = 0$, since we are given that $\lambda_n \neq 0$. Hence, in fact $v = a_1v_1 + \dots + a_{n-1}v_{n-1}$. Now the same argument shows that $a_{n-1} = 0$ and so on. By induction, we see that $v = 0$, as desired.

Conversely, let's assume that at least one diagonal entry is 0 and prove that T is not injective. Let the $(j, j)^{th}$ entry of $M(T, v_1, \dots, v_n)$ be zero. Then, we see that the image of $T|_{span(v_1,\dots,v_j)}$ is contained in $span(v_1, \dots, v_{j-1})$. Using the rank nullity theorem, we see that the nullspace of $T|_{span(v_1,\dots,v_j)}$ is at least one dimensional. This finishes the proof. □

Exercise 58. Deduce part (2) of this proposition from part (1). □

19. Lecture 19: Eigenspaces, Jordan blocks, generalized eigenvectors/spaces

Let us start with a definition.

Definition 27. Let V be a vector space over $\mathbb{F}$, T be an operator on V and $\lambda \in \mathbb{F}$. Then we call the subspace of vectors $v \in V$ such that $Tv = \lambda v$ the eigenspace of λ and denote it by

$$E(T, \lambda) \subset V.$$

□

Note that the eigenspace of λ is the same as the nullspace of $T - \lambda\text{Id}$. If λ is an eigenvalue, then $E(T, \lambda)$ has non-zero vectors in it.

Lemma 26. *Let V be a vector space over $\mathbb{F}$, T be an operator on V and $\lambda \in \mathbb{F}$. Assume that $\lambda_1, \dots, \lambda_n \in \mathbb{F}$ are pairwise distinct eigenvalues of T. Then*

$$E(T, \lambda_1) + \dots + E(T, \lambda_n)$$

is a direct sum.

Proof. We have to show that if $v_1, \dots, v_n \in V$ are so that $Tv_i = \lambda_i v_i$ for all $i = 1, \dots, n$ and λ_i pairwise distinct, then $v_1 + \dots + v_n = 0$ implies that all v_i are zero. Let us prove this by induction on n. For $n = 1$, it is trivial. Let us assume that it is true for $n - 1$ and prove it for n.

From $v_1 + \dots + v_n = 0$, we obtain

$$T(v_1 + \dots + v_n) = Tv_1 + \dots + Tv_n = \lambda_1v_1 + \dots + \lambda_nv_n = 0.$$

Subtracting $\lambda_1(v_1 + \ldots + v_n) = 0$ from the last equality, we end up with

$$(\lambda_2 - \lambda_1)v_2 + \ldots + (\lambda_n - \lambda_1)v_n = 0.$$

Note that

$$v_i = 0 \text{ if and only if } (\lambda_i - \lambda_1)v_i = 0$$

for all $i = 2, \ldots, n$.

Applying the induction hypothesis to vectors

$$v_2' = (\lambda_2 - \lambda_1)v_2, \ldots, v_n' = (\lambda_n - \lambda_1)v_n,$$

we obtain that $v_i' = 0$ and therefore $v_i = 0$ for all $i = 2, \ldots, n$. The initial equality also gives $v_1 = 0$, which finishes the induction. □

In the statement we do not need to assume that λ_i are eigenvalues but this does not lead to a stronger statement. Here is a corollary.

Corollary 9. *Assume that V is an n-dimensional vector space over $\mathbb{F}$ and $T \in \mathcal{L}(V)$ has n distinct eigenvalues. Then, one can choose a basis $v_1, \ldots, v_n$ such that $M(T, v_1, \ldots, v_n)$ is a diagonal matrix.*

Proof. We take v_i to be a non-zero eigenvector with eigenvalue λ_i for $i = 1, \ldots, n$. Lemma 26 shows that $v_1, \ldots, v_n$ is linearly independent, and hence a basis. The statement follows. □

We have already seen an operator on a finite dimensional vector space that is not "diagonalizable":

$$T_A \in \mathcal{L}(\mathbb{R}^2) \text{ with } A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

This operator did not even have an eigenvalue.

We now introduce another example, which holds over all fields.

Definition 28. A Jordan matrix of size n over $\mathbb{F}$ with diagonal entries $\lambda \in \mathbb{F}$ is the $n \times n$ matrix

$$J(n, \mathbb{F}, \lambda) := \begin{pmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \cdots & 1 \\ & & & \lambda \end{pmatrix},$$

with all the other entries 0. □

The linear map

$$T_{J(n,\mathbb{F},\lambda)} : \mathbb{F}^n \to \mathbb{F}^n$$

has only one eigenvalue λ by Lemma 25. A non-zero eigenvector is given by $(1, 0, \ldots, 0)$ and the eigenspace of λ contains its multiplies. It turns out that it contains nothing else.

Proposition 16. *$E(T_{J(n,\mathbb{F},\lambda)}, \lambda)$ is one dimensional.*

Proof. The image of $(a_1, \ldots, a_n)$ is

$$\lambda(a_1, \ldots, a_n) + (a_2, \ldots, a_n, 0).$$

Therefore, if $(a_1, \ldots, a_n)$ is an eigenvector $a_2 = \ldots = a_n = 0$. □

So even for complex numbers there are matrices that are not diagonalizable. Colloquially, the linear map defined by a Jordan matrix cannot be simplified further by using another basis.

On the bright side, for $\mathbb{F} = \mathbb{C}$, Jordan blocks are really all that there is as an obstacle to diagonalization. We will prove the following theorem next class.

Theorem 6. *Let V be a finite dimensional vector space over $\mathbb{C}$ and T be an operator on V. Then we can find a basis $v_1, \ldots, v_n$ of V such that*

$$M(T, v_1, \ldots, v_n)$$

is in Jordan normal form

$$\left(\begin{array}{c|c|c|c} J_1 & 0 & & \\ \hline 0 & J_2 & & \\ \hline & & \cdots & \\ \hline & & & J_k \end{array}\right).$$

Here all the non-diagonal blocks have all zero entries and

$$J_i = J(n_i, \mathbb{C}, \lambda_i),$$

for some n_i positive integer and $\lambda_i \in \mathbb{C}$ for all $i = 1, \ldots, k$.

We are going to find subspaces U_i corresponding to each of these Jordan blocks such that $U_1 \oplus \ldots \oplus U_k = V$. Here is how they will come about.

Definition 29. Let V be a vector space over an arbitrary field $\mathbb{F}$ and T be a linear operator on V. A vector $v \in V$ is called a generalized eigenvector with eigenvalue λ if

$$(T - \lambda \mathrm{Id})^N v = 0,$$

for some $N \geq 1$.

We call the set of all generalized eigenvectors of λ the generalized eigenspace of λ and denote it by $E^{gen}(T, \lambda)$. □

Notice that the generalized eigenvector equation for $N = 1$ is nothing but the eigenvector equation.

Exercise 59. Prove that $E^{gen}(T, \lambda)$ is a subspace. □

Exercise 60. Show that $E^{gen}(T_{J(n,\mathbb{F},\lambda)}, \lambda) = \mathbb{F}^n$, so in this case all vectors are generalized eigenvectors with eigenvalue λ. □

20. Lecture 20: Proof of Jordan normal form theorem I

V is a finite dimensional vector space over $\mathbb{C}$ throughout this lecture. Recall that we were after proving the Jordan normal form Theorem, which is Theorem 6 from last time. If we have an operator $T : V \to V$, we will say that T can be brought into Jordan normal form (JNF), if Theorem 6 holds for T.

Today, our goal will be to reduce the general statement to the case of nilpotent operators, that is operators S such that for some positive integer N,

$$S^N = \underbrace{S \circ \ldots \circ S}_{N\ S's} = 0.$$

In other words we will show that if nilpotent operators on V can be brought into JNF, then all operators can be brought into JNF. Nilpotent operators will be covered in the next lecture.

Let us now assume that nilpotent operators can be brought into JNF and deduce the general case. We will do this by induction on the dimension of V. The statement is trivial for V one dimensional. We assume that if V has dimension less than n, then all operators on V can be brought into JNF, and prove that $T : V \to V$ can be brought into JNF for V n dimensional. In case you already forgot we are assuming that all nilpotent operators (for all dimensions) can be brought into JNF.

Let us start with a simple observation. An operator T on V can be brought into JNF, if $T - \lambda Id$ can be brought into JNF for some $\lambda \in \mathbb{C}$. This is because if we use exactly same basis, then the matrix we obtain for T is the entrywise sum of the matrix of $T - \lambda Id$ and the matrix of λId, which is also in JNF - the latter simply adds λ to all diagonal entries.

The upshot is that we can reduce to the case where the operator has 0 as an eigenvalue by replacing T with $T - \lambda Id$ with λ an eigenvalue of T. From now on we assume that T has 0 as an eigenvalue.

Now consider the generalized eigenspace of 0:

$$E := E^{gen}(T, 0).$$

This is the set of vectors in V such that $T^N v = 0$ for some positive integer N. Clearly E is an invariant subspace of T. Also note that E is at least one dimensional.

Lemma 27. *There exists a positive integer N_0 such that if $T^N v = 0$ for some positive integer $N > N_0$ and $v \in V$, then in fact $T^{N_0} v = 0$.*

Proof. For every positive integer N,

$$E_N := null(T^N)$$

is a subspace. Morever E_{N+1} contains E_N, since $T^N v = 0$ implies $T^{N+1} v = 0$. To finish note that the dimension of E_N has to stabilize after some N_0, as otherwise we would have arbitrarily large dimensional subspaces of V, which is finite dimensional. This implies that E_N stabilizes after N_0 as desired. □

We fix an N_0 as in Lemma 27. We define the subspace

$$U := im(T^{N_0}).$$

Proposition 17. *U is a complement to E, i.e. $E \oplus U = V$.*

Proof. Because of Lemma 27, we have

$$E = null(T^{N_0}).$$

Using the rank-nullity theorem we get that the dimensions of E and U add up to the dimension of V. Therefore, all we need to show is that if $v \in E \cap U$, then $v = 0$.

If $v \in E \cap U$. then $v = T^{N_0} w$ for some $w \in V$ and $T^{N_0} v = 0$. Combining the two we get $T^{2N_0} w = 0$. This implies by Lemma 27 that $T^{N_0} w = 0$. since $2N_0 > N_0$. Hence, we have $v = 0$ as desired. □

The point is that U is not just some complement to E. It is trivially invariant under T as $T(T^{N_0}v) = T^{N_0}(Tv)$. Therefore we have for ourselves an invariant complement and we can use Corollary 8's second part.

We consider the map $T' : U \to U$ defined by restricting and "projecting". By the induction hypothesis T' can be brought into JNF, since the dimension of U is less than n. On the other hand $T'' : E \to E$ defined in the same way is nilpotent, so that can also be brought into JNF. This finishes today's work by Corollary 8's second part.

21. Lecture 21: Proof of Jordan normal form theorem II: Nilpotent operators

Recall that we reduced the proof of the Jordan normal form theorem over complex numbers to nilpotent operators. Today we finish the proof. It turns out that this part works over an arbitrary field. Let us state what we are proving for clarity.

Theorem 7. *Let V be a finite dimensional vector space over $\mathbb{F}$ and T be a nilpotent operator on V. Then we can find a basis $v_1, \dots, v_n$ of V such that*

$$M(T, v_1, \dots, v_n)$$

is in Jordan normal form

$$\left(\begin{array}{c|c|c|c} J_1 & 0 & & \\ \hline 0 & J_2 & & \\ \hline & & ... & \\ \hline & & & J_k \end{array}\right).$$

Here all the non-diagonal blocks have all zero entries and

$$J_i = J(n_i, \mathbb{F}, 0),$$

for some n_i positive integer $i = 1, \dots, k$.

Remember that T being nilpotent meant that there exists an $N > 0$ such that $T^N = 0$ and

$$J(n, \mathbb{F}, 0) = \begin{pmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & ... & 1 \\ & & & 0 \end{pmatrix},$$

with all the other entries 0.

We make a definition that will be useful in the proof. For any $v \in V$, we know that $T^N v = 0$ for some $N > 0$. Therefore there exists a unique non-negative integer N_0 such that $T^{N_0}v = 0$ but $T^{N_0 - 1}v \neq 0$. We call this N_0 the exponent of v.

We restate Theorem 7 in a way that is more convenient for the proof.

Exercise 61. Prove that Theorem 7 and Theorem 8 are equivalent to each other.
□

Theorem 8. *Let V be a finite dimensional vector space over $\mathbb{F}$ and T be a nilpotent operator on V. Then we can find a list of vectors $w_1, \dots, w_k$ in V with exponents $n_1, \dots, n_k$ such that*

- *The list of vectors*

$$T^{n_1 - 1}w_1, \dots, w_1, \dots, T^{n_k - 1}w_k, \dots, w_k$$

form a basis of V.

Proof. We use induction on the dimension of V. For V one dimensional, the statement is trivial. Let us assume that it is true for $dimV < n$ and prove it for $dimV = n$.

Claim 3. $null(T) \neq \{0\}$

To see this, take any non-zero $v \in V$. We know that $T^N v = 0$ for some $N > 0$, so we can take the smallest integer $m > 0$ such that $T^m v = 0$. This implies that $T^{m-1}v$ is non-zero and in the nullspace.

Therefore, by the rank-nullity theorem $im(T)$ has smaller dimension than n. Since $im(T)$ is an invariant subspace for T, we can apply the induction hypothesis to the map $im(T) \to im(T)$ obtained by restricting T to $im(T)$.

Let $u_1, \ldots, u_l \in im(T)$ be the list of vectors that we obtain with exponents $m_1, \ldots, m_l$. By definition,

$$T^{m_1-1}u_1, \ldots, u_1, \ldots, T^{m_l-1}u_l, \ldots, u_l$$

is a basis of $im(T)$

We define $v_i \in V$ to be any vector such that $Tv_i = u_i$ for all $i = 1, \ldots, l$. Notice that the exponent of v_i is $m_i + 1$ for all $i = 1, \ldots, l$.

Claim 4. $T^{m_1}v_1, \ldots, v_1, \ldots, T^{m_l}v_l, \ldots, v_l$ *is a linearly independent set of vectors.*

We take a linear relation

$$a_{10}v_1 + \ldots + a_{1m_1}T^{m_1}v_1 + \ldots + a_{lm_l}T^{m_l}v_l = 0.$$

We apply T to this equality and obtain

$$a_{10}u_1 + \ldots + a_{1m_1-1}T^{m_1-1}u_1 + \ldots + a_{lm_l-1}T^{m_l-1}u_l = 0.$$

This implies that all the coefficients that appear in the last equation are zero. We therefore have

$$a_{1m_1}T^{m_1}v_1 + \ldots + a_{lm_l}T^{m_l}v_l = 0,$$

which is the same as

$$a_{1m_1}T^{m_1-1}u_1 + \ldots + a_{lm_l}T^{m_l-1}u_l = 0.$$

Therefore, these coefficients are also zero, which finishes the proof of the claim.

We define the subspace

$$W = span(T^{m_1}v_1, \ldots, v_1, \ldots, T^{m_l}v_l, \ldots, v_l).$$

Claim 5. $W + null(T) = V$

Let $v \in V$. We know that Tv is a linear combination of $T^{m_1-1}u_1, \ldots, u_1, \ldots, T^{m_l-1}u_l, \ldots, u_l$. We define $w \in W$ by replacing each u_i in this expression with v_i, which satisfies $Tw = Tv$. This means that $v - w$ is in the nullspace finishing the proof of the claim.

We are almost done. Choose a complement U to $W \cap null(T)$ inside $null(T)$ and let $v_{l+1}, \ldots, v_k$ be a basis for it. It is easy to see that U is in fact a complement to W inside V. Therefore

$$T^{m_1}v_1, \ldots, v_1, \ldots, T^{m_l}v_l, \ldots, v_l, v_{l+1}, \ldots, v_k$$

form a basis for V.

Noticing that the exponents of $v_{l+1}, \ldots, v_k$ are all 1, we see that

$$v_1, \ldots, v_l, v_{l+1}, \ldots, v_k$$

is the kind of list of vectors that we were after to prove the statement. □

We have finally finished the proof of the Jordan normal form theorem for complex numbers (Theorem 6).

There is a Jordan normal form theorem for real numbers as well. I will state it below. Its proof is actually not hard given the version for complex numbers, but I will omit it.

We define

$$\tilde{J}(n, \mathbb{R}, a, b) := \left(\begin{array}{c|c|c|c|c} A(a,b) & I_2 & & & \\ \hline 0 & A(a,b) & I_2 & & \\ \hline & & \dots & & I_2 \\ \hline & & & & A(a,b) \end{array}\right)$$

where the matrix is $2n \times 2n$,

$$A(a,b) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \text{ and } I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

All the other blocks are zero.

Theorem 9. *Let V be a finite dimensional vector space over $\mathbb{R}$ and T be an operator on V. Then we can find a basis $v_1, \dots, v_n$ of V such that*

$$M(T, v_1, \dots, v_n)$$

is in real Jordan normal form

$$\left(\begin{array}{c|c|c|c} J_1 & 0 & & \\ \hline 0 & J_2 & & \\ \hline & & \dots & \\ \hline & & & J_k \end{array}\right).$$

Here all the non-diagonal blocks have all zero entries and

$$J_i = J(n_i, \mathbb{R}, \lambda) \textit{ for some } \lambda \in \mathbb{R} \textit{ or } J_i = J(n_i, \mathbb{R}, a, b) \textit{ for some } a, b \in \mathbb{R}$$

for some $n_i > 0$ for all $i = 1, \dots, k$.

22. Lecture 22: Inner product spaces, orthonormal bases

For the rest of the quarter, we are going to be interested in analyzing finite dimensional vector spaces over real numbers equipped with an inner product.

Here is the prototypical example. On the vector space $\mathbb{R}^n$, we can define the dot product, which is a map $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$(a_1, \dots, a_n) \cdot (b_1, \dots, b_n) = a_1 b_1 + \dots + a_n b_n.$$

The dot product is an inner product on $\mathbb{R}^n$, of which definition we give now.

Definition 30. Let V be a vector space over $\mathbb{R}$. Then an inner product on V is a map

$$< \cdot, \cdot >: V \times V \to \mathbb{R}$$

with the following three properties:

- (positive definiteness) For every $v \in V$, $< v, v > \geq 0$ and the equality holds if and only if $v = 0$.
- (symmetry) For every $v, w \in V$, $< v, w >=< w, v >$.
- (linearity) For $a, b \in \mathbb{R}$ and $v_1, v_2, w \in V$,

$$< av_1 + bv_2, w >= a < v_1, w > +b < v_2, w > .$$

□

Exercise 62. State and prove the linearity in the second slot. □

Exercise 63. Check that indeed the dot product is an inner product on $\mathbb{R}^n$. □

Remark 17. The definition did not assume finite dimensionality but we will make this assumption from now on. □

For the rest of the lecture, V is a finite dimensional vector space over $\mathbb{R}$ with inner product $< \cdot, \cdot >$. It is customary to define the associated norm via

$$\|v\| := \sqrt{< v, v >}.$$

If two vectors v, w satisfy $< v, w >= 0$, we say that they are orthogonal.

Exercise 64. Show that if v and w are orthogonal, then for any scalar $c \in \mathbb{R}$, cv and w are also orthogonal. □

Definition 31. A basis $v_1, \dots, v_n$ for V is called orthonormal if

- $< v_i, v_i >= 1$ for all $i = 1, \dots, n$
- $< v_i, v_j >= 0$ for all $i \neq j \in \{1, \dots, n\}$

□

It is easy to see that for $\mathbb{R}^n$ with the dot product, the standard basis is orthonormal.

We can turn any basis of V into an orthonormal basis by an algorithmic procedure called the Gramm-Schmidt process. It goes as follows. Let $v_1, \dots, v_n$ be a basis of V. We will modify the vectors in this basis one by one, starting from v_1, and getting to

$$v_1', \dots, v_k', v_{k+1}, \dots, v_n$$

after the kth step, so that $v_1', \dots, v_n'$ is an ON basis at the end.

The key point of the procedure is that we will require

$$span(v_1', \dots, v_k') = span(v_1, \dots, v_k)$$

at all steps $k = 1, \dots n$. This essentially determines what the procedure is (up to a sign ambiguity at every step). If you remember that we can achieve our goal under this requirement, you will have no trouble coming up with the Gramm-Schmidt process yourself.

So we start: all we get to do in the first step is to multiply v_1 by a scalar so that the result has norm 1. This is possible of course, we define $v_1' := \frac{v_1}{\|v_1\|}$.

Exercise 65. Show that v_1' has norm 1. □

We could have also chosen $-v_1'$ in this step, which is the ambiguity I mentioned. We will have this ambiguity at every step but I will not mention it again.

We come to the second step. We get to multiply v_2 by a scalar and also add a multiple of v_1' (equivalently of v_1) such that the result has norm 1 and it is orthogonal to v_1'. Let us first make it orthogonal, since we already know how to deal with the norm after that. Here is what we do

$$v_2'' = v_2 - < v_1', v_2 > v_1'.$$

Exercise 66. Show that v_2'' is orthogonal to v_1'. □

Notice that v_2'' is not zero because v_1, v_2 and therefore v_1', v_2 is linearly independent, so we define $v_2' := \frac{v_2''}{\|v_2''\|}$. Orthogonality still holds of course.//

I will also do the third step and leave the proof by induction to you. To modify v_3 we will add a linear combination of v_1', v_2' so that it becomes orthogonal to v_1' and v_2'. The result cannot be zero, therefore we can then divide by the norm and get what we want. What linear combination do we add? We follow the same pattern:

$$v_3'' := v_3 - < v_2', v_3 > v_2' - < v_1', v_3 > v_1'.$$

Exercise 67. Show that v_3'' is orthogonal to v_1' and v_2'. □

Exercise 68. Rigorously define the Gramm-Schmidt process inductively. □

An immediate corollary is the following.

Corollary 10. *V admits an orthonormal basis.*

Recall that given subspace $U \subset V$ we could also always find a complement subspace for it. Using the inner product, we can define a special complement called the orthogonal complement.

Definition 32. The orthogonal complement to a subspace $U \subset V$ is defined as

$$U^{\perp} := \{v \in V \mid < v, u >= 0 \text{ for all } u \in U\}.$$

□

Exercise 69. Show that $U^{\perp}$ is a subspace. □

Lemma 28. $U \oplus U^{\perp} = V$

Proof. First of all note that if $v \in U \cap U^{\perp}$, then $< v, v >= 0$, or equivalently $v = 0$. Hence it suffices to show $U + U^{\perp} = V$. This is very similar to Gramm-Schmidt process.

Choose an orthonormal basis $u_1, \ldots, u_k$ of U. Let $v \in V$ and define

$$u' := v - < u_1, v > u_1 - \ldots - < u_k, v > u_k.$$

It follows easily that $u' \in U^{\perp}$, which finishes the proof. □

Just as for any splitting we obtain linear maps $V \to U$ and $V \to U^{\perp}$. The map $V \to U$ is called the projection to U. Note that we can define the projection map $V \to U^{\perp}$ in two different ways: using the splittings $U \oplus U^{\perp} = V$ or $U^{\perp} \oplus (U^{\perp})^{\perp} = V$. It turns out that these two splittings are the same up to changing the order of the subspaces.

Lemma 29. $(U^{\perp})^{\perp} = U$

Proof. We first show that $U \subset (U^{\perp})^{\perp}$. We need to show that for $u \in U$, and every $u' \in U^{\perp}$, $< u, u' >= 0$. This is immediate from symmetry.

Note that just like $U \oplus U^{\perp} = V$, we have $U^{\perp} \oplus (U^{\perp})^{\perp} = V$. This means that the dimensions of $(U^{\perp})^{\perp}$ and U are the same which finishes the proof. □

Exercise 70. Let $U \subset \mathbb{R}^n$ be a subspace that is the span of a subset of the vectors in the standard basis. Compute the projection map to U with respect to the dot product. □

23. Lecture 23: Projection formula, self-adjoint operators, isometries

Throughout this lecture: V is a finite dimensional vector space over $\mathbb{R}$ with inner product $< \cdot, \cdot >$.

Before we start analyzing operators on V, let us point out a couple of useful things about orthonormal bases and projections.

Lemma 30. *Let $v_1, \dots v_n$ be an ON basis of V. Then for all $v \in V$, we have*

$$v =< v, v_1 > v_1 + \ldots + < v, v_n > v_n.$$

Proof. Since $v_1, \dots v_n$ is a basis, there exists $a_1, \dots, a_n \in \mathbb{R}$ such that

$$v = a_1 v_1 + \ldots + a_n v_n.$$

Now take the inner product of both sides with v_i, for $i = 1, \dots, n$. Using that $v_1, \dots v_n$ is orthonormal, we get that

$$< v, v_i >= a_i,$$

proving the lemma. □

Therefore, we know exactly how to compute which linear combination of an orthonormal basis is equal to a given vector.

Now let $U \subset V$ be a subspace. Last time we proved that we have a splitting $U \oplus U^{\perp} = V$, which defines for us the projection map

$$pr_{U \subset V} : V \to U.$$

Lemma 31 (Projection formula)**.** *Let $U \subset V$ be a subspace and $u_1, \dots, u_m$ be an ON basis for U. Then, for all $v \in V$, we have*

$$pr_{U \subset V} v =< v, u_1 > u_1 + \ldots + < v, u_m > u_m.$$

Proof. We also take an orthonormal basis of $U^{\perp}$: $u_1', \dots, u_k'$. It follows from Lemma 28 and the definition of $U^{\perp}$ that $u_1, \dots, u_m, u_1', \dots, u_k'$ is an ON basis for V. Therefore, we have

$$v = \underbrace{< v, u_1 > u_1 + \ldots + < v, u_m > u_m}_{u} + \underbrace{< v, u_1' > u_1' + \ldots + < v, u_m' > u_k'}_{u'}.$$

This finishes the proof. □

Remark 18. As you can see, it is very pleasant to work in the presence of an inner product. Similar pleasantries are used in the infinite dimensional context as well. For example when you see Fourier series you should try to think of what is going on light of the formulas we just proved. □

It turns out that V is always isomorphic to $\mathbb{R}^n$ in such a way that its inner product is matched with the dot product.

Proposition 18. *Let $v_1, \dots, v_n$ be an ON basis of V, and let $e_1, \dots, e_n$ be the standard basis of $\mathbb{R}^n$. Consider the unique linear map $T : V \to \mathbb{R}^n$ satisfying $Tv_i = e_i$ for all $i = 1, \dots, n$. Then, for all $v, w \in V$,*

$$< v, w >= Tv \cdot Tw,$$

where we used the dot product on the right hand side.

Exercise 71. Prove this proposition using Lemma 30. You just need compute both sides using the bases. □

Since we know that we can find an ON basis of V, this shows that V with its inner product is isomorphic to $\mathbb{R}^n$ with the dot product (without the inner products we already knew this). Since the ON basis is not unique, the isomorphism is also not unique. We find it useful to keep going with the abstract framework.

Let us now start our analysis of operators. We will focus on two special types of operators:

Definition 33. Let T be an operator on V.

- We call T self-adjoint if for all $v, w \in V$,
$$< Tv, w >=< v, Tw >$$
- We call T an isometry if for all $v, w \in V$,
$$< Tv, Tw >=< v, w >$$

□

Note that the definitions do not actually require the finite dimensionality assumption.

Exercise 72. Prove that for $\mathbb{R}^n$ with dot product an operator T is self-adjoint if and only if the matrix $M(T, e_1, \ldots, e_n)$ is a symmetric matrix. □

Exercise 73. Prove that an isometry of V has to be an isomorphism. □

For a couple of lectures we work on choosing ON bases of V such that self-adjoint operators and isometries have "simple" matrices with respect those bases. We will give more intuition as well. Today, we end with a preparatory lemma.

Lemma 32. *Let T be either an isometry or a self-adjoint operator on V. If a subspace $U \subset V$ is invariant under T, then so is $U^{\perp}$.*

Proof. Here is what we need to show: if for some $v \in V$, $< v, u >= 0$ for all $u \in U$, then $< Tv, u >= 0$ for all $u \in U$ as well.

First, assume that T is self-adjoint. Then, we have $< Tv, u >=< v, Tu >$. Since U is an invariant subspace, $Tu \in U$ and $< v, Tu >= 0$.

Second, assume that T is an isometry, which in particular means that it is an isomorphism. Let T^{-1} be the inverse operator. It is easy to see that U is an invariant subspace of T^{-1} as well (see exercise below). Therefore, we have $< Tv, u >=< Tv, TT^{-1}u >=< v, T^{-1}u >= 0$. □

Exercise 74. Let T be an isomorphism $V \to V$ and let $U \subset V$ be an invariant subspace of T. Then, prove that U is an invariant subspace of T^{-1} as well. This requires finite dimensionality of V but not the inner product. You might have fun trying to construct a counter-example when V is infinite dimensional. □

If we assume the real Jordan normal form theorem, we can easily deduce that any operator T on V has an invariant subspace of dimension at most 2. Using what we just proved along with Corollary 8, this shows by the induction that in either case we can make the matrix of T into a block diagonal one, where each block in the diagonal has size at most two, by choosing a not necessarily orthonormal basis. We can in fact make sure that in the self-adjoint case, we can make the blocks size 1 and in the isometry case we can make each block be either ±1 or a rotation matrix. And, we can achieve these with an orthonormal basis. We will cover these in the next two classes - we will also not assume the real Jordan normal form theorem as we did not prove it.

24. Lecture 24: Spectral theorem

Let V be a finite dimensional vector space over $\mathbb{R}$ with inner product $< \cdot, \cdot >$ throughout this lecture.

Perhaps the most intuitive way to think about self-adjoint operators is the following.

Lemma 33. *Let $v_1, \dots v_n$ be an ON basis of V. Then, $T \in \mathcal{L}(V)$ is self-adjoint if and only if $M(T, v_1, \dots v_n)$ is a symmetric matrix.*

Proof. We have $Tv_i = a_{1i}v_1 + \dots + a_{ni}v_n$ where a_{ji} is the ji^{th} entry of $M(T, v_1, \dots v_n)$. We compute

$$< Tv_k, v_l >= a_{lk} \text{ and } < v_k, Tv_l >= a_{kl},$$

which finishes the proof. □

Note that what you have already shown in Exercise 72 gives a proof of this statement as well.

Let us now move on to the main theorem about self-adjoint operators: the spectral theorem. We first need a preparatory lemma.

Lemma 34. *For every real polynomial $p(z) = a_m z^m + \dots + a_1 z + a_0$ there exists $p_1, \dots, p_s, b_1, c_1, \dots b_r, c_r \in \mathbb{R}$ such that*

$$p(z) = a_m(z - p_1) \dots (z - p_s)(z^2 + b_1 z + c_1) \dots (z^2 + b_r z + c_r),$$

and the quadratic polynomials $(z^2 + b_i z + c_i)$ have no real roots (that is $4c_i > b_i^2$).

Proof. Using the inclusion of real numbers into complex numbers as purely real complex numbers, we can think of $p(z)$ as a polynomial over complex numbers. The key point is that if $\lambda \in \mathbb{C}$ is a root of $p(z)$, then should be $\overline{\lambda}$. This can be seen by taking the complex conjugate of

$$a_m \lambda^m + \dots + a_1 \lambda + a_0 = 0$$

and using $\overline{a_i} = a_i$.

Using the fundamental theorem of algebra, we had already shown that there exists $\lambda_1, \dots, \lambda_m \in \mathbb{C}$ such that

$$a_m z^m + \dots + a_1 z + a_0 = a_m(z - \lambda_1) \dots (z - \lambda_m)$$

as polynomials over complex numbers.

If all of these λ_i's are purely real, then we are done, because the same equality then holds as polynomials over real numbers as well. Assume that λ_1 is not purely real. We know that one of the other roots have to be its complex conjugate, so let's assume that $\lambda_2 = \overline{\lambda_1}$.

Let's now consider the real polynomial

$$z^2 + b_1 z + c_1 := z^2 - 2Re(\lambda_1) + |\lambda_1|^2.$$

Over the complex numbers this polynomial is equal to $(z-\lambda_1)(z-\lambda_2)$ and it divides $p(z)$. It is immediate that $4c_1 > b_1^2$ since the imaginary part of λ_1 is non-zero (check it!).

Let us prove that $z^2 + b_1 z + c_1$ divides $p(z)$ over the real numbers as well. We use polynomial division as in Lemma 20 to get that there is a real polynomial $q(z)$ and an at most degree 1 polynomial $r(z)$ such that

$$p(z) = (z^2 + b_1 z + c_1)q(z) + r(z).$$

We now consider this equality as an equality of complex polynomials and substitute λ_1 for z. We obtain $r(\lambda_1) = 0$, but since $r(z)$ is real and has degree at most 1, and λ_1 is not real, it follows that $r(z)$ has to be the zero polynomial.

We can now apply the same argument for $q(z)$. Since the degree of $q(z)$ is less than $p(z)$ the procedure terminates at some point and we obtain the desired factorization. □

Theorem 10 (Spectral theorem)**.** *Let V be a finite dimensional vector space over $\mathbb{R}$ with inner product $< \cdot, \cdot >$ and T be a self-adjoint operator on V. Then, there is an ON basis of V which makes the matrix of T diagonal.*

Proof. First, it suffices to show that T has an eigenvalue using the inductive argument relying on Corollary 8. This is because of the important Lemma 28 we proved in the previous lecture and the simple fact that the restriction of a self-adjoint operator to an invariant subspace is self-adjoint. Make sure you understand why we automatically get that the basis we produce after the induction is ON if we use a unit norm eigenvector for our eigenvalue.

The proof that T has an eigenvalue uses the same strategy we used to prove the existence of eigenvalues over complex numbers. Let V be n-dimensional. Choose a non-zero $v \in V$ and consider the list of vectors

$$T^n v, T^{n-1} v, \ldots, Tv, v.$$

Since this list has more elements than the dimension, it has to be linearly dependent. Therefore, there exists $n \geq m \geq 1$ and real numbers $a_m, \ldots, a_1, a_0$ such that

$$a_m T^m v + a_{m-1} T^{m-1} v + \ldots + a_1 Tv + a_0 v = 0$$

and $a_m \neq 0$. This means that the operator

$$a_m T^m + a_{m-1} T^{m-1} + \ldots + a_1 Tv + a_0 \mathrm{Id}$$

is not an isomorphism.

We just showed that there exists $p_1, \ldots, p_s, b_1, c_1, \ldots b_r, c_r \in \mathbb{R}$ such that

$$p(z) = a_m (z - p_1) \ldots (z - p_s)(z^2 + b_1 z + c_1) \ldots (z^2 + b_r z + c_r),$$

and $4c_i > b_i^2$. We substitute T for z and obtain that

$$(T - p_1 Id) \circ \ldots \circ (T - p_s Id) \circ (T^2 + b_1 T + c_1 Id) \circ \ldots \circ (T^2 + b_r T + c_r Id)$$

is not an isomorphism.

If $T - p_i Id$ is not an isomorphism for some $i = 1, \ldots, s$, we are done. Let us show that $T^2 + b_i T + c_i Id$ has to be an isomorphism for all $i = 1, \ldots, r$, which finishes the proof.

Note that

$$T^2 + b_i T + c_i Id = (T + \frac{b_i}{2} Id)^2 + \delta Id,$$

for some $\delta > 0$ using $4c_i > b_i^2$. Clearly, $T + \frac{b_i}{2} Id$ is self-adjoint. Therefore, for any $v \in V$:

$$< ((T + \frac{b_i}{2} Id)^2 + \delta)v, v >=< Tv + \frac{b_i}{2} v, Tv + \frac{b_i}{2} v > + \delta < v, v > .$$

It follows using positive definiteness that if $T^2 v + b_i T v + c_i v = 0$, then $v = 0$, which shows that $T^2 + b_i T + c_i Id$ is injective and hence an isomorphism, as desired. □

25. Lecture 25: Hermitian inner products on complex vector spaces, eigenvalues of self-adjoint operators, isometries as distance preserving maps

Let's talk a little bit about complex vector spaces equipped with a Hermitian inner product before we move on to isometries.

Definition 34. Let V be a vector space over $\mathbb{C}$. Then a Hermitian inner product on V is a map

$$< \cdot, \cdot >: V \times V \to \mathbb{C}$$

with the following three properties:

- (conjugate symmetry) For every $v, w \in V$, $< v, w >= \overline{< w, v >}$. In particular, $< v, v >$ is purely real.
- (positive definiteness) For every $v \in V$, $< v, v > \geq 0$ and the equality holds if and only if $v = 0$.
- (linearity) For $a, b \in \mathbb{R}$ and $v_1, v_2, w \in V$,

$$< av_1 + bv_2, w >= a < v_1, w > +b < v_2, w > .$$

□

Exercise 75. A Hermitian inner product is not linear in the second slot, but it is conjugate-linear (anti-linear). Formulate what this should mean using the conjugate symmetry property. □

Definition 35. Let T be an operator on V. We call T self-adjoint if for all $v, w \in V$,

$$< Tv, w >=< v, Tw >$$

□

Remark 19. One can define complexifications of real vector spaces (which are complex vector spaces) and linear maps between them (which are linear maps). You can also turn an inner product on a real vector space into a Hermitian inner product on its complexification. Then, the complexification of self-adjoint operator ends up being self adjoint as well. This essentially boils down to the following: if you consider a symmetric real matrix as a complex matrix, then it is still symmetric. I avoided discussing this complexification stoty in detail but it is actually quite useful, especially if one introduces the notion of a normal operator over $\mathbb{C}$. Hopefully you will learn about them later in your life. □

We know that complex operators have eigenvalues. An important point is that for self-adjoint ones these have to be purely real.

Lemma 35. *Let V be a vector space over $\mathbb{C}$ with Hermitian inner product $< \cdot, \cdot >$ and T be a self-adjoint operator on V. Then, the eigenvalues of T are purely real.*

Proof. Let $\lambda \in \mathbb{C}$ be an eigenvalue of T with non-zero eigenvector v. We want to show $\overline{\lambda} = \lambda$.

We compute $< Tv, v >$ in two different ways. Here is the first one:

$$< Tv, v > = < \lambda v, v > = \lambda < v.v > .$$

Using self-adjoint property it is also equal to

$$< v, Tv > = < v, \lambda v > = \overline{\lambda} < v, v > .$$

Therefore

$$\lambda < v.v > = \overline{\lambda} < v, v > .$$

Using positive definiteness we finish the proof. □

Remark 20. This can also be used in the proof of Theorem 10 to obtain a real eigenvalue using the remark about complexification above but we omit the details. □

One can use the same proof with Lemma 28 to show that the orthogonal complement (define this!) of an invariant subspace of a self-adjoint operator is also invariant. Hence, we easily obtain the following.

Theorem 11 (Complex spectral theorem)**.** *Let V be a finite dimensional vector space over $\mathbb{C}$ with Hermitian inner product $< \cdot, \cdot >$ and T be a self-adjoint operator on V. Then, there is an ON basis (define this) of V which makes the matrix of T diagonal with purely real entries.*

If one understands the complexification idea sufficiently well, one can deduce the real spectral theorem directly from the complex spectral theorem.

Remark 21*.* The most general spectral theorem is for normal operators over complex vector spaces with Hermitian inner products. These also include complexifications of isometries! Hence, what we will prove next time is also a special case of that generalization. Normal operators are abstract but they are useful, as I already mentioned in a previous remark. □

Now let's go back to real numbers and start our discussion of isometries. For this part we actually just work with $\mathbb{R}^n$ and the dot product. This is actually the same level of generality with finite dimensional vector spaces equipped with an inner product by Proposition 18.

There is a notion of a distance between two points x, y in $\mathbb{R}^n$ defined by

$$d(x, y) = \|x - y\| = \sqrt{(x - y) \cdot (x - y)}.$$

Note that here $x - y$ is a vector in $\mathbb{R}^n$, so we can take its norm with respect to the dot product. Of course, this is nothing but the Pythagoras theorem distance that we are all familiar with. I will assume that you are also familiar with:

Lemma 36. *We have the triangle inequality:*

$$d(x, y) + d(y, z) \geq d(x, z),$$

with equality if and only if y lies inside the straight line segment between x and z.

Let us call a map (not assumed to be linear) $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ distance preserving if for every $x, y \in \mathbb{R}^n$:

$$d(\Phi(x), \Phi(y)) = d(x, y).$$

Here is a very cool statement.

Proposition 19. *A map $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is distance preserving and sends the origin to origin if and only if it is an isometry.*

Proof. Let's start with the main content: if Φ is distance preserving and sends origin to origin, then it has to be linear!

Let $x, z \in \mathbb{R}^n$. We will show that the midpoint y of the straight line segment between x and z is sent to the midpoint of the straight line segment between $\Phi(x)$ and $\Phi(z)$. Let us call this *the claim.*

We know that

$$d(x, y) + d(y, z) = d(x, z),$$

which implies

$$d(\Phi(x), \Phi(y)) + d(\Phi(y), \Phi(z)) = d(\Phi(x), \Phi(z)).$$

This means that $\Phi(y)$ has to lie inside the straight line segment between $\Phi(x)$ and $\Phi(z)$. Since, we also have $d(x, y) = d(y, z)$, which implies

$$d(\Phi(x), \Phi(y)) = d(\Phi(y), \Phi(z)),$$

the claim follows.

Applying the claim to 0 and $2x$, we see that

$$\Phi(2x) = 2\Phi(x),$$

since we know that 0 is sent to 0.

If we think of x, z as vectors, the midpoint y as above is given by $1/2(x + y)$. Therefore the claim implies the additivity of Φ by the previous paragraph.

We move onto the homogeneity. Applying the claim to $-x$ and x, we see that

$$\Phi(-x) = -\Phi(x),$$

again using that 0 is sent to 0. All that is left to do is to prove homogeneity for positive scalars. It suffices to prove this for $0 < c < 1$ (why?).

The proof of this final point follows immediate from the argument that gave us the claim. I leave this to you.

Hence, we know that Φ has to be a norm preserving linear map. To finish we need to show that it preserves the dot product of arbitrary two vectors. This follows from the polarization identity:

$$u \cdot v = \frac{\|u + v\| - \|u - v\|}{4}$$

which is in your homework.

The converse is trivial. □

Next class we will analyze isometries of $\mathbb{R}^n$, which are as we just saw the same as distance preserving maps up to a translation. We stress again that the latter is a notion that does not use the linear structure in its definition.

For example in three dimensions, we will see that if an isometry preserves the orientation of $\mathbb{R}^3$, then it has to be a rotation along an axis passing through the origin. Since the composition of two orientation preserving isometries is an orientation preserving isometry, this means that composition of two rotations along two different axes is a rotation along a third axis. This is not clear visually.

26. Lecture 26: Multiplicativity of determinants, orientations, isometries of $(\mathbb{R}^n, \cdot)$

In what follows $\mathbb{R}^n$ is equippped with its standard vector space structure and the dot product. We denote the standard basis by $e_1, \ldots, e_n$. Note that we will be choosing different bases for $\mathbb{R}^n$ in this lecture as well.

We will first introduce the notion of orientation preserving/reversing isomorphisms of $\mathbb{R}^n \to \mathbb{R}^n$. In fact, we start with a more general discussion. Below is an important property of determinants that we did not discuss in Lecture 13.

Theorem 12. *Let A and B be two $n \times n$ matrices over an arbitrary field. Then*

$$det(AB) = det(A)det(B).$$

We omit the proof as our focus is elsewhere. Giving the correct proof of this involves the notion of the top exterior power of a vector space, which again I hope you will see elsewhere.

We have seen in a homework that matrices of operators with respect to different bases are related by a relation of the form

$$M' = AMA^{-1}$$

for an invertible matrix A that is called the change of basis matrix.

Remark 22. What I defined as the change of basis matrix in that homework is the inverse of what most people call the change of basis matrix. I believe that my definition is more intuitive: it is the matrix that turns the old column vector representation to the new column vector representation. On the other hand the more common definition has the virtue that it is easy to say what the matrix is: you write the old column vector representations of the new basis and put them in a matrix as the columns. The more conventional definition is so that the matrix turns the new column vector representation to the old column vector representation.

I took my first class in this stuff from Michael Artin, whose book Algebra covers this topic. He says that in the first version he used my convention to define what the change of basis matrix is but then in the second version he switched to the more conventional one. As long as you know what you are doing, all is fine. □

Corollary 11. *Let T be an operator on a finite dimensional vector space V over $\mathbb{F}$. Then*

$$det(M(T, v_1, \ldots, v_n))$$

is independent of the chosen basis.

Exercise 76. Prove this using the multiplicativity of determinant. □

As a result, we can talk about the determinant of an operator: $det(T)$.

Definition 36. Let T be an operator on a finite dimensional vector space V over $\mathbb{R}$, which we assume is an isomorphism. We call T orientation preserving if $det(T) > 0$, and orientation reversing if $det(T) < 0$. □

Let us call an isometry of $\mathbb{R}^n$ positive if it is orientation preserving, and negative if it is orientation reversing. We start by discussing isometries in low dimensions.

Take $n = 1$. What are the isometries of $\mathbb{R}$? It is trivial that these are the identity map and multiplication by -1. The former is positive, whereas the latter is negative. Multiplication by -1 is the simplest example of a reflection.

How about the isometries of $\mathbb{R}^2$? Let T be one and let $Te_1 = v_1 = (\cos\theta, \sin\theta)$ for some angle θ. Then, Te_2 must be a vector that is of norm 1 and is perpendicular to v_1. Here I am using perpendicular to mean orthogonal with respect to the dot product, which matches exactly the meaning of the word as you always used. There are two such vectors $(-\sin\theta, \cos\theta)$ and $(\sin\theta, -\cos\theta)$. The matrix $M(T, e_1, e_2)$ are either

$$Rot_\theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ or } \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} = Rot_\theta \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The first option gives rise to a counter clockwise rotation by θ radians and it is positive. The second one gives rise to a reflection along the "x-axis" followed by the same rotation. This one is negative.

The rotation isometry has the same matrix with respect to any ON basis of $\mathbb{R}^2$ (show this!). It has no eigenvalues. On the other hand the negative isometry above has 1 and -1 as eigenvalues with perpendicular non-zero eigenvectors!

Here is a geometric way to see what these eigenvectors are. We assume that v_1 is in the first quadrant - other cases can be dealt with in similar manner. Take the unit vector w that bisects the angle from e_1 to v_1 in the first quadrant. It is easy to see that reflection along x-axis and then rotation by θ brings w back to itself. Now rotate w positively $\pi/2$ degrees to get w'. It is slightly harder but still easy to see that reflection along x-axis and then rotation by θ brings w' to $-w'$. Hence, with respect to the ON basis w, w' the matrix becomes

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

Summing it all, all positive isometries of $\mathbb{R}^2$ are rotations and all negative isometries of $\mathbb{R}^2$ are reflections.

Before we go to $n = 3$, let us discuss reflections in all dimensions. On $\mathbb{R}^n$ one can reflect along any $n-1$ dimensional subspace (a hyperplane). The most convenient way to describe a hyperplane is through a normal vector (or the normal line). Given a unit vector $v \in \mathbb{R}^n$, define

$$H_v := span(v)^\perp.$$

Any hyperplane is of this form as its orthogonal complement has to be one dimensional and $H_v = H_w$ if and only if $w = \pm v$.

First notice that projection $\mathbb{R}^n \to H_v$ is given by

$$u \mapsto u - (u \cdot v)v.$$

This agrees with our preconception of what this projection should be.

Using this it is easy to see that we can define the reflection $\mathbb{R}^n \to \mathbb{R}^n$ along H_v by

$$u \mapsto u - 2(u \cdot v)v.$$

Exercise 77. Check that reflection along any hyperplane is a negative isometry. Show that with respect to some ON basis of $\mathbb{R}^n$, the matrix of a reflection along a hyperplane is diagonal with all entries equal to 1 but the last one, which is equal to -1. □

Note that the composition of two reflections (along possibly different hyperplanes) is a positive isometry.

Let us now go back to isometries of $\mathbb{R}^3$. Of course we have all the reflections that we just described, but we have a lot more. Choose a one dimensional subspace (a line) $l \subset \mathbb{R}^3$. We can write any vector in $\mathbb{R}^3$ uniquely as $v + w$, where $v \in l$ and w is orthogonal to l. We define the θ radian rotation along l operator on $\mathbb{R}^3$ as

$$v + w \mapsto v + rot_\theta(w),$$

where rot_θ is defined using any ON basis on $l^\perp$.

Exercise 78. Think about what the θ radian rotation along l does and make sure it deserves the name. Check that it is a positive isometry. □

There is one more class of isometries of $\mathbb{R}^3$. These are the ones where you follow a rotation along an axis with a reflection along the plane perpendicular to the same axis. In particular, all positive isometries of $\mathbb{R}^3$ are rotations along an axis. We will deduce the classification of the isometries of $\mathbb{R}^3$ into three classes (which were...?) from the following theorem that holds in all dimensions.

Theorem 13. *Let V be a finite dimensional vector space over $\mathbb{R}$ equipped with an inner product $< \cdot, \cdot >$ and T be an isometry on V. Then we can find an ON basis $v_1, \ldots, v_n$ of V such that*

$$M(T, v_1, \ldots, v_n)$$

is of the form

$$\left(\begin{array}{c|c|c|c} A & 0 & & \\ \hline 0 & Rot_{\theta_1} & & \\ \hline & & \cdots & \\ \hline & & & Rot_{\theta_k} \end{array}\right).$$

Here all the non-diagonal blocks have all zero entries, A is a diagonal matrix with all entries equal to 1 or -1, and Rot_{θ_k} are 2×2 rotation matrices for all $i = 1, \ldots, k$.

Note that even though I preferred stating everything for the dot product until now, I switched to the abstract framework for the theorem. Here is the reason: when you take a subspace W of $\mathbb{R}^n$ and, using the dot product, equip W with an inner product, it is not canonically an $\mathbb{R}^k$ with the dot product. It becomes one if you choose an ON basis, but that is a choice and it does not need be made.

Exercise 79. Show that Theorem 13 implies what we claimed about the isometries of $\mathbb{R}^3$. Check also that it reproduces what we proved by hand for $n = 1, 2$. □

The final covers up to here. Next lecture I will finish the proof. This will be the last lecture of the course.

27. Lecture 27: Proof of the normal form theorem for isometries, polar decomposition, SVD

Let us finish the proof from last time.

Proof of Theorem 13. All we need to prove is that T either has ± 1 as an eigenvalue or has a two dimensional invariant subspace U. In the latter option, we already know that $T|_U : U \to U$ has to be either a rotation or a reflection in an arbitrary ON basis of U (using column vector representations via Proposition 18 and our analysis of the isometries of $(\mathbb{R}^2, \cdot)$) We then finish using Lemma 28 and Corollary 8.

We use the argument that we used for the proof of real spectral theorem. We find that there exists $p_1, \ldots, p_s, b_1, c_1, \ldots b_r, c_r \in \mathbb{R}$ such that $4c_i > b_i^2$ and

$$(T - p_1 Id) \circ \ldots \circ (T - p_s Id) \circ (T^2 + b_1 T + c_1 Id) \circ \ldots \circ (T^2 + b_r T + c_r Id)$$

is not an isomorphism.

If $T - p_i Id$ is not an isomorphism for some $i = 1, \ldots, s$, then it follows that T has an eigenvalue. Because T is an isometry this eigenvalue has to be 1 or -1, so we are done.

If $T^2 + b_i T + c_i Id$ is not an isomorphism for some $i = 1, \ldots, r$, then we obtain that for some $v \in V$,

$$T^2 v + b_i T v + c_i v = 0.$$

This implies that $span(v, Tv)$ is an invariant subspace, which ends the proof. □

I want to now briefly explain polar decomposition for operators on real vectors spaces with inner products.

Let V be a finite dimensional vector space over $\mathbb{R}$ equipped with an inner product $< \cdot, \cdot >$ and T be an operator on V.

Lemma 37. *There exists a unique operator T^* on V, which satisfies*

$$< Tv, w > = < v, T^* w >,$$

for all $v, w \in V$

Proof. There are many ways to prove this but I will tell you the one that is most concrete (not the best proof). We choose an orthonormal basis $v_1, \ldots, v_n$ of V. Let M be the matrix of T with respect to $v_1, \ldots, v_n$. One then easily sees that if there is an operator T^* as in the statement then its matrix with respect to $v_1, \ldots, v_n$ has to be the transpose of M. Defining the unique operator T^* with matrix M^t, we also prove the existence. □

Definition 37. Let V be a finite dimensional vector space over $\mathbb{R}$ equipped with an inner product $< \cdot, \cdot >$ and T be an operator on V. The unique operator T^* on V, which satisfies

$$< Tv, w > = < v, T^* w >,$$

for all $v, w \in V$, is called the adjoint of T. □

Remark 23. We called T self-adjoint above if $T = T^*$. We had used the little computation above on the relationship between the matrices of T and T^* when we showed that the matrix of a self-adjoint operator with respect to an ON basis is symmetric. □

Exercise 80. Prove that T is an isometry if and only if $T \circ T^* = Id$. □

Exercise 81. Prove that if T and S are operators on V as above, then $(T \circ S)^* = S^* \circ T^*$. □

Let us make one more definition.

Definition 38. Let V be a finite dimensional vector space over $\mathbb{R}$ equipped with an inner product $< \cdot, \cdot >$ and T be an operator on V. We call T positive semi-definite, if $< Tv, v > \geq 0$ for all $v \in V$. We call it positive definite, if in addition equality implies $v = 0$. □

Here is the statement of polar decomposition. I will not give a full proof.

Theorem 14. *Let V be a finite dimensional vector space over $\mathbb{R}$ equipped with an inner product $< \cdot, \cdot >$ and T be an operator on V. Then, there exists a unique positive semi-definite self-adjoint operator P and an isometry U such that*

$$T = P \circ U.$$

If T is an isomorphism, then P is positive-definite and U too is unique.

One easily sees that $T = P \circ U$ implies:

$$T \circ T^* = P \circ U \circ U^* \circ P^* = P^2.$$

$T \circ T^*$ is a positive semi-definite self-adjoint operator which is positive definite if and only if T is an isomorphism (check this!) One needs to show that there exists a unique positive definite self-adjoint P that satisfies this equality.

By the spectral theorem we can find an ON basis $v_1, \dots, v_n$ of V such that

$$(T \circ T^*)v_i = \lambda_i v_i,$$

with $\lambda_i \geq 0$ (inequalities are strict if T is an isomorphism). It immediately follows that there exists a square root P, since we can define it by

$$Pv_i = \sqrt{\lambda_i} v_i,$$

for all $i = 1, \dots, n$. Uniqueness is easy when λ_i are all distinct and the general case follows by a continuity argument.

When T is an isomorphism, we need to choose

$$U = P^{-1} \circ T.$$

Using that P^{-1} is also self-adjoint, one shows that U is an isometry. When T is not an isomorphism more analysis is required to choose U.

Note that a positive semi-definite self-adjoint operator P is an operator that stretches or contracts in its orthonormal eigenbasis directions by the spectral theorem. Hence, polar decomposition says that every operator is an isometry composed with such an operator.

Let's think about the case $\mathbb{R}^n$ with the dot product and restate the combination above of polar decomposition and spectral theorem in terms of matrices. Any square matrix with real entries can be written as

$$ODO^t U,$$

where D is a diagonal matrix with non-negative entries, $UU^t = I$ and $OO^t = I$ (matrices satisfying this equality are called orthogonal matrices). Combining $O^t U$ into a single orthogonal matrix, we obtain what is called the singular value decomposition of a real square matrix. SVD generalizes to non-square matrices as well, which requires more care.

28. Some problems

28.1. **Constructing the rational numbers.**

i) We can construct $\mathbb{Q}$ as a set using an equivalence relation on $\mathbb{Z} \times (\mathbb{Z} - \{0\})$. Namely, let $(a, b) \sim (c, d)$ if and only if $ad = bc$. Then define $\mathbb{Q} := \mathbb{Z} \times (\mathbb{Z} - \{0\})/\sim$.

It is customary to denote the equivalence class of an element (a, b) in $\mathbb{Z} \times (\mathbb{Z} - \{0\})/\sim$ by $\frac{a}{b}$. Make sure you understand why. Where does this equivalence relation come from?

ii) Define the field operations $+$ and $\cdot$ on $\mathbb{Q}$ as they should be using the intuition you have. Check that these definitions indeed make $\mathbb{Q}$ a field.

28.2. **Finite fields.** Let $\mathbb{F}$ be a finite field.

i) Prove that there must exist a positive integer n such that $\overbrace{1 + 1 + \cdots + 1}^{n \text{ times}} = 0$. Define $c_{\mathbb{F}}$ to be the smallest such n.

ii) Prove that $c_{\mathbb{F}}$ must divide $|\mathbb{F}|$, the number of elements in k. Feel free to look up Lagrange's theorem and use it (without proving) for this problem. Everything else, including the fact that the theorem applies, needs to be proven of course.

iii) Now assume that $|\mathbb{F}| = p$, where p is a prime number. Prove that $c_{\mathbb{F}} = p$.

iv) Prove that there exists exactly one field with p elements.

(Hint: To do this, first show that at least one field with p elements exists. We have already constructed two operations on a set of p elements in class. You only need to check that the field axioms are satisfied. Secondly, you will need to show that this is the only field with p elements. For this part, start by representing the elements of the field as $1 + 1 + \cdots + 1$)

28.3. **Subfields.**

i) Let k be a field and $F \subset k$ be a subset. Assume that:
- $0, 1 \in F$.
- If $a \in F$, then $-a \in F$.
- If $a \in F - \{0\}$, then $a^{-1} \in F$.
- If $a, b \in F$, then $a + b \in F$.
- If $a, b \in F$, then $a \cdot b \in F$.

Prove that if one endows F with the same operations as k, then F is itself a field. We call F a *subfield* of k.

ii) Inside $\mathbb{C}$, we have the set of purely real numbers:

$$A := \{a + i \cdot 0 \mid a \in \mathbb{R}\} \subset \mathbb{C}.$$

We also have the set of purely imaginary numbers:

$$B := \{0 + i \cdot b \mid b \in \mathbb{R}\} \subset \mathbb{C}.$$

Which of A and B are subfields of $\mathbb{C}$?

28.4. **Solving an equation over different fields.** For which of the following fields does the equation $x^2 + 2x + 2 = 0$ have a solution? $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{F}_p$ with p prime (Note: By x^2 we mean $x \cdot x$, and by 2 we mean $1 + 1$ where 1 is the multiplicative identity in whichever field we are considering.)

You can use the following fact without proof. For every prime p, there exists an element g in $\mathbb{F}_p$ such that the smallest positive integer k so that $g^k = 1$ is $k = p - 1$.

28.5. **A weird vector space.** We can make $\mathbb{R}$ into a vector space over $\mathbb{Q}$. The vector addition is the usual addition in $\mathbb{R}$, and scalar multiplication is given by multiplying a real number with a rational number in the usual way.

i) Prove that this makes $\mathbb{R}$ a vector space over $\mathbb{Q}$.
ii) Prove that $\mathbb{Q} \subset \mathbb{R}$ is a subspace of $\mathbb{R}$ as a vector space over $\mathbb{Q}$.
iii) Find a subspace V such that $\mathbb{Q} \subsetneq V \subsetneq \mathbb{R}$. You can use without proof that $\sqrt{2}$, $\sqrt{3}$, and $\sqrt{6}$ are not rational numbers.
iv) Can you find infinitely many subspaces $V_1, V_2, \ldots$ of $\mathbb{R}$ such that for every positive integer i, $V_i \subsetneq V_{i+1}$? You do not need to rigorously prove your statement but indicate why you think your answer is true. If you think the answer is yes, you should at least provide a candidate example. Feel free to read about transcendental numbers, and use that there exists transcendental numbers for this part.

28.6. **Subspaces of $\mathbb{F}_2^5$.**

i) Let $V \subset \mathbb{F}_2^5$ be a subspace. What are the possible values of $|V|$?
ii) Find subspaces V_1, V_2, V_3 of $\mathbb{F}_2^5$ such that $V_1 \cap V_2 = V_2 \cap V_3 = V_1 \cap V_3 = \{0\}$ but $V_1 + V_2 + V_3$ is not a direct sum.

28.7. **When do we consider two vector spaces to be the same?** Let $\mathbb{F}$ be a field and let V, W be two vector spaces over $\mathbb{F}$. An *isomorphism* from V to W is a bijective (meaning one-to-one and onto) map $\phi : V \to W$ such that $\phi(v + v') = \phi(v) + \phi(v')$ for all $v, v' \in V$ and $\phi(c \cdot v) = c \cdot \phi(v)$ for all $c \in \mathbb{F}$ and all $v \in V$.

i) Prove that if there is an isomorphism $V \to W$ then there is also an isomorphism $W \to V$. In this case we say that V and W are isomorphic.

   In the following two parts, $\mathbb{F}$ means the vector space $\mathbb{F}^1$.
ii) Prove that the subspace $\{(t, t, t) \mid t \in \mathbb{F}\} \subset \mathbb{F}^3$ and $\mathbb{F}$ are isomorphic.
iii) Prove that $\mathbb{F}$ and $\mathbb{F}^2$ are not isomorphic.
iv) Find subspaces $U, U', V, V' \subset \mathbb{R}^3$ such that no two of them are isomorphic as vector spaces, but $U + U'$ and $V + V'$ are isomorphic. You can use your intuition about lines and planes in three dimensional space as long as you understand their connection to our definitions in this problem. Example of the kind of fact you can use: if you have a point p on a line, then the line contains other points which are not equal to p.

28.8. **Polynomial and functional vector spaces.** Let $\mathbb{F}$ be a field and $P(\mathbb{F})$ be the vector space of polynomials as explained in pg. 30-31 of your book. For a set S, we defined in class the vector space $\mathbb{F}^S$ of maps $S \to \mathbb{F}$, also described in pg. 14 of your book.

i) Let $\mathbb{Z}$ be the set of integers. Prove that $\mathbb{F}^{\mathbb{Z}}$ is not finite dimensional.
ii) Is $P(\mathbb{F})$ finite dimensional? Give four subspaces of $P(\mathbb{F})$ which are not isomorphic as vector spaces (pairwise). Can you find finitely many finite dimensional subspaces of $P(\mathbb{F})$ whose sum equals $P(\mathbb{F})$?
iii) If $\mathbb{F}^S$ is finite dimensional, what can you say about S? Prove your result.

iv) Let $\mathbb{N}$ be the set of nonnegative integers. Do you think $\mathbb{F}^{\mathbb{N}}$ is isomorphic to $P(\mathbb{F})$? Think about it and make an educated guess. You don't need to justify your guess. (Worth 5% of the homework.)

28.9. **Matrix multiplication.** Let A be an $m \times n$ matrix and B be a $k \times m$ matrix. Consider the linear maps $T_A : \mathbb{F}^n \to \mathbb{F}^m$ and $T_B : \mathbb{F}^m \to \mathbb{F}^k$ as defined in Lecture 11. Prove that

$$T_B \circ T_A = T_{BA}$$

as linear maps $\mathbb{F}^n \to \mathbb{F}^k$, where we BA is the matrix multiplication of B and A.

28.10. **Finite fields revisited.** Let $\mathbb{F}$ be a field with finitely many elements. Recall the definition of $c_{\mathbb{F}}$ from your first problem set before you proceed.

i) Prove that $c_{\mathbb{F}}$ must be a prime number.
ii) Let $c_{\mathbb{F}} = p$. Show that $\mathbb{F}$ can be equipped with a scalar multiplication and vector addition which makes it a vector space over $\mathbb{F}_p$. (Hint: You might try to find $\mathbb{F}_p$ as a subfield inside $\mathbb{F}$ in order to define the vector space structure, similar to how $\mathbb{R}$ was a vector space over $\mathbb{Q}$ in PSet 2.)
iii) Show that $|\mathbb{F}| = p^n$ for some positive integer n.
iv) Construct a field with 4 elements.

Remarks (You do not have to prove these statements, they are just remarks)

- There is exactly one field with p^n elements for each prime p and positive integer n.
- The fields where $\overbrace{1 + 1 + \cdots + 1}^{m \text{ times}} = 0$ for some $m > 1$ are called finite characteristic fields. They do not have to have finitely many elements. One can again show that the smallest such m has to be a prime p. Therefore, if the field is not finite, we have an infinite dimensional vector space over $\mathbb{F}_p$.
- If you have free time, you might want to think or read about fields with p^n elements or infinite fields with finite characteristic.

28.11. **Matrix represenations of linear maps.** Let V be an n-dimensional vector space over $\mathbb{F}$. Think of the elements of $\mathbb{F}^n$ as column vectors. Choosing a basis $v_1, \dots, v_n$ of V determines a way of representing vectors of V as column vectors as well. Namely, let $T : V \to \mathbb{F}^n$ be the unique linear map such that $Tv_i = e_i$ (the i-th standard basis vector of $\mathbb{F}^n$). The representation of v as a column vector is given by Tv. Let $v'_1, \dots, v'_n$ be another basis for V. This gives rise to another representation of the vectors in V as column vectors, now via

$$\begin{aligned} T' : V &\to k^n \\ v'_i &\mapsto e_i. \end{aligned}$$

- Prove that there exists an $n \times n$ matrix A such that for every $v \in V$, the two column vector representations $Tv = (a_1, \dots, a_n)^\intercal$ and $T'v = (a'_1, \dots, a'_n)^\intercal$ satisfy

$$A \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix}.$$

You might want to do this by considering a diagram of linear maps of the form

$$\begin{tikzcd}
 & V \arrow[dl, "T"'] \arrow[dr, "T'"] & \\
k^n \arrow[rr] & & k^n.
\end{tikzcd}$$

I call the matrix A is the change of basis matrix. Other people have other conventions.

- Let S be an operator on V. Express the relationship between
$$M(S, v_1, \dots, v_n) \text{ and } M(S, v'_1, \dots, v'_n)$$
using the change of basis matrix.

Here is how I think about the second part of this problem. I am spelling it out in case it helps anyone.

Let $T : V \to \mathbb{F}^n$ as in the problem. An alternative way to think about $M(S, v_1, \dots, v_n)$ is the following. This is similar to my suggestion about how to think of A from the first part. Consider

$$\begin{tikzcd}
V \arrow[r, "S"] \arrow[d, "T"] & V \arrow[d, "T"] \\
\mathbb{F}^n & \mathbb{F}^n
\end{tikzcd}$$

Since all three arrows are isomorphisms, there is unique bottom horizontal arrow that makes this diagram commutative. The matrix of that unique linear map with respect to the standard basis is precisely $M(S, v_1, \dots, v_n)$ (check please!).

Of course, we can consider the same diagram for the primed versions. We can put everything together in one diagram:

$$\begin{tikzcd}
 & V \arrow[rr] \arrow[d] \arrow[ddl] & & V \arrow[d] \arrow[ddl] \\
 & \mathbb{F}^n \arrow[rr] \arrow[dl] & & \mathbb{F}^n \arrow[dl] \\
\mathbb{F}^n \arrow[rr] & & \mathbb{F}^n &
\end{tikzcd}$$

The vertical square face is the square diagram we had above. The slanted square face is the same for T' instead of T. The triangular faces are the diagram that I gave you in the suggestion for the first part of the problem. The bottom face therefore is the commutative diagram:

$$\begin{tikzcd}
\mathbb{F}^n \arrow[r, "M"] \arrow[d, "A"] & \mathbb{F}^n \arrow[d, "A"] \\
\mathbb{F}^n \arrow[r, "M'"] & \mathbb{F}^n
\end{tikzcd}$$

Here the labels of the arrows denote the matrices that give the corresponding linear maps and I defined

$$M := M(S, v_1, \dots, v_n) \text{ and } M' := M(S, v_1', \dots, v_n').$$

The commutative diagram says $M'A = AM$ or equivalently

$$M' = AMA^{-1}.$$

The point here is that we had

$$\begin{tikzcd}
& V \arrow[rr] \arrow[d] \arrow[ddl] & & V \arrow[d] \arrow[ddl] \\
& \mathbb{F}^n & & \mathbb{F}^n \\
\mathbb{F}^n & & \mathbb{F}^n &
\end{tikzcd}$$

where all the maps were isomorphisms. Therefore we could complete this diagram to the one above by adding isomorphisms for the edges of the bottom square face so that the slanted and horizontal square faces and the triangular faces become commutative. We then get for free (check please!) that the bottom face is also a commutative diagram.