Lecture notes - Math 110 Lec 002, Summer 2016

BW

The reference [LADR] stands for Axler's *Linear Algebra Done Right*, 3rd edition.

Contents

Sets and fields - 6/20

Set notation

Definition 1. A **set** is a collection of distinguishable objects, called **elements**.

Actually, this is not the correct definition. There are a number of paradoxes that arise when you use the word "set" too carelessly. However, as long as you avoid self-referential constructions like "the set of all sets", or even "the set of all real vector spaces", there should be no problem.

Notation: We write

$$x \in M$$

to say that x is an element of the set M, and $x \notin M$ to say that it is not. We write

$$N \subseteq M \text{ or equivalently } N \subset M$$

to say that a set N is a **subset** of a set M; that means that every element of N is also an element of M. Not all authors agree on exactly what the symbol "$\subset$" means, but "$\subseteq$" is always read the same way.
Sets can be defined either by listing their elements inside brackets $\{ \}$, or by specifying properties that define the set with a colon $:$ or bar $|$.

Example 1.

$$\Big\{x : \ x \in \{1,2,3,4,5,6\}, \ x \,\text{is even}\Big\} = \{2,4,6\}.$$

Here are some common constructions with two subsets M and N of a set A.

Name	Symbol	Definition
Complement	M^c	$\{x \in A : \ x \notin M\}$
Union	$M \cup N$	$\{x \in A : \ x \in M \,\text{or}\, x \in N\}$
Intersection	$M \cap N$	$\{x \in A : \ x \in M \,\text{and}\, x \in N\}$
Difference	$M \backslash N$	$\{x \in A : \ x \in M \,\text{and}\, x \notin N\} = M \cap N^c$
Cartesian product	$M \times N$	$\{(x,y) : \ x \in M, \ y \in N\}$.

Elements of the product $M \times N$ are pairs, or lists of two elements. The order matters: $(1,2)$ and $(2,1)$ are different elements of $\mathbb{N} \times \mathbb{N}$!

Definition 2. Let M and N be sets. A **function** $f : M \to N$ associates an element $f(x) \in N$ to every element $x \in M$.

Three properties of a function $f : M \to N$ are worth mentioning:
(i) f is **injective**, or **one-to-one**, if for any $x, y \in M$, $f(x) \neq f(y)$ unless $x = y$.
(ii) f is **surjective**, or **onto**, if for any $z \in N$, there is at least one element $x \in M$ such that $f(x) = z$.
(iii) f is **bijective** if it is both injective and surjective.

Fields

Fields are the number systems we will use as coefficients throughout the course. There are several axioms that our number systems have to obey. Most of these axioms are very natural, and are common to all reasonable number systems - the axioms you should pay particular attention to are (vi) and (vii) about existence of inverses. Don't worry about memorizing these axioms.

Definition 3. A **field** $\mathbb{F} = (\mathbb{F}, +, \cdot, 0, 1)$ is a set $\mathbb{F}$, together with two distinct elements $0, 1 \in \mathbb{F}$ and functions

$$+ : \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}, \quad \cdot : \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}$$

which we write $x + y$ instead of $+(x, y)$ and xy instead of $\cdot(x, y)$, such that

(i) Addition and multiplication are commutative:

$$x + y = y + x \text{ and } xy = yx \text{ for all } x, y \in \mathbb{F}.$$

(ii) Addition and multiplication are associative:

$$x + (y + z) = (x + y) + z \text{ and } x(yz) = (xy)z \text{ for all } x, y, z \in \mathbb{F}.$$

(iii) 0 is the additive identity: $x + 0 = x$ for all $x \in \mathbb{F}$.
(iv) 1 is the multiplicative identity: $1 \cdot x = x$ for all $x \in \mathbb{F}$.
(v) The distributive law holds:

$$x \cdot (y + z) = xy + xz \text{ for all } x, y, z \in \mathbb{F}.$$

(vi) For any $x \in \mathbb{F}$, there is an additive inverse $-x \in F$ such that $(-x) + x = 0$.
(vii) For any nonzero $x \in \mathbb{F} \backslash \{0\}$, there is a multiplicative inverse x^{-1} such that $x^{-1} \cdot x = 1$.

The notation suggests that identities and inverses are unique. This is true. If $0'$ is another additive identity, then

$$0 = 0 + 0' = 0';$$

similarly, if $1'$ is another multiplicative identity, then

$$1 = 1 \cdot 1' = 1'.$$

If x has two additive inverses $-x$ and $(-x)'$, then

$$-x = -x + 0 = -x + (x + (-x)') = (-x + x) + (-x)' = 0 + (-x)' = (-x)';$$

similarly, multiplicative inverses are also unique.

Example 2. The rational numbers $\mathbb{Q}$ form a field with the usual addition and multiplication.

The real numbers $\mathbb{R}$ contain $\mathbb{Q}$ and many more numbers that are not in $\mathbb{Q}$. They also form a field with the usual addition and multiplication.

The integers $\mathbb{Z}$ are not a field, because elements other than ± 1 do not have multiplicative inverses in $\mathbb{Z}$.

Example 3. Complex numbers $\mathbb{C}$ are polynomials with real coefficients in the variable i, but with the understanding that $i^2 = -1$. The operations on $\mathbb{C}$ are the usual addition and multiplication of polynomials.
For example,

$$(1+i)^3 = 1 + 3i + 3i^2 + i^3 = 1 + 3i - 3 - i = -2 + 2i$$

and

$$(5+i) \cdot (4+3i) = 20 + (4+15)i + 3i^2 = 17 + 19i.$$

$\mathbb{C}$ contains $\mathbb{R}$: any real number $x \in \mathbb{R}$ is interpreted as the complex number $x + 0i \in \mathbb{C}$. $\mathbb{C}$ is a field: most of the axioms should be familiar from working with real polynomials, and the condition that remains to be checked is that every nonzero element is invertible. Let $a + bi \in \mathbb{C}\backslash\{0\}$; then either a or b is nonzero, so $a^2 + b^2 > 0$. Then we can multiply

$$(a+bi) \cdot \Big(\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2} i \Big) = \frac{(a+bi)(a-bi)}{a^2+b^2} = \frac{a^2+b^2}{a^2+b^2} = 1.$$

The complex numbers come with another important structure. **Complex conjugation** is defined by

$$\overline{a+bi} := a - bi, \quad a, b \in \mathbb{R}.$$

Proposition 1. *Let $w, z \in \mathbb{C}$ be complex numbers. Then:*
(i) $\overline{w+z} = \overline{w} + \overline{z}$;
(ii) $\overline{wz} = \overline{w} \cdot \overline{z}$;
(iii) $\overline{z} = z$ if and only if z is real.
(iv) $z \cdot \overline{z}$ is always real and nonnegative.

Proof. Write $w = a + bi$ and $z = c + di$ with $a, b, c, d \in \mathbb{R}$. Then:

(i) $\overline{w + z} = \overline{(a + c) + (b + d)i} = (a + c) - (b + d)i = (a - bi) + (c - di) = \overline{w} + \overline{z}$;

(ii) $\overline{wz} = \overline{(ac - bd) + (ad + bc)i} = (ac - bd) - (ad + bc)i = (a - bi)(c - di) = \overline{w} \cdot \overline{z}$;

(iii) $z - \overline{z} = (c + di) - (c - di) = 2di$, which is 0 if and only if $d = 0$; and that is true if and only if $z = c$ is real.

(iv) $z \cdot \overline{z} = (c + di) \cdot (c - di) = (c^2 + d^2) + (dc - cd)i = c^2 + d^2$ is real and nonnegative. □

The fourth property makes the formula for inverting a complex number more clear. For example,

$$\frac{1}{3 + 4i} = \frac{3 - 4i}{(3 + 4i)(3 - 4i)} = \frac{3 - 4i}{25}.$$

There are many other examples of fields that are used in math. For example, there are fields where the set $\mathbb{F}$ is finite. The smallest possible example of this is when $\mathbb{F}$ contains nothing other than 0 and 1, and addition and multiplication are defined by

+	0	1
0	0	1
1	1	0

·	0	1
0	0	0
1	0	1

Incredibly, most of the theorems in linear algebra (at least until around chapter 5 of our book) do not care whether $\mathbb{F}$ represents $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ or any of the other possible fields we could come up with. In most direct applications, it's enough to study linear algebra over $\mathbb{R}$ or $\mathbb{C}$, but applying linear algebra to $\mathbb{Q}$ and to finite fields is a very useful technique in areas like number theory, abstract algebra and cryptography.

Vector spaces - 6/21

Vector spaces

Definition 4 (LADR 1.19)**.** Let $\mathbb{F}$ be a field. An **$\mathbb{F}$-vector space** is a set V, together with operations

$$+ : V \times V \longrightarrow V \quad \text{and} \quad \cdot : \mathbb{F} \times V \longrightarrow V$$

called **vector addition** resp. **scalar multiplication**, such that:

(i) Addition is commutative: $v + w = w + v$ for all $v, w \in V$.

(ii) Addition is associative: $(v + w) + x = v + (w + x)$ for all $v, w, x \in V$.

(iii) There is an additive identity ("zero element") $0 \in V$ such that $v + 0 = v$ for all $v \in V$.

(iv) For any $v \in V$, there is an additive inverse $-v \in V$ such that $v + (-v) = 0$.

(v) Scalar multiplication is associative: $(\lambda\mu) \cdot v = \lambda \cdot (\mu \cdot v)$ for all $\lambda, \mu \in \mathbb{F}$ and $v \in V$.

(vi) The distributive laws hold:

$$\lambda \cdot (v + w) = \lambda v + \lambda w \quad \text{and} \quad (\lambda + \mu) \cdot v = \lambda v + \mu v$$

for all $\lambda, \mu \in \mathbb{F}$ and $v, w \in V$.

(vii) $1 \cdot v = v$ for every $v \in V$.

Again, the additive identity and additive inverses are unique. This is the same argument as uniqueness for a field. See LADR 1.25 and 1.26 for details.

Condition (vii) may look unimpressive but it must not be left out. Among other things, it makes sure scalar multiplication doesn't always return 0.

Example 4. The basic example of a vector space you should keep in mind is the set

$$\mathbb{F}^n = \Big\{(x_1, ..., x_n) : \ x_1, ..., x_n \in \mathbb{F}\Big\}$$

of **lists**, or **tuples**, of n elements from $\mathbb{F}$. We add lists and multiply by scalars componentwise:

$$(x_1, ..., x_n) + (y_1, ..., y_n) := (x_1 + y_1, ..., x_n + y_n), \quad \lambda \cdot (x_1, ..., x_n) := (\lambda x_1, ..., \lambda x_n).$$

In particular, the field $\mathbb{F}$ itself is an $\mathbb{F}$-vector space: it's just $\mathbb{F}^1$.
The zero element in $\mathbb{F}^n$ has the zero of $\mathbb{F}$ in each component: it's

$$0 = (0, ..., 0).$$

Try not to get the vector $0 \in \mathbb{F}^n$ and the number $0 \in \mathbb{F}$ confused.

Example 5. Let S be any set. Then the set of functions

$$\mathbb{F}^S := \{f : S \to \mathbb{F}\}$$

is a vector space: we add functions by

$$(f + g)(x) := f(x) + g(x), \quad x \in S, \ f, g \in \mathbb{F}^S$$

and multiply by scalars by

$$(\lambda f)(x) := \lambda \cdot f(x), \quad x \in S, \ \lambda \in \mathbb{F}.$$

The zero element of $\mathbb{F}^S$ is the constant 0 function:

$$0 : S \longrightarrow \mathbb{F}, \quad 0(x) = 0 \,\forall\, x.$$

If $S = \{1, ..., n\}$, then we can identify $\mathbb{F}^S$ with $\mathbb{F}^n$ by

$$f \in \mathbb{F}^S \ \Leftrightarrow \ (f(1), ..., f(n)) \in \mathbb{F}^n.$$

When $S = \mathbb{N} = \{1, 2, 3, 4, ...\}$, we can elements of $\mathbb{F}^S$ as sequences by

$$f \in \mathbb{F}^S \ \Leftrightarrow \ \text{the sequence } (f(1), f(2), f(3), ...).$$

When we interpret $\mathbb{F}^\mathbb{N}$ as a space of sequences, we will denote it $\mathbb{F}^\infty$.

Example 6. $\mathbb{C}$ is a real vector space. The addition is the usual addition; and scalar multiplication is the usual multiplication but only allowing reals as scalars.
Similarly, $\mathbb{C}$ and $\mathbb{R}$ can be interpreted as $\mathbb{Q}$-vector spaces.

The following observations are not quite as obvious as they might appear.

Proposition 2 (LADR 1.29-1.31). *Let V be an $\mathbb{F}$-vector space. Then:*
(i) $0 \cdot v = 0$ *for every* $v \in V$*;*
(ii) $\lambda \cdot 0 = 0$ *for every* $\lambda \in \mathbb{F}$*;*
(iii) $(-1) \cdot v = -v$ *for every* $v \in V$*.*

Proof. (i) Since

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

we can subtract $0 \cdot v$ from both sides to get $0 \cdot v = 0$.
(ii) Since

$$\lambda \cdot 0 = \lambda \cdot (0+0) = \lambda \cdot 0 + \lambda \cdot 0,$$

we can subtract $\lambda \cdot 0$ from both sides to get $\lambda \cdot 0 = 0$.
(iii) This is because

$$(-1) \cdot v + v = (-1) \cdot v + 1 \cdot v = (-1+1) \cdot v = 0 \cdot v = 0$$

by (i), so $(-1) \cdot v$ is the additive inverse of v. □

Subspaces

Definition 5 (LADR 1.34). Let V be an $\mathbb{F}$-vector space. A **subspace** of V is a subset $U \subseteq V$ such that:
(i) $0 \in U$;
(ii) $\lambda v + w \in U$ for all $v, w \in U$ and $\lambda \in \mathbb{F}$.

In particular, subspaces are vector spaces in their own right, with the same addition and scalar multiplication.

Example 7. Consider the space $V = \mathbb{R}^{(0,1)}$ of all functions $f : (0,1) \to V$. The subset of continuous functions and the subset of differentiable functions are both subspaces.

Proposition 3 (LADR 1.C.10). *Let U and W be subspaces of an $\mathbb{F}$-vector space V. Then the intersection $U \cap W$ is also a subspace.*

Proof. (i) Since $0 \in U$ and $0 \in W$, the intersection also contains 0.
(ii) Let $v, w \in U \cap W$ and $\lambda \in \mathbb{F}$. Since $\lambda v + w \in U$ and $\lambda v + w \in W$, the intersection also contains $\lambda v + w$. □

In general, the union of two subspaces is *not* another subspace. The correct analogue of the union in linear algebra is the sum:

Proposition 4 (LADR 1.39)**.** *Let U and W be subspaces of an $\mathbb{F}$-vector space V. Then their* ***sum***

$$U + W = \{v \in V : \exists\, u \in U,\, w \in W \text{ with } v = u + w\}$$

is a subspace of V, and it is the smallest subspace of V that contains the union $U \cup W$.

Proof. $U + W$ contains $0 = 0 + 0$, and: let $u_1, u_2 \in U$, $w_1, w_2 \in W$ and $\lambda \in \mathbb{F}$. Then

$$\lambda(u_1 + w_1) + (u_2 + w_2) = (\lambda u_1 + u_2) + (\lambda w_1 + w_2) \in U + W.$$

$U + W$ is the smallest subspace of V containing the union $U \cup W$ in the following sense: let X be any subspace of V containing $U \cup W$. Let $u \in U$ and $w \in W$ be any elements; then

$$u, w \in U \cup W \subseteq X.$$

Since X is closed under addition, $u + w \in X$; since u and w were arbitrary, $U + W \subseteq X$. $\square$

Be careful not to push the analogy between union and sum too far, though. Some relations that are true for sets, such as the distributive law

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

are not true for subspaces with union replaced by sum: you can find a counterexample with three lines in the plane $\mathbb{R}^2$ for the claim

$$U_1 \cap (U_2 + U_3) = (U_1 \cap U_2) + (U_1 \cap U_3).$$

The special case of disjoint unions is important when studying sets, and this also has an analogue to vector spaces:

Definition 6. Let U and W be subspaces of an $\mathbb{F}$-vector space V. The sum of U and W is **direct** if $U \cap W = \{0\}$. In this case, $U + W$ is denoted $U \oplus W$.

Of course, vector subspaces can never be truly disjoint because they always share 0.

In general, we can take sums of more than 2 subspaces:

$$U_1 + ... + U_m := \{u_1 + ... + u_m : \; u_1 \in U_1, \; ..., \; u_m \in U_m\}.$$

We call the sum direct, written $U_1 \oplus ... \oplus U_m$, if the sum is direct when interpreted as

$$\Big[\Big((U_1 \oplus U_2) \oplus U_3\Big) \oplus ...\Big] \oplus U_m.$$

In other words, $U_1 \cap U_2 = \{0\}$, $U_3 \cap (U_1 + U_2) = \{0\}$, ..., $U_m \cap (U_1 + ... + U_{m-1}) = \{0\}$.

Be careful: a union of sets is disjoint if and only if each pairwise intersection is empty. But a sum is not necessarily direct when each pairwise sum is direct. This comes down to the failure of the distributive law. The example of three lines in the plane $\mathbb{R}^2$ is also a counterexample for this.

Proposition 5 (LADR 1.44,1.45). *Let $U_1, ..., U_m$ be subspaces of a vector space V. The following are equivalent:*
(i) The sum $U_1 \oplus ... \oplus U_m$ is direct;
*(ii) If $v \in U_1 + ... + U_m$ is any element, then there are **unique** $u_1 \in U_1, ..., u_m \in U_m$ with $v = u_1 + ... + u_m$;*
(iii) There do not exist elements $u_1 \in U_1, ..., u_m \in U_m$, not all of which are zero, such that $u_1 + ... + u_m = 0$.

Proof. (i) $\Rightarrow$ (ii): $u_1, ..., u_m$ must exist by definition of the sum $U_1 + ... + U_m$. They are unique, because: assume that

$$v = u_1 + ... + u_m = \tilde{u}_1 + ... + \tilde{u}_m$$

with $\tilde{u}_k \in U_k$ for all k. Then

$$u_m - \tilde{u}_m = (\tilde{u}_1 - u_1) + ... + (\tilde{u}_{m-1} - u_{m-1}) \in U_m \cap (U_1 + ... + U_{m-1}) = \{0\},$$

so $u_m = \tilde{u}_m$. Then

$$u_{m-1} - \tilde{u}_{m-1} = (\tilde{u}_1 - u_1) + ... + (\tilde{u}_{m-1} - u_{m-1}) \in U_{m-1} \cap (U_1 + ... + U_{m-2}) = \{0\},$$

so $u_{m-1} = \tilde{u}_{m-1}$. Continuing in this way, we find that $\tilde{u}_k = u_k$ for all k.

(ii) $\Rightarrow$ (iii): Certainly, $0 = 0 + ... + 0$ is one way to write 0 as a sum of elements from $U_1, ..., U_m$. Claim (ii) implies that this is the only way.

(iii) $\Rightarrow$ (i): Assume that $U_k \cap (U_1 + ... + U_{k-1}) \neq \{0\}$ for some index k, and choose an element

$$0 \neq u_k = u_1 + ... + u_{k-1} \in U_k \cap (U_1 + ... + U_{k-1}), \quad \text{with } u_1 \in U_1, ..., u_{k-1} \in U_{k-1}.$$

Then $u_1 + ... + u_{k-1} - u_k + 0 + ... + 0$ is a combination of 0 by elements that are not all 0, contradicting claim (iii). $\square$

Linear independence and span - 6/22

Linear independence, span and basis

Definition 7 (LADR 2.17)**.** Let V be a vector space. A finite set $\{v_1, ..., v_m\}$ of vectors is **linearly independent** if, given that

$$\lambda_1 v_1 + ... + \lambda_m v_m = 0 \text{ for some } \lambda_1, ..., \lambda_m \in \mathbb{F},$$

we can conclude that all scalars λ_k are 0.

The empty set $\emptyset$ vacuously fulfills this condition, so it is also linearly independent. A set containing the zero vector can *never* fulfill this condition!

Sums of the form

$$\lambda_1 v_1 + ... + \lambda_m v_m, \quad \lambda_1, ..., \lambda_m \in \mathbb{F}$$

are called **linear combinations** of $v_1, ..., v_m$.

Definition 8 (LADR 2.5, 2.8, 2.27, 2.29)**.** Let V be a vector space.
(i) A finite set $\{v_1, ..., v_m\}$ of vectors is a **spanning set**, or **spans** V, if every $v \in V$ can be written as a linear combination

$$v = \lambda_1 v_1 + ... + \lambda_m v_m, \quad \lambda_1, ..., \lambda_m \in \mathbb{F}.$$

(ii) A finite set $\{v_1, ..., v_m\}$ is a **basis** of V if it is a linearly independent spanning set. In other words, every $v \in V$ can be written in a unique way as a linear combination

$$v = \lambda_1 v_1 + ... + \lambda_m v_m, \quad \lambda_1, ..., \lambda_m \in \mathbb{F}.$$

The two conditions for being a basis are equivalent: because having two different representations

$$v = \lambda_1 v_1 + ... + \lambda_m v_m = \mu_1 v_1 + ... + \mu_m v_m$$

is the same as having the nontrivial combination

$$0 = (\lambda_1 - \mu_1) v_1 + ... + (\lambda_m - \mu_m) v_m$$

to zero. See LADR 2.29 for details.

More generally, the **span** of a set of vectors $\{v_1, ..., v_m\}$ is the set of all linear combinations:

$$\mathrm{Span}(v_1, ..., v_m) := \Big\{\lambda_1 v_1 + ... + \lambda_m v_m : \ \lambda_1, ..., \lambda_m \in \mathbb{F}\Big\}.$$

In other words,

$$\mathrm{Span}(v_1, ..., v_m) = U_1 + ... + U_m, \text{ where } U_k := \mathbb{F} \cdot v_k = \{\lambda v_k : \ \lambda \in \mathbb{F}\}.$$

The sum $U_1 + ... + U_m$ is direct if and only if $\{v_1, ..., v_m\}$ was a basis.

Example 8. (i) One basis of $\mathbb{F}^n$ is given by the set

$$\Big\{e_1 := (1, 0, 0, ..., 0), \ e_2 := (0, 1, 0, ..., 0), \ ..., \ e_n := (0, 0, 0, ..., 1)\Big\}.$$

(ii) The span of the sequences

$$(1, 1, 1, 1, ...), \ (0, 1, 2, 3, 4, ...) \in \mathbb{R}^\infty$$

is the set

$$\Big\{(a_0, a_0 + d, a_0 + 2d, a_0 + 3d, ...) = a_0(1, 1, 1, ...) + d(0, 1, 2, ...), \ \ a_0, d \in \mathbb{R}\Big\}$$

of "arithmetic sequences".

(iii) The empty set $\emptyset$ is the only basis of the zero vector space $\{0\}$.

In most of the course, we will want to consider vector spaces that are spanned by finite lists of vectors. Vector spaces that can be spanned by finitely many vectors are called **finite-dimensional**. There are fewer interesting results that hold for all infinite-dimensional vector spaces (but see the remarks at the end).

Proposition 6 (LADR 2.21)**.** *Let V be a vector space and let $\{v_1, ..., v_m\}$ be linearly dependent. Then:*
(i) There is some index j such that $v_j \in \mathrm{Span}(v_1, ..., v_{j-1})$.
(ii) For any such index j,

$$\mathrm{Span}(\{v_1, ..., v_m\} \backslash \{v_j\}) = \mathrm{Span}(v_1, ..., v_m).$$

Proof. (i) Choose a linear combination

$$\lambda_1 v_1 + ... + \lambda_m v_m = 0$$

with scalars $\lambda_1, ..., \lambda_m$ that are not all 0. Let j be the largest index such that $\lambda_j \neq 0$; so the linear combination is actually

$$\lambda_1 v_1 + ... + \lambda_j v_j + 0 + ... + 0 = 0.$$

Then we can divide by λ_j and see

$$v_j = -\frac{\lambda_1}{\lambda_j} v_1 - ... - \frac{\lambda_{j-1}}{\lambda_j} v_{j-1} \in \mathrm{Span}(v_1, ..., v_m).$$

(ii) For any index j such that $v_j \in \mathrm{Span}(v_1, ..., v_{j-1})$, we can find scalars $c_1, ..., c_{j-1}$ with

$$v_j = c_1 v_1 + ... + c_{j-1} v_{j-1}.$$

Now let $v = \lambda_1 v_1 + ... + \lambda_m v_m \in \mathrm{Span}(v_1, ..., v_m)$ be any linear combination; then

$$v = \lambda_1 v_1 + ... + \lambda_{j-1} v_{j-1} + \lambda_j (c_1 v_1 + ... + c_{j-1} v_{j-1}) + \lambda_{j+1} v_{j+1} + ... + \lambda_m v_m$$

is a linear combination only invovling $\{v_1, ..., v_m\} \backslash \{v_j\}$, so

$$v \in \mathrm{Span}(\{v_1, ..., v_m\} \backslash \{v_j\}).$$

Since v was arbitrary,

$$\mathrm{Span}(v_1, ..., v_m) \subseteq \mathrm{Span}(\{v_1, ..., v_m\} \backslash \{v_j\}).$$

The converse inclusion

$$\mathrm{Span}(\{v_1, ..., v_m\} \backslash \{v_j\}) \subseteq \mathrm{Span}(v_1, ..., v_m)$$

is obvious. □

Proposition 7 (LADR 2.31)**.** *Let V be a finite-dimensional vector space and let $\{v_1, ..., v_m\}$ be a spanning set. Then some subset of $\{v_1, ..., v_m\}$ is a basis of V.*

Proof. Consider the collection of all linearly independent subsets of $\{v_1, ..., v_m\}$, and pick any such subset $\mathcal{B}$ that has the largest possible size. Certainly, $\mathcal{B}$ is linearly independent, so we need to show that it spans V. Assume that it doesn't, and choose an index k such that $v_k \notin \mathrm{Span}(\mathcal{B})$. (If $\mathrm{Span}(\mathcal{B})$ contained $v_1, ..., v_m$, then it would contain $\mathrm{Span}(v_1, ..., v_m) = V$.) Then $\mathcal{B} \cup \{v_k\}$ does not satisfy claim (i) of the previous proposition, so it must be linearly independent. This is a contradiction, because $\mathcal{B}$ had the largest possible size. □

Proposition 8 (LADR 2.33)**.** *Let V be a finite-dimensional vector space and let $\{v_1, ..., v_m\}$ be linearly independent. Then there is a basis of V containing $\{v_1, ..., v_m\}$.*

Proof. Consider the collection of all spanning subsets of V that contain $v_1, ..., v_m$, and pick any such subset $\mathcal{B} = \{v_1, ..., v_m, v_{m+1}, ..., v_r\}$ that has the smallest possible size. Certainly, $\mathcal{B}$ spans V, so we need to show that it is linearly independent. Assume it is not; then by the linear dependence lemma (2.21), there is some index j such that $v_j \in \mathrm{Span}(v_1, ..., v_{j-1})$. Since $\{v_1, ..., v_m\}$ is linearly independent, j must be greater than m. Then, also by the linear dependence lemma,

$$\{v_1, ..., v_m, v_{m+1}, ..., v_r\} \backslash \{v_j\}$$

is a set containing v_j whose span is still $\mathrm{Span}(v_1, ..., v_r) = V$. Contradiction, because $\mathcal{B}$ had the smallest possible size. □

Finally, here is an important result relating the sizes of linearly independent and spanning sets. We'll use this tomorrow.

Proposition 9 (LADR 2.23)**.** *Let V be a finite-dimensional vector space, and let $\{v_1, ..., v_n\}$ be a spanning set. Then every linearly independent set contains n vectors or fewer.*

Proof. Let $\{u_1, ..., u_m\}$ be any linearly independent set, and assume that $m > n$. Then $u_1 \neq 0$ is nonzero. If we write $u_1 = \sum_{i=1}^n \lambda_i v_i$, then there is some coefficient $\lambda_i \neq 0$ that is not zero. Without loss of generality, assume $\lambda_1 \neq 0$. Then $\{u_1, v_2, ..., v_n\}$ also spans V, since

$$v_1 = \lambda_1^{-1}\Big(u_1 - \sum_{i=2}^n \lambda_i v_i\Big).$$

Now assume we know that $\{u_1, ..., u_k, v_{k+1}, ..., v_n\}$ is a spanning set for some $1 \leq k < m$. Then we can write

$$u_{k+1} = \sum_{i=1}^k \lambda_i u_i + \sum_{i=k+1}^n \mu_i v_i, \quad \lambda_i, \mu_i \in \mathbb{F}.$$

By linear independence of $\{u_1, ..., u_{k+1}\}$, at least one μ_i is nonzero; without loss of generality, $\mu_{k+1} \neq 0$. Then we can write

$$v_{k+1} = \mu_{k+1}^{-1}\Big(u_{k+1} - \sum_{i=1}^k \lambda_i u_i - \sum_{i=k+2}^n \mu_i v_i\Big),$$

so $\{u_1, ..., u_{k+1}, v_{k+2}, ..., v_n\}$ also spans V.

By induction, we see that $\{u_1, ..., u_n\}$ is a spanning set. This is impossible, because: it implies that u_m is a linear combination

$$u_m = \sum_{i=1}^n \lambda_i u_i,$$

so we get the nontrivial linear combination

$$\lambda_1 u_1 + ... + \lambda_n u_n - u_m = 0$$

to zero. $\square$

Example 9. No set of three vectors in $\mathbb{F}^2$ can ever be linearly independent, since $\mathbb{F}^2$ has a spanning set with fewer than three vectors. This explains the counterexample yesterday of three lines in $\mathbb{R}^2$ - they can never form a direct sum.

Remarks on infinite-dimensional spaces

This section will not be tested: no homework or test problems will refer to infinite linearly independent or spanning sets.

Generally speaking, linear algebra doesn't allow any sort of limit processes. That includes infinite sums. In an infinite-dimensional vector space, "linear independent" and "spanning" sets can contain infinitely many vectors, but the definitions have to be changed such that all linear combinations are finite.

Definition 9. Let V be an $\mathbb{F}$-vector space.
(i) A subset $M \subseteq V$ is **linearly independent** if all of its finite subsets are linearly independent.
(ii) A subset $M \subseteq V$ is a **spanning set** if V is the union of the spans of all finite subsets of M.
(iii) M is a **basis** of V if it is a linearly independent spanning set.

In other words, M is linearly independent if the only *finite* linear combination giving 0 is trivial; and it is a spanning set if every element of V is a *finite* linear combination of the vectors in M.

Example 10. The infinite set $M = \{1, x, x^2, x^3, x^4, ...\}$ is a basis of the space $\mathcal{P}(\mathbb{R})$ of polynomials. Every polynomial has only finitely many terms, so it is a finite combination of M; and a polynomial is zero if and only if its coefficients are all zero.

Example 11. Let $V = \mathbb{F}^\infty$ be the space of sequences. Let e_k be the sequence with 1 at position k and 0 elsewhere, and consider

$$M = \{e_1, e_2, e_3, ...\} = \Big\{(1,0,0,0,...), (0,1,0,0,...), (0,0,1,0,...), ...\Big\}.$$

M is linearly independent, but it is not a basis: the sequence $(1,1,1,1,1,...)$ is not a *finite* linear combination of M.

Proposition 7 and proposition 8 still apply to infinite-dimensional spaces: every linearly independent set can be extended to a basis, and every spanning set shrinks to a basis. In particular, *every* vector space (even an infinite-dimensional space) has a basis. The proofs are similar - but picking the subset $\mathcal{B}$ that has the largest/smallest possible size is no longer possible to do directly. Its existence depends on Zorn's lemma, or equivalently the axiom of choice. If you don't see why this is a difficulty, then try writing down a basis of $\mathbb{F}^\infty$!

Proposition 9 still applies in the sense that there is always an injective map from any linearly independent set into any spanning set.

Dimension - 6/23

Dimension

Proposition 10 (LADR 2.35)**.** *Let V be a finite-dimensional vector space. Then any two bases of V have the same size.*

Proof. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be bases. We use LADR 2.23 twice.
(i) Since $\mathcal{B}_1$ is linearly independent and $\mathcal{B}_2$ is spanning, $\#\mathcal{B}_1 \leq \#\mathcal{B}_2$.
(ii) Since $\mathcal{B}_2$ is linearly independent and $\mathcal{B}_1$ is spanning, $\#\mathcal{B}_2 \leq \#\mathcal{B}_1$. □

This is still true for infinite-dimensional vector spaces, but it is harder to prove.

Definition 10. Let V be a vector space. The **dimension** of V is the size of any basis of V.

Example 12. The dimension of $\mathbb{F}^n$ is n. The basis

$$e_1 = (1, 0, 0, ..., 0),\ e_2 = (0, 1, 0, ..., 0),\ ...\ e_n = (0, 0, 0, ..., 1)$$

consists of n vectors.

Example 13. Let $\mathcal{P}_r(\mathbb{F})$ denote the space of polynomials of degree less than or equal to r. Then the dimension of $\mathcal{P}_r(\mathbb{F})$ is $(r + 1)$, because this is the size of the basis $\{1, x, x^2, ..., x^r\}$.

Example 14. Over any field, the zero vector space $V = \{0\}$ has dimension 0. The only basis is the empty set $\emptyset$, which has zero elements.

Example 15. $\mathbb{C}^2$ can be interpreted either as a $\mathbb{C}$- or an R-vector space. The dimension of $\mathbb{C}^2$ over $\mathbb{C}$ is 2. The dimension of $\mathbb{C}^2$ over $\mathbb{R}$ is 4; one example of a real basis is

$$(1, 0),\ (i, 0),\ (0, 1),\ (0, i).$$

If we know the dimension of V in advance, then it becomes easier to test whether sets are a basis:

Proposition 11 (LADR 2.39, 2.42)**.** *Let V be a finite-dimensional vector space of dimension $d = \dim(V)$.*
(i) Any linearly independent set of d vectors is a basis.
(ii) Any spanning set of d vectors is a basis.

Proof. (i) Let M be a linearly independent set of d vectors, and extend M to a basis $\mathcal{B}$. Then $\mathcal{B}$ also has d vectors, so $M = \mathcal{B}$.
(ii) Let M be a spanning set of d vectors, and shrink M to a basis $\mathcal{B}$. Then $\mathcal{B}$ also has d vectors, so $M = \mathcal{B}$. $\square$

This claim fails dramatically for infinite-dimensional vector spaces. It cannot be fixed.

Example 16. Let $a \in \mathbb{R}$ be a real number. Then $\{1, x-a, (x-a)^2, ..., (x-a)^n\}$ is a linearly independent subset of $\mathcal{P}_n(\mathbb{R})$, since all of the elements have different degrees: there is no way to write a polynomial as a sum of polynomials of lower degree. It must be a basis of $\mathcal{P}_n(\mathbb{R})$, because it consists of $(n+1)$ vectors.
Without input from calculus, it is not that easy to verify directly that this is a spanning set. However, it is clear in the context of Taylor's theorem:

$$p(x) = p(a) \cdot 1 + p'(a) \cdot (x-a) + \frac{p''(a)}{2} \cdot (x-a)^2 + ... + \frac{p^{(n)}(a)}{n!}(x-a)^n,$$

since all derivatives of order greater than n of a polynomial $p \in \mathcal{P}_n(\mathbb{R})$ are 0.

Proposition 12 (LADR 2.38)**.** *Let V be a finite-dimensional vector space, and let $U \subseteq V$ be a subspace. Then* $\dim U \leq \dim V$; *and* $\dim(U) = \dim(V)$ *if and only if $U = V$.*

Proof. Any basis of U is still a linearly independent subset in V, since linear independence doesn't depend on the ambient space. Therefore, the size of any basis of U must be less than or equal to the size of any spanning set of V; in particular, this includes any basis of V.
If $\dim(U) = \dim(V)$, then any basis of U is a linearly independent set of $\dim(V)$ vectors in V, and therefore a basis of V. Since U and V are spanned by the same basis, they are equal. $\square$

Finally, we will work out the relationship between the dimensions of the intersection and sum of two subspaces. We need to understand how to choose bases for two subspaces in a way that makes them compatible together.

Proposition 13. *Let V be a finite-dimensional vector space and let $U, W \subseteq V$ be subspaces. Let $\mathcal{C}$ be any basis of $U \cap W$. Then there is a basis $\mathcal{B}$ of $U + W$ such that $\mathcal{B} \cap U$ is a basis of U, $\mathcal{B} \cap W$ is a basis of W, and $\mathcal{B} \cap U \cap W = \mathcal{C}$.*

Proof. Since $\mathcal{C}$ is linearly independent in $U \cap W$, it is also linearly independent in each of U and W. We can extend $\mathcal{C}$ to a basis $\mathcal{C}_U$ of U and $\mathcal{C}_W$ of W. Written explicitly, let

$$\mathcal{C} = \{v_1, ..., v_r\}, \;\; \mathcal{C}_U = \{v_1, ..., v_r, u_1, ..., u_k\}, \;\; \mathcal{C}_W = \{v_1, ..., v_r, w_1, ..., w_l\}.$$

Then

$$\mathcal{B} := \{v_1, ..., v_r, u_1, ..., u_k, w_1, ..., w_l\}$$

is a basis of $U + W$, because:
(i) Let $u + w \in U + W$ be any element. Then u is a linear combination of $\{v_1, ..., v_r, u_1, ..., u_k\}$ and w is a linear combination of $\{v_1, ..., v_r, w_1, ..., w_l\}$, so adding these combinations together, we see that $u + w$ is a linear combination of $\{v_1, ..., v_r, u_1, ..., u_k, w_1, ..., w_l\}$.
(ii) Let

$$\lambda_1 v_1 + ... + \lambda_r v_r + \mu_1 u_1 + ... + \mu_k u_k + \nu_1 w_1 + ... + \nu_l w_l = 0$$

be any combination to 0. Then

$$v := \underbrace{\lambda_1 v_1 + ... + \lambda_r v_r + \mu_1 u_1 + ... + \mu_k u_k}_{\in U} = \underbrace{-\nu_1 w_1 - ... - \nu_l w_l}_{\in W} \in U \cap W,$$

so there are coefficients $\alpha_i \in \mathbb{F}$ such that

$$v = \alpha_1 v_1 + ... + \alpha_r v_r.$$

Then

$$\begin{aligned} &(\lambda_1 - \alpha_1)v_1 + ... + (\lambda_r - \alpha_r)v_r + \mu_1 u_1 + ... + \mu_k u_k \\ &= \Big(\lambda_1 v_1 + ... + \lambda_r v_r + \mu_1 u_1 + ... + \mu_k u_k\Big) - \Big(\alpha_1 v_1 + ... + \alpha_r v_r\Big) \\ &= v - v = 0. \end{aligned}$$

Since $\{v_1, ..., v_r, u_1, ..., u_k\}$ was linearly independent, it follows that $\lambda_i = \alpha_i$ and $\mu_j = 0$ for all i and j. Therefore,

$$\begin{aligned} 0 &= \lambda_1 v_1 + ... + \lambda_r v_r + \mu_1 u_1 + ... + \mu_k u_k + \nu_1 w_1 + ... + \nu_l w_l \\ &= \lambda_1 v_1 + ... + \lambda_r v_r + 0 + ... + 0 + \nu_1 w_1 + ... + \nu_l w_l \end{aligned}$$

is a linear combination to 0 of the basis $\{v_1, ..., v_r, w_1, ..., w_l\}$ of W. Therefore, $\lambda_i = 0$ and $\nu_j = 0$ for all i, j. □

Here is the corollary:

Proposition 14 (LADR 2.43)**.** *Let U and W be subspaces of a finite-dimensional vector space V. Then*

$$\dim(U+W)+\dim(U\cap W)=\dim(U)+\dim(W).$$

Proof. This follows from looking at the sizes of the bases in the previous proof. In that notation,

$$\dim(U\cap W)=r,\ \dim(U)=r+k,\ \dim(W)=r+l,\ \dim(U+W)=r+k+l.$$

We see that $(r+k+l)+r=(r+k)+(r+l)$. □

In concrete examples, we are often given a basis of U and W. In this case, it is not that hard to find a basis of $U+W$: we know that the union of the two bases will span $U+W$, so we shrink it to a basis by eliminating unnecessary vectors. On the other hand, it is not so easy to find a basis of $U\cap W$ directly.

Example 17. Consider the two planes

$$U=\mathrm{Span}\Big((1,1,0),(0,1,1)\Big),\ \ W=\mathrm{Span}\Big((1,2,2),(2,2,1)\Big)$$

in $\mathbb{F}^3$. It is straightforward to check that $(1,2,2)$ is not in the span of $(1,1,0)$ and $(0,1,1)$ over any field: if we could write

$$(1,2,2)=\lambda(1,1,0)+\mu(0,1,1),$$

then comparing the first coefficient shows that $\lambda=1$ and comparing the last coefficient shows that $\mu=2$, but then $(1,1,0)+2\cdot(0,1,1)=(1,3,2)\neq(1,2,2)$. Therefore,

$$(1,1,0),\ (0,1,1),\ (1,2,2)$$

is a basis of $U+W$.
The formula shows that these planes intersect in a line:

$$\dim(U\cap W)=\dim(U)+\dim(W)-\dim(U+W)=2+2-3=1.$$

In fact, $U\cap W=\mathrm{Span}(-1,0,1)$, but it takes more work to figure that out.

Example 18 (LADR 2.C.17)**.** For a sum of three subspaces, we can use this formula twice to see that

$$\begin{aligned}\dim(U_1+U_2+U_3)&=\dim(U_1+U_2)+\dim(U_3)-\dim((U_1+U_2)\cap U_3)\\&=\dim(U_1)+\dim(U_2)+\dim(U_3)-\dim(U_1\cap U_2)-\dim((U_1+U_2)\cap U_3).\end{aligned}$$

Unfortunately, it is impossible to simplify $\dim((U_1+U_2)\cap U_3)$ further. In particular, the inclusion-exclusion principle for counting the elements in the union of three sets does not carry over. The usual example (three lines in $\mathbb{F}^2$) is a counterexample to the exact claim in 2.C.17.

When all these intersections are $\{0\}$, the formula simplifies considerably:

Proposition 15 (LADR 2.C.16). *Let $U_1, ..., U_m \subseteq V$ be subspaces that form a direct sum. Then*

$$\dim(U_1 \oplus ... \oplus U_m) = \dim(U_1) + ... + \dim(U_m).$$

Proof. Induction on m.
(i) When $m = 1$, this is obvious: $\dim(U_1) = \dim(U_1)$.
(ii) Assume this is true for a direct sum of $(m-1)$ subspaces, for $m \geq 2$. Then

$$\begin{aligned}\dim(U_1 \oplus ... \oplus U_m) &= \dim((U_1 \oplus ... \oplus U_{m-1}) + U_m) \\ &= \dim(U_1 \oplus ... \oplus U_{m-1}) + \dim(U_m) - \dim(\underbrace{(U_1 \oplus ... \oplus U_{m-1}) \cap U_m}_{=\{0\}}) \\ &= \dim(U_1) + ... + \dim(U_{m-1}) + \dim(U_m).\end{aligned}$$

□

Linear maps - 6/27

Linear maps

Let $\mathbb{F}$ be a field. U, V, W and X will denote $\mathbb{F}$-vector spaces. S and T will denote linear maps.

Definition 11 (LADR 3.2)**.** A function $T : V \to W$ is **linear** if it respects linear combinations:

$$T(\lambda_1 v_1 + ... + \lambda_n v_n) = \lambda_1 T(v_1) + ... + \lambda_n T(v_n), \text{ for } v_1, ..., v_n \in V, \ \lambda_1, ..., \lambda_n \in \mathbb{F}.$$

It is enough to know that $T(v + w) = T(v) + T(w)$ and $T(\lambda v) = \lambda T(v)$ for all $v, w \in V$ and $\lambda \in \mathbb{F}$, because every linear combination is built up from sums and scalar multiplication. Equivalently, T is linear if and only if

$$T(\lambda v + w) = \lambda T(v) + T(w), \ \ v, w \in V, \ \lambda \in \mathbb{F}.$$

Example 19. There are lots of examples! Matrix maps are linear (we'll see more about this tomorrow); differentiation between appropriate function spaces is linear; and the map sending a periodic function to its sequence of Fourier coefficients is linear.

Proposition 16 (LADR 3.7)**.** *The set of linear maps from V to W is a vector space with respect to the usual addition and scalar multiplication of functions, denoted $\mathcal{L}(V, W)$.*

Proof. Let $S, T : V \to W$ be linear maps and $\mu \in \mathbb{F}$. Then the linear combination

$$(\mu S + T)(v) := \mu \cdot S(v) + T(v)$$

is linear, because for any $v, w \in V$ and $\lambda \in \mathbb{F}$,

$$\begin{aligned}(\mu S + T)(\lambda v + w) &= \mu \cdot S(\lambda v + w) + T(\lambda v + w) \\ &= \mu(\lambda S(v) + S(w)) + \lambda T(v) + T(w) \\ &= \lambda \cdot (\mu S + T)(v) + (\mu S + T)(w).\end{aligned}$$

The additive identity is the zero function $0(v) = 0$, $v \in V$. This makes $\mathcal{L}(V, W)$ a subspace of the vector space of all functions from V to W. $\square$

Proposition 17 (LADR 3.8)**.** *Let $T : U \to V$ and $S : V \to W$ be linear maps. Then their* ***composition***

$$ST : U \longrightarrow W, \;\; ST(u) := S(T(u))$$

is linear.

Proof. For any $\lambda \in \mathbb{F}$ and $u, v \in U$,

$$ST(\lambda u+v) = S(T(\lambda u+v)) = S(\lambda T(u)+T(v)) = \lambda S(T(u))+S(T(v)) = \lambda ST(u)+ST(v).$$

$\square$

The composition is sometimes called the product - particularly when S, T are matrix maps between spaces of the form $\mathbb{F}^n$ - because it satisfies the basic axioms we expect a product to have:

(1) associativity: $(T_1T_2)T_3 = T_1(T_2T_3)$ for all linear maps $T_3 : U \to V$, $T_2 : V \to W$, $T_3 : W \to X$;

(2) existence of identity: $T\mathrm{id}_V = \mathrm{id}_W T$ for all linear maps $T : V \to W$, where id_V is the identity on V ($\mathrm{id}_V(v) = v \, \forall \, v$) and id_W is the identity on W. If the vector space on which id_V is clear, then we denote it by I.

(3) distributive law: $(S_1 + S_2)T = S_1T + S_2T$ and $S(T_1 + T_2) = ST_1 + ST_2$.

Certainly, composition is not commutative in general: even if they are both defined, ST and TS are not necessarily maps between the same vector spaces.

Definition 12 (LADR 3.67)**.** A linear map $T : V \to V$ from a vector space to itself is called an **operator** on V. The space of operators on V is denoted $\mathcal{L}(V)$, rather than $\mathcal{L}(V, V)$.

Even operators do not generally commute. Physics students may recognize the following example:

Example 20. Let $V = \{\text{infinitely differentiable functions } f : \mathbb{R} \to \mathbb{R}\}$, and define operators $\hat{X}, \hat{P} \in \mathcal{L}(V)$ by

$$\hat{X}(f)(x) := x \cdot f(x), \;\; \hat{P}(f)(x) := f'(x), \;\; x \in \mathbb{R}.$$

Then $(\hat{P}\hat{X} - \hat{X}\hat{P})(f) = (xf)' - xf' = f$, so $\hat{P}\hat{X} - \hat{X}\hat{P} = I \neq 0$.

Now we will prove that, given a basis $\mathcal{B}$ of V, there is a bijection

$$\mathcal{L}(V, W) \leftrightarrow \{\text{functions}\, \mathcal{B} \to W\}.$$

This is proved in a few steps.

Definition 13. Let $T : V \to W$ and let $U \subseteq V$ be a subspace. The **restriction** of T to U is the linear function

$$T|_U : U \longrightarrow W, \;\; T|_U(u) := T(u), \;\; u \in U \subseteq V.$$

Proposition 18. *Let $U_1, ..., U_m \subseteq V$ be subspaces that form a direct sum. Let $T_i : U_i \to W$ be linear functions. Then there is a unique linear function*

$$T : U_1 \oplus ... \oplus U_m \longrightarrow V \text{ with } T|_{U_i} = T_i \text{ for all } i.$$

Proof. The only way to define T is

$$T(u_1 + ... + u_m) = T(u_1) + ... + T(u_m) := T_1(u_1) + ... + T_m(u_m), \quad u_i \in U_i.$$

There is no ambiguity in this definition, since every element $v \in U_1 \oplus ... \oplus U_m$ can be written in only one way as $v = u_1 + ... + u_m$. The map T defined above is linear, because: for any $u_i, v_i \in U_i$ and $\lambda \in \mathbb{F}$,

$$\begin{aligned}
&T\Big(\lambda(u_1 + ... + u_m) + (v_1 + ... + v_m)\Big) \\
&= T\Big((\lambda u_1 + v_1) + ... + (\lambda u_m + v_m)\Big) \\
&= T_1(\lambda u_1 + v_1) + ... + T_m(\lambda u_m + v_m) \\
&= \lambda T_1(u_1) + T_1(v_1) + ... + \lambda T_m(u_m) + T_m(v_m) \\
&= \lambda\Big(T_1(u_1) + ... + T_m(u_m)\Big) + \Big(T_1(v_1) + ... + T_m(v_m)\Big) \\
&= \lambda T(u_1 + ... + u_m) + T(v_1 + ... + v_m).
\end{aligned}$$

□

Proposition 19 (LADR 3.5)**.** *Let $v_1, ..., v_n$ be a basis of V and $w_1, ..., w_n \in W$. Then there is a unique linear map $T : V \to W$ such that $T(v_j) = w_j$ for all j.*

Proof. Consider the subspaces $U_j := \mathbb{F} \cdot v_j$. There is a unique linear map $T_j : U_j \to W$ with $T_j(v_j) = w_j$, and it is

$$T_j : U_j \longrightarrow W, \;\; T_j(\lambda v_j) := \lambda w_j, \;\; \lambda \in \mathbb{F}.$$

By the previous proposition, there is a unique linear map

$$T : U_1 \oplus ... \oplus U_n = V \longrightarrow W \;\; \text{with } T|_{U_j} = T_j,$$

or equivalently $T(v_j) = w_j$ for all j. □

Range and null space

Definition 14 (LADR 3.12, 3.17)**.** Let $T : V \to W$ be a linear map.
(i) The **null space**, or **kernel**, of T is

$$\mathrm{null}(T) = \ker(T) = \{v \in V : \ T(v) = 0\}.$$

(ii) The **range**, or **image**, of T is

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

Example 21. Consider the differentiation operator $D \in \mathcal{L}(\mathcal{P}_4(\mathbb{R}))$. The null space of D is

$$\mathrm{null}(D) = \{p : \ p' = 0\} = \mathrm{Span}(1).$$

The range is all of $\mathcal{P}_3(\mathbb{R})$: since $\{1, x, x^2, x^3, x^4\}$ spans $\mathcal{P}_4(\mathbb{R})$,

$$\{D(1), D(x), D(x^2), D(x^3), D(x^4)\} = \{0, 1, 2x, 3x^2, 4x^3\}$$

spans $\mathrm{range}(T)$.

Proposition 20 (LADR 3.14, 3.19)**.** *Let $T \in \mathcal{L}(V, W)$. Then $\mathrm{null}(T) \subseteq V$ and $\mathrm{range}(T) \subseteq W$ are subspaces.*

Proof. (i) Since $T(0_V) = 0_W$, we know that $0_V \in \mathrm{null}(T)$ and $0_W \in \mathrm{range}(T)$.
(ii) For any $v_1, v_2 \in \mathrm{null}(T)$ and $\lambda \in \mathbb{F}$,

$$T(\lambda v_1 + v_2) = \lambda T(v_1) + T(v_2) = \lambda \cdot 0 + 0 = 0,$$

so $\lambda v_1 + v_2 \in \mathrm{null}(T)$.
(iii) For any $w_1 = T(v_1)$ and $w_2 = T(v_2) \in \mathrm{range}(T)$ and $\lambda \in \mathbb{F}$,

$$\lambda w_1 + w_2 = \lambda T(v_1) + T(v_2) = T(\lambda v_1 + v_2) \in \mathrm{range}(T).$$

□

Definition 15. The dimensions of $\mathrm{null}(T)$ resp. $\mathrm{range}(T)$ are called the **nullity** resp. the **rank** of T.

Nullity and rank are more precise variants of the concepts of "injective" and "surjective", in the following sense:

Proposition 21 (LADR 3.16, 3.20)**.** *Let $T \in \mathcal{L}(V, W)$. Then T is injective if and only if $\mathrm{null}(T) = \{0\}$, and T is surjective if and only if $\mathrm{range}(T) = W$.*

Proof. (i) Assume that T is injective, and $v \in \mathrm{null}(T)$. Then $T(v) = 0 = T(0)$, so $v = 0$. On the other hand, assume that $\mathrm{null}(T) = \{0\}$. If $T(v) = T(w)$, then $T(v - w) = 0$, so $v - w \in \mathrm{null}(T) = \{0\}$ and $v = w$.
(ii) T is surjective if and only if $\mathrm{range}(T) = W$: this is clear from the definition of "surjective". □

The following result is called the Fundamental Theorem of Linear Maps in the textbook. Many other books call it the rank-nullity theorem. The rank-nullity theorem itself is only a statement about dimensions; for some applications, the particular basis we construct is important, so it is included in the statement below.

Proposition 22 (LADR 3.22)**.** *Let V be a finite-dimensional vector space and let $T \in \mathcal{L}(V, W)$. Then there is a basis $\{u_1, ..., u_m, v_1, ..., v_n\}$ of V such that $\{u_1, ..., u_m\}$ is a basis of* $\mathrm{null}(T)$ *and $T(v_1), ..., T(v_n)$ is a basis of* $\mathrm{range}(T)$. *In particular,*

$$\dim(V) = \dim \mathrm{null}(T) + \dim \mathrm{range}(T).$$

Proof. Choose any basis $\{u_1, ..., u_m\}$ of $\mathrm{null}(T)$ and extend it to a basis $\{u_1, ..., u_m, v_1, ..., v_n\}$ of V. Then $T(v_1), ..., T(v_n)$ is a basis of $\mathrm{range}(T)$, because:
(i) It is linearly independent: assume that

$$\lambda_1 T(v_1) + ... + \lambda_n T(v_n) = 0.$$

Then $T(\lambda_1 v_1 + ... + \lambda_n v_n) = 0$, so

$$\lambda_1 v_1 + ... + \lambda_n v_n \in \mathrm{null}(T) = \mathrm{Span}(u_1, ..., u_m).$$

If we write $\lambda_1 v_1 + ... + \lambda_n v_n = \mu_1 u_1 + ... + \mu_m u_m$, then

$$\mu_1 u_1 + ... + \mu_m u_m - \lambda_1 v_1 - ... - \lambda_n v_n = 0$$

is a combination to 0; since $\{u_1, ..., u_m, v_1, ..., v_n\}$ is linearly independent, we see that all μ_i and λ_j are 0.
(ii) It spans $\mathrm{range}(T)$: let $w = T(v) \in \mathrm{range}(T)$ be any element, and write

$$v = \lambda_1 v_1 + ... + \lambda_n v_n + \mu_1 u_1 + ... + \mu_m u_m.$$

Then

$$T(v) = \lambda_1 T(v_1) + ... + \lambda_n T(v_n) + \mu_1 \underbrace{T(u_1)}_{=0} + ... + \mu_m \underbrace{T(u_m)}_{=0} = \lambda_1 T(v_1) + ... + \lambda_n T(v_n). \quad \square$$

Matrices - 6/28

Matrices

For working with matrices, we use the convention that $\mathbb{F}^n$ consists of column vectors, rather that row vectors.

Definition 16 (LADR 3.30,3.39)**.** Let $m, n \in \mathbb{N}$. An $(m \times n)$-**matrix** is a rectangular array

$$A = (a_{ij})_{i,j} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$$

with entries $a_{ij} \in \mathbb{F}$. The set of matrices is denoted $\mathbb{F}^{m,n}$.

It is straightforward to check that $\mathbb{F}^{m,n}$ is an $\mathbb{F}$-vector space with entrywise addition and scalar multiplication, just as $\mathbb{F}^n$ is. A natural basis is given by the matrices $E_{i,j}$, with a 1 in the ith row and jth column, and 0 elsewhere. In particular, $\dim(\mathbb{F}^{m,n}) = mn$.

Example 22. In $\mathbb{C}^{2,3}$,

$$\begin{pmatrix} 1+i & 3+2i & 4 \\ 5 & i & 2+i \end{pmatrix} + \begin{pmatrix} 4-2i & 5 & 1 \\ 3 & 2 & i \end{pmatrix} = \begin{pmatrix} 5-i & 8+2i & 5 \\ 8 & 2+i & 2+2i \end{pmatrix},$$

and

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

Example 23. A natural basis for $\mathbb{F}^{2,2}$ is

$$E_{1,1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},\ E_{1,2} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},\ E_{2,1} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},\ E_{2,2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Definition 17 (LADR 3.32)**.** Let $\mathcal{B} = \{v_1, ..., v_n\}$ be a basis of V and let $\mathcal{C} = \{w_1, ..., w_m\}$ be a basis of W. Let $T : V \to W$. The **representation matrix** of T with respect to $\mathcal{B}$ and $\mathcal{C}$ is the matrix

$$\mathcal{M}(T) = \mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(T) = (a_{ij})_{i,j},$$

where the entries a_{ij} are defined by

$$T(v_j) = a_{1j}w_1 + ... + a_{mj}w_m = \sum_{i=1}^{m} a_{ij}w_i.$$

Example 24. Consider the linear map

$$T : \mathcal{P}_3(\mathbb{C}) \longrightarrow \mathbb{C}^4, \;\; T(p) := (p(1), p(i), p(-1), p(-i)).$$

We will calculate its representation matrix with respect to the usual bases $\{1, x, x^2, x^3\}$ and $\{e_1, e_2, e_3, e_4\}$.
The first column of $\mathcal{M}(T)$ is found by applying T to the constant 1, and writing the result as a column vector; we get

$$\mathcal{M}(T) = \begin{pmatrix} 1 & ? & ? & ? \\ 1 & ? & ? & ? \\ 1 & ? & ? & ? \\ 1 & ? & ? & ? \end{pmatrix}.$$

The second column of $\mathcal{M}(T)$ is found by applying T to x; we get

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 1 & ? & ? \\ 1 & i & ? & ? \\ 1 & -1 & ? & ? \\ 1 & -i & ? & ? \end{pmatrix}.$$

The third column of $\mathcal{M}(T)$ is found by applying T to x^2; we get

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 1 & 1 & ? \\ 1 & i & -1 & ? \\ 1 & -1 & 1 & ? \\ 1 & -i & -1 & ? \end{pmatrix}.$$

Finally, apply T to x^3 to find

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}.$$

The representation matrix gives a correspondence between matrices and linear functions. It is useful in two ways: it is easier to do practical calculations with matrices, but it is easier to prove theorems about linear functions.

Proposition 23. *Let V and W be finite-dimensional vector spaces. Let $\mathcal{B} = \{v_1, ..., v_n\}$ be a basis of V and let $\mathcal{C} = \{w_1, ..., w_m\}$ be a basis of W. There is a bijection*
$$\mathcal{L}(V, W) \leftrightarrow \mathbb{F}^{m,n}, \;\; T \mapsto \mathcal{M}(T).$$

This bijection also preserves the vector space structure; i.e. $\mathcal{M}(S+T) = \mathcal{M}(S)+\mathcal{M}(T)$ and $\mathcal{M}(\lambda T) = \lambda \cdot \mathcal{M}(T)$.

Proof. For any elements $x_1, ..., x_n \in W$, there is a unique linear map $T \in \mathcal{L}(V, W)$ with $T(v_j) = x_j$ for all j; and for any elements $x_1, ..., x_n \in W$, there is a unique matrix $(a_{ij})_{i,j}$ with $x_j = \sum_{i=1}^m a_{ij} w_i$ for all j. In other words, there are two bijections
$$\mathcal{L}(V, W) \leftrightarrow \{\text{lists}\, x_1, ..., x_n \in W\} \leftrightarrow \text{matrices},$$
$$T \mapsto \Big(T(v_1), ..., T(v_n)\Big) \mapsto \mathcal{M}(T).$$
□

Definition 18 (LADR 3.62)**.** Let V be a finite-dimensional vector space with an ordered basis $\mathcal{B} = \{v_1, ..., v_n\}$. For any $v \in V$, the **coordinates** of v with respect to $\mathcal{B}$ are
$$\mathcal{M}(v) := \mathcal{M}_B(v) := \begin{pmatrix} \lambda_1 \\ ... \\ \lambda_n \end{pmatrix}, \;\; \text{where } v = \lambda_1 v_1 + ... + \lambda_n v_n.$$

In other words, $\mathcal{M}(v) = \sum_{i=1}^n \lambda_i e_i$ if $v = \sum_{i=1}^n \lambda_i v_i$.

Example 25. The coordinates of $1 + 2x + 3x^2 \in \mathcal{P}_2(\mathbb{R})$ with respect to the basis $\{1, x, x^2\}$ are $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

The coordinates of $1 + 2x + 3x^2$ with respect to the basis $\{1 + x, 1 + x^2, x + x^2\}$ are $\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$, because
$$1 + 2x + 3x^2 = 0 \cdot (1 + x) + 1 \cdot (1 + x^2) + 2 \cdot (x + x^2).$$

Definition 19 (LADR 3.65)**.** We define the product of a matrix and a vector by

$$\mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(T) \cdot \mathcal{M}_{\mathcal{B}}(v) := \mathcal{M}_{\mathcal{C}}(T(v))$$

for any linear map $T : V \to W$ and vector $v \in V$, and any bases $\mathcal{B}$ of V and $\mathcal{C}$ of W.

We should check that this doesn't depend on the linear maps and bases, but only on the matrix $\mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(T)$ and column vector $\mathcal{M}_{\mathcal{B}}(v)$. Let $\mathcal{B} = \{v_1, ..., v_n\}$ and $\mathcal{C} = \{w_1, ..., w_m\}$. Let $\mathcal{M}(T) = (a_{ij})_{i,j}$ and $v = \sum_{j=1}^{n} \lambda_j v_j$. Then

$$T(v) = \sum_{j=1}^{n} \lambda_j T(v_j) = \sum_{j=1}^{n} \Big(\lambda_j \sum_{i=1}^{m} a_{ij} w_i \Big) = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij} \lambda_j \Big) w_i,$$

so

$$\mathcal{M}_{\mathcal{C}}(T(v)) = \sum_{j=1}^{n} a_{ij} \lambda_j,$$

which depends only on the matrix $(a_{ij})_{i,j}$ and the column vector $(\lambda_j)_{j=1}^{n}$.

Example 26.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix} = \begin{pmatrix} 1 \cdot 5 + 2 \cdot 7 + 3 \cdot 9 \\ 2 \cdot 5 + 3 \cdot 7 + 4 \cdot 9 \end{pmatrix} = \begin{pmatrix} 46 \\ 67 \end{pmatrix}.$$

Example 27. Let $T : \mathcal{P}_3(\mathbb{C}) \to \mathbb{C}^4,\ T(p) := (p(1), p(i), p(-1), p(-i))$ be the linear map from above. To calculate $T(1+2x+x^2)$, we can express $1+2x+x^2$ by coordinates in the basis $\{1, x, x^2, x^3\}$, i.e. $\begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix}$, and use matrix multiplication:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 2i \\ 0 \\ -2i \end{pmatrix}.$$

The product (composition) of functions gives us a product of matrices:

Definition 20 (LADR 3.43)**.** We define the product of two matrices by

$$\mathcal{M}_{\mathcal{C}}^{\mathcal{A}}(ST) := \mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(S) \cdot \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(T),$$

for any linear maps $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, and bases $\mathcal{A}$ of U; $\mathcal{B}$ of V; and $\mathcal{C}$ of W.

We should check that this doesn't depend on the linear maps and bases, but only on the matrices $\mathcal{M}(S)$ and $\mathcal{M}(T)$ themselves. If $\mathcal{A} = \{u_1, ..., u_l\}$, $\mathcal{B} = \{v_1, ..., v_m\}$ and $\mathcal{C} = \{w_1, ..., w_n\}$, and $\mathcal{M}(S) = (a_{ij})_{i,j}$ and $\mathcal{M}(T) = (b_{jk})_{j,k}$, then $\mathcal{M}(ST) = (c_{ik})_{i,k}$ is defined as follows:

$$\begin{aligned}
\sum_{i=1}^n c_{ik} w_i &= ST(u_k) \\
&= S(T(u_k)) \\
&= S\Big(\sum_{j=1}^m b_{jk} v_j\Big) \\
&= \sum_{j=1}^m b_{jk} S(v_j) \\
&= \sum_{j=1}^m b_{jk} \sum_{i=1}^n a_{ij} w_i \\
&= \sum_{i=1}^n \Big(\sum_{j=1}^m a_{ij} b_{jk}\Big) w_i,
\end{aligned}$$

or in other words $c_{ik} = \sum_{j=1}^m a_{ij} b_{jk}$.

Example 28. Let $T : \mathcal{P}_3(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ be the differentiation map and let $S : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_3(\mathbb{R})$ be the antiderivative without constant term; i.e. $S(p)(x) = \int_0^x p(t)\,\mathrm{d}t$. With respect to the standard bases, the composition ST has the matrix

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Note that matrix-vector multiplication is a special case of matrix-matrix multiplication: if we let $T : \mathbb{F} \to V$ be the linear map with $T(1) = v$ and choose the standard basis $\{1\}$ of $\mathbb{F}$, then $\mathcal{M}(T) = \mathcal{M}(v)$, and

$$\mathcal{M}(ST) = \mathcal{M}(S) \cdot \mathcal{M}(T) = \mathcal{M}(S) \cdot \mathcal{M}(v) = \mathcal{M}(S(v)).$$

The following properties are easier to check for the underlying linear functions, rather than for matrix multiplication directly:

Proposition 24.
(i) Let $A \in \mathbb{F}^{m,n}$, $B \in \mathbb{F}^{n,r}$, $C \in \mathbb{F}^{r,s}$ *be matrices. Then* $(AB) \cdot C = A \cdot (BC)$.
(ii) Let $A \in \mathbb{F}^{m,n}$ *and* $B, C \in \mathbb{F}^{n,r}$. *Then* $A(B+C) = AB + AC$.
(iii) Let $A, B \in \mathbb{F}^{m,n}$ *and* $C \in \mathbb{F}^{n,r}$. *Then* $(A+B)C = AC + BC$.
(iv) The identity matrix

$$I = \mathcal{M}(\mathrm{id}_V) = \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 \end{pmatrix} \in \mathbb{F}^{n,n}$$

satisfies $AI = A$ *and* $IB = B$ *for all* $A \in \mathbb{F}^{m,n}$ *and* $B \in \mathbb{F}^{n,r}$.

Isomorphisms - 6/29

Isomorphisms

Definition 21 (LADR 3.53, 3.58)**.** Let V and W be vector spaces. A linear map $T : V \to W$ is an **isomorphism** if there is a linear map $S : W \to V$ such that $TS = \mathrm{id}_W$ and $ST = \mathrm{id}_V$.
If there is an isomorphism $T : V \to W$, we call V and W **isomorphic** and write $V \cong W$.

Isomorphic vector spaces are, for most purposes, the same. The map $T : V \to W$ can be thought of as a "relabeling" of the elements of V.

As usual, the inverse S, if it exists, is unique:

Proposition 25 (LADR 3.54)**.** *Let $T : V \to W$ be an isomorphism. Then the inverse is unique, and is denoted $T^{-1} : W \to V$.*

Proof. Assume that $S_1, S_2 : W \to V$ are two linear maps such that $TS_1 = \mathrm{id}_W = TS_2$ and $S_1T = \mathrm{id}_V = S_2T$. Then

$$S_1 = S_1(TS_2) = (S_1T)S_2 = S_2. \qquad \square$$

What we have shown here is a little stronger: if T has both a "left-inverse" and a "right-inverse", then the left- and right-inverses are equal. It follows that any two left-inverses are equal: they both equal the right-inverse. Similarly, any two right-inverses are equal.

We can't prove this if we don't know that T has an inverse on both sides; for example, the matrix $A = \begin{pmatrix} 1 & 2 \end{pmatrix}$ has no left-inverse and infinitely many right-inverses:

$$\begin{pmatrix} 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 - 2x \\ x \end{pmatrix} = 1 = \mathrm{id}_{\mathbb{F}}, \quad x \in \mathbb{F}.$$

Proposition 26 (LADR 3.56)**.** *Let $T : V \to W$ be a linear map. Then T is an isomorphism if and only if it is both injective and surjective.*

Proof. (i) Assume that T is an isomorphism. Then T is injective (i.e. $\mathrm{null}(T) = \{0\}$), because:

$$v \in \mathrm{null}(T) \implies v = T^{-1}(T(v)) = T^{-1}(0) = 0.$$

T is surjective, because: any $w \in W$ has the form

$$w = T(T^{-1}(w)) \in \mathrm{range}(T).$$

(ii) Assume that T is injective and surjective. Define T^{-1} such that

$$T^{-1}(w) = v \text{ is the unique vector with } T(v) = w;$$

so $T(T^{-1}(w)) = w$ and $T^{-1}(T(v)) = v$. Then T^{-1} is linear, because: for any $\lambda \in \mathbb{F}$ and $w_1, w_2 \in W$,

$$T(T^{-1}(\lambda w_1 + w_2)) = \lambda w_1 + w_2 = \lambda T(T^{-1}(w_1)) + T(T^{-1}(w_2)) = T(\lambda T^{-1}(w_1) + T^{-1}(w_2))$$

shows that $T^{-1}(\lambda w_1 + w_2) = \lambda T^{-1}(w_1) + T^{-1}(w_2)$. $\square$

Proposition 27 (LADR 3.59)**.** *Two finite-dimensional vector spaces are isomorphic if and only if they have the same dimension.*

This is also true for infinite-dimensional vector spaces, although the dimension needs to be made more precise than simply "∞".

Proof. (i) Assume that $T : V \to W$ is an isomorphism; then $W = \mathrm{range}(T)$ and $\mathrm{null}(T) = \{0\}$. By the rank-nullity theorem,

$$\dim(V) = \dim \mathrm{null}(T) + \dim \mathrm{range}(T) = 0 + \dim(W).$$

(ii) Assume that V and W have the same dimension, and choose bases $\{v_1, ..., v_n\}$ of V and $\{w_1, ..., w_n\}$ of W. Let

$$T : V \longrightarrow W \text{ and } T^{-1} : W \longrightarrow V$$

be the unique linear maps such that $T(v_j) = w_j$ and $T^{-1}(w_j) = v_j$. Then

$$TT^{-1}(w_j) = T(v_j) = w_j = \mathrm{id}_W(w_j) \text{ and } T^{-1}T(v_j) = T^{-1}(w_j) = v_j = \mathrm{id}_V(v_j),$$

so $T^{-1}T$ and id_V, and TT^{-1} and id_W agree on a basis; so $T^{-1}T = \mathrm{id}_V$ and $TT^{-1} = \mathrm{id}_W$. Therefore, T is an isomorphism. $\square$

If we already know that the vector spaces V and W have the same finite dimension, then it becomes easier to test whether a linear map $T : V \to W$ is an isomorphism. (Often, proving surjectivity is the harder part, since for any $w \in W$ we have to "guess" an element $v \in V$ with $T(v) = w$.)

Proposition 28 (LADR 3.69)**.** *Assume that* $\dim(V) = \dim(W) < \infty$*, and let* $T \in \mathcal{L}(V, W)$*. The following are equivalent:*
(i) T *is an isomorphism;*
(ii) T *is injective;*
(iii) T *is surjective.*

Proof. (i) $\Rightarrow$ (ii): This follows from LADR 3.56.
(ii) $\Rightarrow$ (iii): By the rank-nullity theorem,

$$\dim \mathrm{range}(T) = \dim(V) - \underbrace{\dim \mathrm{null}(T)}_{=0} = \dim(V) = \dim(W);$$

therefore, $\mathrm{range}(T) = W$, and T is surjective.
(iii) $\Rightarrow$ (i): By the rank-nullity theorem,

$$\dim \mathrm{null}(T) = \dim(V) - \dim \mathrm{range}(T) = \dim(V) - \dim(W) = 0.$$

Therefore, $\mathrm{null}(T) = \{0\}$, so T is also injective. By LADR 3.56, it is an isomorphism. $\square$

Example 29. Let $(x_0, y_0), ..., (x_n, y_n)$ be any $(n+1)$ points in the real plane $\mathbb{R}^2$, where $x_0, ..., x_n$ are distinct. We will show that there is a unique interpolating polynomial of degree at most n; i.e. a polynomial $p \in \mathcal{P}_n(\mathbb{R})$ with $p(x_j) = y_j$ for all j.
The map

$$T : \mathcal{P}_n(\mathbb{R}) \longrightarrow \mathbb{R}^{n+1}, \;\; T(p) := (p(x_0), ..., p(x_n))$$

is injective, because: any nonzero polynomial with $(n+1)$ distinct zeros $x_0, ..., x_n$ must have all of $(x - x_0), ..., (x - x_n)$ as factors, and therefore have degree at least $n+1$. This means that $\mathrm{null}(T) = \{0\}$. By the previous theorem, it is also surjective.

Example 30. This fails for infinite-dimensional vector spaces. For example, the differentiation map

$$T : \mathcal{P}(\mathbb{R}) \longrightarrow \mathcal{P}(\mathbb{R}), \;\; p \mapsto p'$$

is surjective, but not injective.

Taking inverses reverses the order of multiplication:

Proposition 29 (LADR 3.D.1, 3.D.9)**.** *Let* $S, T \in \mathcal{L}(V)$ *be operators.*
(i) If S *and* T *are both invertible, then* ST *is invertible with inverse* $(ST)^{-1} = T^{-1}S^{-1}$.
(ii) If ST *is invertible, then* S *is surjective and* T *is injective.*

In particular, if V is finite-dimensional, then S and T are both invertible, and by (i), TS is also invertible. This does not need to be true when V is not finite-dimensional.

Proof. (i) We can check that

$$(T^{-1}S^{-1}) \cdot (ST) = T^{-1}(S^{-1}S)T = T^{-1}T = \mathrm{id}_V$$

and

$$(ST) \cdot (T^{-1}S^{-1}) = S(TT^{-1})S^{-1} = SS^{-1} = \mathrm{id}_V.$$

(ii) Assume that ST is invertible. Then S is surjective, because: any $v \in V$ can be written as

$$v = (ST)(ST)^{-1}(v) = S\Big(T(ST)^{-1}v\Big) \in \mathrm{range}(S).$$

Also, T is injective, because: if $T(v) = 0$, then

$$v = (ST)^{-1}(ST)(v) = (ST)^{-1}S(0) = 0. \qquad \square$$

Note that even if V is infinite-dimensional and ST is invertible, the fact that T is injective means that T defines an isomorphism

$$T : V \longrightarrow \mathrm{range}(T).$$

It follows by (i) that

$$S|_{\mathrm{range}(T)} = (ST) \cdot T^{-1} : \mathrm{range}(T) \longrightarrow V$$

is also invertible, and we can write

$$(ST)^{-1} = T^{-1}(S|_{\mathrm{range}(T)})^{-1}.$$

In this sense, the formula for the inverse is still valid.

Example 31. Let $V = \mathcal{P}(\mathbb{R})$ with linear maps

$$S(p) := p', \;\; T(p)(x) := \int_0^{x+1} p(t)\,\mathrm{d}t.$$

Then $ST(p)(x) = p(x+1)$, so ST is invertible with inverse $p(x) \mapsto p(x-1)$. The range of T is

$$\mathrm{range}(T) = \{p \in \mathcal{P}(\mathbb{R}) : \; p(-1) = 0\} =: U,$$

and the inverse of ST is the composition of

$$(S|_U)^{-1} : V \longrightarrow U, \; (S|_U)^{-1}(p)(x) = \int_{-1}^{x} p(t)\,\mathrm{d}t$$

and

$$T^{-1} : U \longrightarrow V, \; T^{-1}(p)(x) = p'(x-1).$$

Notice that TS is not invertible because it sends all constants to 0.

Here is a useful corollary:

Proposition 30 (LADR 3.D.10)**.** *Let V be a finite-dimensional vector space and let $S, T \in \mathcal{L}(V)$ be operators. Then $ST = I$ if and only if $TS = I$.*

Proof. Assume that $ST = \mathrm{id}_V$. Then T is injective and S is surjective by the previous proposition. By LADR 3.69, S and T are isomorphisms, and

$$S = S(TT^{-1}) = (ST)T^{-1} = T^{-1},$$

so $TS = TT^{-1} = I$.
The other direction is the same proof, with the roles of S and T swapped. $\square$

Example 32. This is also false for infinite-dimensional spaces. The operators

$$S, T \in \mathcal{L}(\mathcal{P}(\mathbb{R})), \;\; S(p) := p', \; T(p)(x) := \int_0^x p(t)\,\mathrm{d}t$$

satisfy $ST = \mathrm{id}$ but not $TS = \mathrm{id}$.

Product and quotient space - 6/30

Product

Recall that the Cartesian product of two sets A and B is the set of ordered pairs

$$A \times B = \{(a, b) : \ a \in A, \, b \in B\};$$

more generally, if $A_1, ..., A_m$ are sets, their product is the set of lists

$$\prod_{i=1}^{m} A_i = A_1 \times ... \times A_m = \{(a_1, ..., a_m) : \ a_i \in A_i\}.$$

Proposition 31 (LADR 3.73)**.** *Let $V_1, ..., V_m$ be vector spaces. Then their product*

$$V_1 \times ... \times V_m$$

is a vector space with the componentwise operations

$$(v_1, ..., v_m) + (w_1, ..., w_m) := (v_1 + w_1, ..., v_m + w_m), \quad v_i, w_i \in V_i$$

and

$$\lambda \cdot (v_1, ..., v_m) := (\lambda v_1, ..., \lambda v_m).$$

Proof. All of the axioms are straightforward to verify. The zero element is $0 = (0, ..., 0)$, and the additive inverse of $(v_1, ..., v_m)$ is $(-v_1, ..., -v_m)$. $\square$

Example 33. $\mathbb{R}^n$ is the product $\underbrace{\mathbb{R} \times ... \times \mathbb{R}}_{n \text{ times}}$.

Example 34. There is a natural identification

$$\mathbb{R}^m \times \mathbb{R}^n \xrightarrow{\sim} \mathbb{R}^{m+n}, \quad \Big((x_1, ..., x_m), (y_1, ..., y_n)\Big) \mapsto (x_1, ..., x_m, y_1, ..., y_n).$$

In general, even though $V_1 \times (V_2 \times V_3)$ and $(V_1 \times V_2) \times V_3$ and $V_1 \times V_2 \times V_3$ are technically not the same spaces, there are natural ways to identify them with each other and mathematicians often do not bother distinguishing these spaces. The same is true for $V \times W$ and $W \times V$.

Let $V_1, ..., V_m$ be vector spaces. Each V_i is isomorphic to the subspace

$$\{0\} \times ... \times \{0\} \times V_i \times \{0\} ... \times \{0\} = \{(0, ..., 0, v^{(i)}, 0, ..., 0) : v^{(i)} \in V_i\} \subseteq V_1 \times ... \times V_m$$

via the obvious map

$$V_i \longrightarrow \{0\} \times ... \times \{0\} \times V_i \times \{0\} ... \times \{0\}, \quad v^{(i)} \mapsto (0, ..., 0, v^{(i)}, 0, ..., 0).$$

We will abbreviate $\{0\} \times ... \times \{0\} \times V_i \times \{0\} \times ... \times \{0\}$ by $V_{(i)}$.

Proposition 32. *Let $V_1, ..., V_m$ be vector spaces. Then $V_1 \times ... \times V_m$ is a direct sum of its subspaces $V_{(i)}$.*

For this reason, the product is also called the **exterior direct sum**. In fact, when V and W are unrelated vector spaces (not subspaces of an ambient vector space), the notation $V \oplus W$ is used equivalently to $V \times W$ - although we will avoid using "$\oplus$" that way in this course.

Proof. Every $v = (v^{(1)}, ..., v^{(m)}) \in V_1 \times ... \times V_m$ can be written in exactly one way as a sum

$$v = (v^{(1)}, 0, ..., 0) + ... + (0, ..., 0, v^{(m)}).$$

□

Proposition 33 (LAD+ 3.76)**.** *Let $V_1, ..., V_m$ be finite-dimensional vector spaces. Then $V_1 \times ... \times V_m$ is finite-dimensional, and*

$$\dim(V_1 \times ... \times V_n) = \dim(V_1) + ... + \dim(V_m).$$

The dimension of the product is *not* the product of the dimensions! There is another construction, called the tensor product, whose dimension is actually the product of the dimensions of each space. We will not discuss this here, but you have already seen one example of it: the space of linear maps $\mathcal{L}(V, W)$, which has dimension $\dim(V) \cdot \dim(W)$.

Proof. This follows from the formula for the dimension of a direct sum:

$$\begin{aligned} \dim(V_1 \times ... \times V_m) &= \dim(V_{(1)} \oplus ... \oplus V_{(m)}) \\ &= \dim(V_{(1)}) + ... + \dim(V_{(m)}) \\ &= \dim(V_1) + ... + \dim(V_m). \end{aligned}$$

□

Here is the relationship between sums and products:

Proposition 34 (LADR 3.77)**.** *Let $U_1, ..., U_m \subseteq V$ be subspaces of a vector space V. Then there is a surjective linear map*

$$\Gamma : U_1 \times ... \times U_m \longrightarrow U_1 + ... + U_m, \quad \Gamma(u_1, ..., u_m) := u_1 + ... + u_m,$$

and it is injective if and only if $U_1, ..., U_m$ form a direct sum.

Proof. Γ is surjective by definition of the sum $U_1 + ... + U_m$. It is injective if and only if the only way to write

$$0 = u_1 + ... + u_m \text{ with } u_i \in U_i$$

is when $u_i = 0$ for all i. This is equivalent to $U_1, ..., U_m$ forming a direct sum. $\square$

This gives us a very clean proof of problem 8 on the first problem set: since Γ is surjective, we know that

$$\dim(U_1 + ... + U_m) = \dim \operatorname{range}(\Gamma) \leq \dim(U_1 \times ... \times U_m) = \dim(U_1) + ... + \dim(U_m),$$

since the rank-nullity theorem implies that

$$\dim \operatorname{range}(\Gamma) = \dim(U_1 \times ... \times U_m) - \dim \operatorname{null}(\Gamma).$$

We also see that those expressions are equal if and only if $\dim \operatorname{null}(\Gamma) = 0$; or equivalently, if Γ is injective; or equivalently, if $U_1, ..., U_m$ form a direct sum.

Affine subsets

Definition 22 (LADR 3.79, 3.81)**.** Let V be a vector space. A subset $A \subseteq V$ is called **affine** if it can be written in the form

$$A = v + U = \{v + u : u \in U\}$$

for some vector $v \in V$ and subspace $U \subseteq V$.

Example 35. In $\mathbb{R}^2$, the line $\ell = \{(x, y) : x + y = 1\}$ is an affine subset. It has the form

$$\ell = (1, 0) + \operatorname{Span}((-1, 1)).$$

Definition 23 (LADR 3.81)**.** Let $U \subseteq V$ be a vector space. An affine subset $A \subseteq V$ is **parallel** to U if it has the form $A = v + U$ for some $v \in V$.

More generally, two affine subsets are called **parallel** if they are parallel to the same subspace of V.

Example 36. In calculus, taking indefinite integrals results in expressions of the form

$$\int x^2 \, \mathrm{d}x = \frac{1}{3}x^3 + C.$$

The algebraic point of view is that $\int x^2 \, \mathrm{d}x$ is an affine subset of the space of polynomials (or whatever function space we are working with) that is parallel to the space of constants $\mathbb{R}$:

$$\int x^2 \, \mathrm{d}x = \frac{1}{3}x^3 + \mathbb{R}.$$

Proposition 35 (LADR 3.E.7)**.** *An affine subset $A \subseteq V$ is parallel to only one subspace of V.*

Proof. Assume that $A = v + U = x + W$ with $v, x \in V$ and subspaces $U, W \subseteq V$. In problem 8 on the current problem set, we will show that $U = W$. □

Proposition 36 (LADR 3.85)**.** *Let $U \subseteq V$ be a subspace and $v, w \in V$. The following are equivalent:*
(i) $v - w \in U$;
(ii) $v + U = w + U$;
(iii) $(v + U) \cap (w + U) \neq \emptyset$.

Proof. (i) $\Rightarrow$ (ii): Let $u \in U$ be arbitrary; then

$$v + u = w + \underbrace{(v - w) + u}_{\in U} \in w + U.$$

Therefore, $v + U \subseteq w + U$. The same argument shows that $w + U \subseteq v + U$.
(ii) $\Rightarrow$ (iii): Since $0 \in U$, it follows that $v = v + 0 \in w + U$ and therefore $v \in (v + U) \cap (w + U)$.
(iii) $\Rightarrow$ (i): Choose $u_1, u_2 \in U$ such that

$$v + u_1 = w + u_2 \in (v + U) \cap (w + U).$$

Then $v - w = u_2 - u_1 \in U$. □

Definition 24. Let $U \subseteq V$ be a subspace of a vector space V. The **quotient space** V/U is the set of all affine subsets of V that are parallel to U.

The vector space operations are defined by

$$(v+U)+(w+U) := (v+w)+U \text{ and } \lambda \cdot (v+U) := (\lambda v)+U.$$

It is not immediately clear that these definitions are valid, because we aren't allowed to extract the vector v from the affine subset $v+U$. We need to check that this is *well-defined*: that

$$\text{if } v_1+U = v_2+U \text{ and } w_1+U = w_2+U, \text{ then } (v_1+w_1)+U = (v_2+w_2)+U,$$

and similarly for scalar multiplication.

Proof. If $v_1+U = v_2+U$ and $w_1+U = w_2+U$, then $v_1-v_2, w_1-w_2 \in U$. Therefore,

$$(v_1+w_1)-(v_2+w_2) = (v_1-v_2)+(w_1-w_2) \in U,$$

so $(v_1+w_1)+U = (v_2+w_2)+U$.
Similarly, if $v_1+U = v_2+U$, then for any $\lambda \in \mathbb{F}$,

$$\lambda v_1 - \lambda v_2 = \lambda \cdot (v_1-v_2) \in U,$$

so $\lambda v_1 + U = \lambda v_2 + U$. □

It is common to think of elements of V/U simply as elements of V, but with a different definition of "=": we can no longer distinguish elements that differ by U. For example, if $U = \mathrm{Span}((0,1)) \subseteq \mathbb{R}^2$, then taking the quotient $\mathbb{R}^2/U$ means intuitively that we are not allowed to look at the y-component, so the vectors $(1,2)$ and $(1,4)$ become "equal".

Quotient space II - 7/5

Quotient space cont.

Recall that the quotient space V/U is the set of affine subsets $v+U$ parallel to U. Intuitively, it is just the set V with a more inclusive definition of "=": two vectors $v, w \in V$ are considered the same in V/U if they differ only by an element $v - w \in U$.

Definition 25 (LADR 3.88)**.** Let $U \subseteq V$ be a subspace. The **quotient map** is the surjective linear map

$$\pi : V \longrightarrow V/U, \quad \pi(v) := v + U.$$

The null space of π is U, because

$$v \in \mathrm{null}(\pi) \;\Leftrightarrow\; \pi(v) = v + U = 0 + U \;\Leftrightarrow\; v = v - 0 \in U.$$

Proposition 37 (LADR 3.89)**.** *Let V be finite-dimensional and let $U \subseteq V$ be a subspace. Then*

$$\dim(V/U) = \dim(V) - \dim(U).$$

Proof. The canonical projection $\pi : V \to V/U$ is a surjective map with null space U. By the rank-nullity theorem,

$$\dim(V) = \dim \mathrm{null}(\pi) + \dim \mathrm{range}(\pi) = \dim(U) + \dim(V/U). \qquad \square$$

When V is infinite-dimensional, it is harder to know what $\dim(V/U)$ is - it can be either finite- or infinite-dimensional.

The following theorem is a cleaner way to formulate the rank-nullity theorem. It is also called the **first isomorphism theorem**. There are analogous statements in many other areas of algebra.

Proposition 38 (LADR 3.91)**.** *Let $T \in \mathcal{L}(V, W)$ be a linear map. Then there is a natural isomorphism $\overline{T} : V/\mathrm{null}(T) \to \mathrm{range}(T)$ such that $T = \overline{T} \circ \pi$.*

In other words, $\overline{T}$ fits into the commutative diagram

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\pi"] & W \\ V/\mathrm{null}(T) \arrow[r, dashed, "\overline{T}"] & \mathrm{range}(T) \arrow[u, "\iota"'] \end{tikzcd}$$

where $\iota : \mathrm{range}(T) \to W$ is the inclusion $\iota(w) = w$.

Proof. Define $\overline{T}$ by $\overline{T}(v + \mathrm{null}(T)) := T(v)$. This is well-defined, because: if $v + \mathrm{null}(T) = w + \mathrm{null}(T)$, then $v - w \in \mathrm{null}(T)$, so $T(v) - T(w) = T(v - w) = 0$. It is clear that $\overline{T}$ is also linear.

It is injective, because: assume that $\overline{T}(v + \mathrm{null}(T)) = 0$. Then $T(v) = 0$, so $v \in \mathrm{null}(T)$, and $v + \mathrm{null}(T) = 0 + \mathrm{null}(T)$ is the zero element of $V/\mathrm{null}(T)$.

It is surjective: any element $w = T(v) \in \mathrm{range}(T)$ also has the form $w = \overline{T}(v + \mathrm{null}(T))$. □

We immediately recover the classical rank-nullity theorem when V is finite-dimensional:

$$\dim(V) - \dim \mathrm{null}(T) = \dim(V/\mathrm{null}(T)) = \dim \mathrm{range}(T).$$

Of course, this is a circular proof, since we used the rank-nullity theorem to calculate the dimension of the quotient $V/\mathrm{null}(T)$ in the first place.

Example 37. The differentiation operator $T : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ is surjective, and its null space is the space of constants, $\mathbb{R}$. The induced isomorphism is the map

$$\overline{T} : \mathcal{P}(\mathbb{R})/\mathbb{R} \longrightarrow \mathcal{P}(\mathbb{R}), \ \ (p + \mathbb{R}) \mapsto p',$$

and its inverse is the indefinite integral

$$\overline{T}^{-1} : \mathcal{P}(\mathbb{R}) \longrightarrow \mathcal{P}(\mathbb{R})/\mathbb{R}, \ \ p \mapsto \int p(x)\,\mathrm{d}x.$$

Remark: You can think of injective maps and subspaces in the same way: any subspace $U \subseteq V$ comes with its inclusion $i : U \to V$, $v \mapsto v$, and any injective map $i : U \to V$ lets you identify U with the subspace $i(U) \subseteq V$.
In the same manner, you can think of surjective maps and quotient spaces in the same way. Any quotient space V/U comes with its quotient map $\pi : V \to V/U$, and any surjective map $T : V \to W$ lets you identify W with the quotient space $V/\mathrm{null}(T)$.

Proposition 39. *If $T : V \to W$ is a linear map and $U_1 \subseteq V$, $U_2 \subseteq W$ are subspaces such that $T(U_1) \subseteq U_2$, then T induces a linear map*

$$\overline{T} : V/U_1 \longrightarrow W/U_2, \;\; \overline{T}(v + U_1) := T(v) + U_2.$$

LADR 3.91 is the special case that $U_1 = \mathrm{null}(T)$ and $U_2 = \{0\}$, since $W/\{0\} \cong W$.

Proof. It is clear that $\overline{T}$ will be linear. We need to verify that it is well-defined. Assume that

$$v + U_1 = w + U_1 \;\text{ for some } v, w \in V;$$

then $v - w \in U_1$, so $T(v) - T(w) = T(v - w) \in U_2$, and

$$T(v) + U_2 = T(w) + U_2. \qquad \square$$

The "dimension formula" relating the dimension of an intersection and of a sum of subspaces is also a special case of a more general statement about quotient spaces, called the **second isomorphism theorem**:

Proposition 40. *Let $U, W \subseteq V$ be subspaces of a vector space V. Then there is a natural isomorphism*

$$\varphi : U/(U \cap W) \longrightarrow (U + W)/W.$$

When V is finite-dimensional, we recover the formula

$$\dim(U) - \dim(U \cap W) = \dim(U/(U \cap W)) = \dim((U + W)/W) = \dim(U + W) - \dim(W).$$

Before proving this, let's make the statement a little clearer. Students are often not sure whether calling an isomorphism "natural" is just an opinion, or a precise mathematical concept.

Short answer: it's natural because you don't have to choose a basis.

Long(er) answer: Saying that two spaces are naturally isomorphic is more than just knowing that they have the same dimension. A natural isomorphism is a method of constructing an isomorphism that works for every vector space at the same time - here, it should be something like a function

$$\{\text{pairs of subspaces}\} \longrightarrow \{\text{isomorphisms}\}, \quad (U, W) \mapsto \varphi.$$

Since the domain here is not "pairs of vector spaces together with bases", we know in particular that we shouldn't be accessing a basis when we define the isomorphism φ.

It is not so easy to motivate why looking at natural isomorphisms is important when our examples are mostly spaces like $\mathbb{F}^n$ and $\mathcal{P}_n(\mathbb{R})$ that are relatively easy to handle. But when your vector spaces are more complicated, say

$$\mathcal{L}\Big(\mathcal{L}(\mathcal{L}(\mathbb{F}^\infty, \mathbb{F}^\infty), \mathcal{L}(\mathbb{F}^\infty, \mathbb{F})), \mathbb{F}\Big),$$

you become grateful for the linear maps you can find without having to understand what the vectors look like.

Proof. Define a linear map

$$T : U \longrightarrow (U + W)/W, \;\; T(u) := u + W.$$

This is surjective, because: for any $u \in U$ and $w \in W$, the affine space $u + w + W$ is also represented by

$$u + w + W = u + W = T(u)$$

(since $(u + w) - u = w \in W$). The null space is $U \cap W$, because:

$$T(u) = u + W = 0 + W \;\Leftrightarrow\; u = u - 0 \in W \;\Leftrightarrow\; u \in U \cap W.$$

The first isomorphism theorem implies that

$$\varphi := \overline{T} : U/(U \cap W) \longrightarrow (U + W)/W, \;\; \varphi(u + (U \cap W)) := u + W$$

is a well-defined isomorphism. □

Finally, the **third isomorphism theorem** is a statement about quotients of quotient spaces. Assume that $W \subseteq U \subseteq V$ is a chain of subspaces. We can think of U/W as the subspace of V/W consisting of those affine spaces $u + W$ with $u \in U$.

Proposition 41. *There is a natural isomorphism*

$$\varphi : (V/W)/(U/W) \xrightarrow{\sim} V/U.$$

This does not lead to any insightful formula when V is finite-dimensional: but we can verify that

$$\begin{aligned} \dim((V/W)/(U/W)) &= \dim(V/W) - \dim(U/W) \\ &= (\dim(V) - \dim(W)) - (\dim(U) - \dim(W)) \\ &= \dim(V) - \dim(U), \end{aligned}$$

as it should be.

Proof. Since the identity $I : V \to V$ sends W into U, it induces a linear map

$$T : V/W \longrightarrow V/U, \;\; v + W \mapsto v + U.$$

It is surjective: every element of V/U has the form $v + U = T(v + W)$ for some $v \in V$. The null space of T is U/W, because:

$$T(v + W) = v + U = 0 + U \;\Leftrightarrow\; v \in U \;\Leftrightarrow\; v + W \in U/W.$$

The first isomorphism theorem implies that

$$\varphi := \overline{T} : (V/W)/(U/W) \longrightarrow V/U, \quad \varphi\Big((v+W)+U/W\Big) := v+U$$

is an isomorphism. $\square$

The third isomorphism theorem can be useful for certain proofs by induction, since V/W generally has a smaller dimension than V.

Dual space - 7/11

Dual space

Definition 26 (LADR 3.92, 3.94)**.** Let V be a vector space. The **dual space** of V is

$$V' = \mathcal{L}(V, \mathbb{F}) = \{\text{linear maps}\, \varphi : V \to \mathbb{F}\}.$$

There are many names for elements of V': depending on the application, they are called **linear functionals**, or **linear forms**, or **one-forms**, or **covectors**.

Remarks: Despite the notation, this has nothing to do with the derivative. If it isn't already clear from the context, it should be clear from the capitalization of the letter which notation is meant.

The dual space is also often denoted V^* - you will see this in other textbooks (including the Wikipedia article on the dual space). Be careful, because we will use the asterisk for a different concept later on.

The different names for elements of V' are similar to the different names for linear functions / maps / transformations - there is no mathematical difference, but the word choice does imply a particular way of thinking about the same object. "One-forms" usually appear in geometry and topology; "linear functionals" usually appear in advanced calculus; etc.

When V is finite-dimensional, we know that V' is isomorphic to V because

$$\dim V' = \dim(V) \cdot \dim(\mathbb{F}) = \dim(V) \cdot 1 = \dim(V).$$

However, this isomorphism is **not natural**. In fact, if V is infinite-dimensional, then V can never be isomorphic to V' (although this is not so easy to prove). Any isomorphism between finite-dimensional V and V' must involve knowledge of what the elements of V look like.

Definition 27 (LADR 3.96)**.** Let V be a finite-dimensional vector space with ordered basis $\{v_1, ..., v_n\}$. The **dual basis** of V' is the list $\varphi_1, ..., \varphi_n \in V'$, defined by

$$\varphi_j(v_k) = \delta_{jk} = \begin{cases} 1 : & j = k; \\ 0 : & j \neq k. \end{cases}.$$

Example 38. Consider $\mathbb{F}^n$ as the space of column vectors with the canonical basis

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, ..., e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

The dual space $(\mathbb{F}^n)' = \mathcal{L}(\mathbb{F}^n, \mathbb{F})$ is the space of $(1 \times n)$-matrices (i.e. row vectors), and the dual basis to $\{e_1, ..., e_n\}$ is the list of row vectors

$$\varphi_1 = \begin{pmatrix} 1 & 0 & ... & 0 \end{pmatrix}, \; ..., \; \varphi_n = \begin{pmatrix} 0 & ... & 0 & 1 \end{pmatrix}.$$

Proposition 42 (LADR 3.98)**.** *Let V be a finite-dimensional vector space with ordered basis $\{v_1, ..., v_n\}$. Then the dual basis $\{\varphi_1, ..., \varphi_n\}$ is a basis of V'.*

Proof. Since $\dim(V) = \dim(V') = n$, it is enough to show that $\varphi_1, ..., \varphi_n$ is linearly independent. Assume that

$$\lambda_1 \varphi_1 + ... + \lambda_n \varphi_n = 0 \text{ with } \lambda_1, ..., \lambda_n \in \mathbb{F}.$$

Plugging in v_k shows that

$$0 = \Big(\lambda_1 \varphi_1 + ... + \lambda_n \varphi_n\Big)(v_k) = \lambda_1 \cdot 0 + ... + \lambda_k \cdot 1 + ... + \lambda_n \cdot 0 = \lambda_k$$

for every k. □

The name "dual" space is appropriate because, although finite-dimensional spaces V are not naturally isomorphic to their dual V, they are naturally isomorphic to their **bidual** space (dual of the dual space) $V'' = (V')'$. In the case of column vectors $\mathbb{F}^n$, passing to the dual space is the same as taking the matrix transpose, and transposing twice leaves us with the column vector we started with. The identification $V \cong V''$ is an abstraction of that idea.

Proposition 43 (LADR 3.F.34)**.** *Let V be a finite-dimensional vector space. Then there is a natural isomorphism*

$$j : V \longrightarrow V'', \;\; j(v)(\varphi) := \varphi(v), \;\; v \in V, \; \varphi \in V'.$$

When V is infinite-dimensional, j is only an injective map.

Proof. Since $\dim(V) = \dim(V') = \dim(V'')$, it is enough to show that j is injective. Let $v \neq 0$ be any nonzero vector; then v can be extended to a basis $\{v = v_1, v_2, ..., v_n\}$ of V. Let $\{\varphi_1, ..., \varphi_n\}$ be the dual basis; then φ_1 is a linear form such that $\varphi_1(v) \neq 0$, and therefore

$$j(v)(\varphi_1) = \varphi_1(v) \neq 0;$$

i.e. $j(v) \neq 0$. Since v was arbitrary, $\mathrm{null}(j) = \{0\}$. $\square$

Dual map

Definition 28. Let $T : V \to W$ be a linear map. The **dual map** is

$$T' : W' \longrightarrow V', \;\; T'(\varphi) := \varphi \circ T.$$

Example 39. Let $T : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ be the map that multiplies by x, and define the linear form

$$\varphi : \mathcal{P}(\mathbb{R}) \to \mathbb{R}, \;\; \varphi(p) := p(2).$$

Then

$$T'(\varphi)(p) = \varphi(T(p)) = \varphi(x \cdot p) = 2 \cdot p(2);$$

i.e. $T'(\varphi) = 2\varphi$ is the linear form $p \mapsto 2p(2)$.

Proposition 44 (LADR 3.101)**.** *Let U, V, W be vector spaces. The dual defines a linear map*

$$' : \mathcal{L}(V, W) \longrightarrow \mathcal{L}(W', V'), \;\; T \mapsto T'.$$

It also reverses products: for any $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$,

$$(ST)' = T'S'.$$

Here is a "fun" thought experiment: since $' : \mathcal{L}(V, W) \to \mathcal{L}(W', V')$ is itself a linear map, we can take its dual. What exactly does its dual map

$$(')' : \mathcal{L}(W', V')' \longrightarrow \mathcal{L}(V, W)'$$

do?

Proof. (i) Let $T_1, T_2 \in \mathcal{L}(V, W)$ be linear maps and let $\lambda \in \mathbb{F}$ be a scalar. For any $\varphi \in W'$ and $v \in V$,

$$\begin{aligned}\Big((\lambda T_1 + T_2)'(\varphi)\Big)(v) &= \Big(\varphi \circ (\lambda T_1 + T_2)\Big)(v) \\ &= \varphi\Big(\lambda T_1(v) + T_2(v)\Big) \\ &= \lambda\varphi(T_1(v)) + \varphi(T_2(v)) \\ &= \Big(\lambda T_1'(\varphi) + T_2'\varphi\Big)(v),\end{aligned}$$

so $(\lambda T_1 + T_2)' = \lambda T_1' + T_2'$.
(ii) For any $\varphi \in W'$,

$$(ST)'\varphi = \varphi \circ (ST) = (\varphi \circ S) \circ T = (S'\varphi) \circ T = T'(S'\varphi),$$

so $(ST)' = T' \circ S'$. □

Proposition 45 (LADR 3.114)**.** *Let $T : V \to W$ be a linear map between finite-dimensional vector spaces. Let $\mathcal{B}$ and $\mathcal{C}$ be ordered bases of V resp. W, and let $\mathcal{B}'$ and $\mathcal{C}'$ be the dual bases of V' and W'. Then*

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{C}'}(T') = \Big(\mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(T)\Big)^T.$$

Here, if $A = (a_{ij})_{i,j}$ is any matrix, $A^T = (a_{ji})_{i,j}$ denotes the transpose.

Proof. Write $\mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(T) = (a_{ij})_{i,j}$; i.e. if

$$\mathcal{B} = \{v_1, ..., v_n\} \text{ and } \mathcal{C} = \{w_1, ..., w_m\},$$

then $T(v_j) = \sum_{i=1}^m a_{ij} w_i$. Denote the dual bases by

$$\mathcal{B} = \{v_1', ..., v_n'\} \text{ and } \mathcal{C}' = \{w_1', ..., w_m'\}.$$

Then

$$T'(w_j')(v_k) = w_j' T(v_k) = w_j'\Big(\sum_{l=1}^m a_{lk} w_l\Big) = a_{jk} = \sum_{i=1}^n a_{ji} v_i'(v_k),$$

so $T'(w_j')$ and $\sum_{i=1}^n a_{ji} v_i'$ agree on the basis $\{v_1, ..., v_n\}$; therefore, $T'(w_j') = \sum_{i=1}^n a_{ji} v_i'$. □

Example 40. Consider the matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} : \mathbb{R}^3 \to \mathbb{R}^2$. The dual map is

$$A' : (\mathbb{R}^2)' \longrightarrow (\mathbb{R}^3)', \quad \begin{pmatrix} x & y \end{pmatrix} \mapsto \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix},$$

and with respect to the standard bases $\{(1,0),(0,1)\}$ and $\{(1,0,0),(0,1,0),(0,0,1)\}$, this is represented by

$$\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

Using this point of view and the fact that $(ST)' = T'S'$ for any linear maps S, T, it follows that $(AB)^T = B^T A^T$ for any matrices A, B.

Annihilator

Definition 29. Let $U \subseteq V$ be a subspace. The **annihilator** of U is

$$U^0 := \{\varphi \in V' : \ \varphi(u) = 0 \text{ for all } u \in U\}.$$

Example 41. The annihilator of the subspace

$$U = \operatorname{Span}\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right) \subseteq \mathbb{Q}^3$$

is $U^0 := \operatorname{Span}\Big((-1,1,-1)\Big)$.

Proposition 46 (LADR 3.107, 3.109)**.** *Let $T : V \to W$ be a linear map. Then:*
(i) $\operatorname{null}(T') = (\operatorname{range}(T))^0$*;*
(ii) $\operatorname{range}(T') = (\operatorname{null}(T))^0$.

Proof. (i) Let $\varphi \in \operatorname{null}(T')$; then, for any $w = T(v) \in \operatorname{range}(T)$,

$$\varphi(w) = \varphi(T(v)) = T'(\varphi)(v) = 0(v) = 0,$$

so $\varphi \in (\operatorname{range}(T))^0$.
On the other hand, let $\varphi \in (\operatorname{range}(T))^0$; then, for any $v \in V$,

$$T'(\varphi)(v) = \varphi(T(v)) = 0,$$

so $T'(\varphi) = 0$ and $\varphi \in \operatorname{null}(T')$.

(ii) Let $\varphi = T'(\psi) \in \operatorname{range}(T')$. Then, for any $v \in \operatorname{null}(T)$,

$$\varphi(v) = T'(\psi)(v) = \psi(Tv) = \psi(0) = 0;$$

therefore, $\varphi \in \mathrm{null}(T)^0$.

On the other hand, let $\varphi \in \mathrm{null}(T)^0$, and define

$$\psi : \mathrm{range}(T) \longrightarrow \mathbb{F}, \;\; \psi(T(v)) := \varphi(v).$$

This is well-defined, because: if $T(v) = T(w)$, then $T(v - w) = 0$, so $v - w \in \mathrm{null}(T)$ and $\varphi(v - w) = 0$. By extending a basis of range(T) to a basis of W, and extending ψ by 0, we can find a linear form $\tilde{\psi} \in W'$ such that $\tilde{\psi}|_{\mathrm{range}(T)} = \psi$. Then

$$T'(\tilde{\psi}) = \tilde{\psi} \circ T = \psi \circ T = \varphi,$$

so $\varphi \in \mathrm{range}(T')$. □

Proposition 47 (LADR 3.108, 3.110)**.** *Let $T : V \to W$ be a linear map. Then:*
(i) T is injective if and only if T' is surjective;
(ii) T is surjective if and only if T' is injective.

Proof. (i) T is injective if and only if $\mathrm{null}(T) = \{0\}$, which is equivalent to $\mathrm{range}(T') = \mathrm{null}(T)^0 = V'$.
(ii) T is surjective if and only if $\mathrm{range}(T) = W$, which is equivalent to $\mathrm{null}(T') = \mathrm{range}(T)^0 = \{0\}$. □

Invariant subspaces - 7/12

Invariant subspaces

Definition 30 (LADR 5.2)**.** Let $T \in \mathcal{L}(V)$ be an operator on a vector space. A subspace $U \subseteq V$ is **invariant** under T if $T(U) \subseteq U$.

In particular, T can be restricted to an operator $T|_U$ on U. It also induces the quotient operator $\overline{T} = T/U$ on V/U.

Example 42. The null space and range of T are invariant subspaces under T; so are $\{0\}$ and V itself.

Example 43. The following matrices are considered as operators on $\mathbb{R}^2$.
(i) The matrix $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ has invariant subspaces $\{0\}$, $\mathrm{Span}(\begin{pmatrix} 1 \\ 0 \end{pmatrix})$, $\mathrm{Span}(\begin{pmatrix} 0 \\ 1 \end{pmatrix})$ and $\mathbb{R}^2$.
(ii) The matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has only the invariant subspaces $\{0\}$ and $\mathbb{R}^2$.
(iii) The matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ has only the invariant subspaces $\{0\}$, $\mathrm{Span}(\begin{pmatrix} 1 \\ 0 \end{pmatrix})$ and $\mathbb{R}^2$.
We will discuss how to find the invariant subspaces later.

Eigenvectors (more precisely, their spans) are the smallest possible invariant subspaces:

Definition 31 (LADR 5.5, 5.7)**.** Let $T \in \mathcal{L}(V)$ be an operator on a vector space. A nonzero vector $v \in V$ is an **eigenvector** of T if $\mathrm{Span}(v)$ is invariant under T.

It follows that $T(v) = \lambda v$ for a unique scalar $\lambda \in \mathbb{F}$, called the **eigenvalue** associated to v.

Definition 32 (LADR 5.36)**.** Let $T \in \mathcal{L}(V)$ be an operator and $\lambda \in \mathbb{F}$. The **eigenspace** of T for λ is

$$E(\lambda, T) = \mathrm{null}(T - \lambda I) = \{0\} \cup \{\text{eigenvectors of } T \text{ with eigenvalue } \lambda\}.$$

It is clear that $E(\lambda, T)$ is an invariant subspace of T: if $v \in E(\lambda, T)$, then $T(v) = \lambda v \in E(\lambda, T)$.

Proposition 48 (LADR 5.21)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional complex vector space. Then T has an eigenvector.*

Both assumptions here are necessary:
(i) The operator $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ on $\mathbb{R}^2$, which acts as counterclockwise rotation by 90 degrees, has no eigenvectors;
(ii) The operator $T : \mathcal{P}(\mathbb{C}) \to \mathcal{P}(\mathbb{C})$, $p \mapsto x \cdot p$ has no eigenvectors, because $T(p)$ never even has the same degree as p for nonzero p.
This theorem is more of a result in calculus, rather than a theorem of linear algebra, since it depends fundamentally on properties of $\mathbb{C}$ (or $\mathbb{R}$). The (relatively short) proof below is adapted from the article [S]. You will probably want to skip it.

[S] Singh, D. *The Spectrum in a Banach Algebra.* The American Mathematical Monthly, vol. 113, no. 8, pp. 756-758.

Proof. Assume that T has no eigenvalues; then $(T - \lambda I)^{-1}$ exists for all $\lambda \in \mathbb{C}$. (In particular, T itself is invertible.) Choose a linear functional $\varphi \in \mathcal{L}(V)'$ such that $\varphi(T^{-1}) \neq 0$; then φ is differentiable, because it is linear. Consider the function

$$F(r) := \int_0^{2\pi} \varphi((T - re^{i\theta} I)^{-1})\, \mathrm{d}\theta, \quad r \in [0, \infty).$$

Then

$$\begin{aligned} irF'(r) &= \int_0^{2\pi} ir\varphi((T - re^{i\theta})^{-2}) e^{i\theta}\, \mathrm{d}\theta \\ &= \int_0^{2\pi} \frac{\partial}{\partial\theta} \Big(\varphi((T - re^{i\theta})^{-1}) \Big)\, \mathrm{d}\theta \\ &= \varphi((T - re^{2\pi i} I)^{-1}) - \varphi((T - re^{0i} I)^{-1}) = 0. \end{aligned}$$

Here we are using the chain rule:

$$\frac{\partial}{\partial r} \varphi((T - re^{i\theta} I)^{-1}) = \varphi((T - re^{i\theta} I)^{-2}) \cdot e^{i\theta}, \quad \frac{\partial}{\partial\theta} \varphi((T - re^{i\theta} I)^{-1}) = \varphi((T - re^{i\theta} I)^{-2}) ire^{i\theta}.$$

It follows that $F'(r) = 0$ everywhere, so $F(r)$ is constant in r. On the other hand, as $r \to \infty$ becomes large, $(T - re^{i\theta} I)^{-1} \approx (-re^{i\theta} I)^{-1} = -\frac{1}{r} e^{-i\theta} I$ tends to 0 (uniformly in θ, since $e^{-i\theta}$ is bounded), so its integral $F(r)$ tends to 0; it follows that

$$0 = \lim_{r \to \infty} F(r) = F(0) = 2\pi \varphi(T^{-1});$$

contradiction. □

Proposition 49 (LADR 5.10)**.** *Let $T \in \mathcal{L}(V)$, and let $v_1, ..., v_m \in V$ be eigenvectors for distinct eigenvalues $\lambda_1, ..., \lambda_m$. Then $\{v_1, ..., v_m\}$ is linearly independent.*

Proof. Assume that T is linearly dependent, and let k be the smallest index such that $v_k \in \mathrm{Span}(v_1, ..., v_{k-1})$. Write

$$v_k = a_1 v_1 + ... + a_{k-1} v_{k-1} \text{ with } a_i \in \mathbb{F}.$$

Then

$$\lambda_k v_k = T(v_k) = T\Big(a_1 v_1 + ... + a_{k-1} v_{k-1}\Big) = \lambda_1 a_1 v_1 + ... + \lambda_{k-1} a_{k-1} v_{k-1}.$$

On the other hand,

$$\lambda_k v_k = \lambda_k \cdot \Big(a_1 v_1 + ... + a_{k-1} v_{k-1}\Big) = \lambda_k a_1 v_1 + ... + \lambda_k a_{k-1} v_{k-1}.$$

Subtracting these equations gives us

$$0 = (\lambda_1 - \lambda_k) a_1 v_1 + ... + (\lambda_{k-1} - \lambda_k) a_{k-1} v_{k-1};$$

here, $\{v_1, ..., v_{k-1}\}$ is linearly independent since k was chosen minimally, and $(\lambda_1 - \lambda_k), ..., (\lambda_{k-1} - \lambda_k)$ are all nonzero, so

$$a_1 = ... = a_{k-1} = 0.$$

This is a contradiction, because it implies that

$$v_k = a_1 v_1 + ... + a_{k-1} v_{k-1} = 0 v_1 + ... + 0 v_{k-1} = 0.$$ □

Example 44. For any $N \in \mathbb{N}$, the functions $\sin(x), \sin(2x), ..., \sin(Nx)$ are linearly independent elements of

$$V = \{f : \mathbb{R} \to \mathbb{R} : \; f \text{ is differentiable infinitely often}\},$$

because they are eigenvectors of the double-differentiation operator $T : V \to V$, $T(f) := f''$ for the distinct eigenvalues $-1, -4, ..., -N^2$.

The following is an immediate corollary:

Proposition 50 (LADR 5.13)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional vector space. Then T has at most $\dim(V)$ distinct eigenvalues.*

Proof. Let $v_1, ..., v_m$ be eigenvectors for distinct eigenvalues $\lambda_1, ..., \lambda_m$. Then $\{v_1, ..., v_m\}$ is linearly independent, so $m \le \dim(V)$. $\square$

Remark for students interested in that sort of thing: this statement also holds for infinite-dimensional vector spaces. For instance, the shift operator

$$T : \mathbb{R}^\infty \longrightarrow \mathbb{R}^\infty, \; (a_1, a_2, a_3, ...) \mapsto (a_2, a_3, a_4, ...)$$

has every real number λ as an eigenvalue, corresponding to the eigenvector $(1, \lambda, \lambda^2, \lambda^3, ...)$. This forces $\dim(\mathbb{R}^\infty)$ to be at least the cardinality of $\mathbb{R}$.

Finally, let's recall how to find eigenvalues and eigenvectors in practice:

Proposition 51. *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional vector space, and let $v_1, ..., v_n$ be an ordered basis of V. Then $\lambda \in \mathbb{F}$ is an eigenvalue of T if and only if it is an eigenvalue of the matrix $\mathcal{M}(T)$, and $v \in V$ is an eigenvector of T for λ if and only if its coordinates $\mathcal{M}(v)$ are an eigenvector of $\mathcal{M}(T)$ for λ.*

Proof. The equation $T(v) = \lambda v$ is equivalent to the equation

$$\mathcal{M}(T)\mathcal{M}(v) = \mathcal{M}(Tv) = \mathcal{M}(\lambda v) = \lambda \mathcal{M}(v).$$ $\square$

The eigenvalues of a square matrix are found by computing the characteristic polynomial and finding its zeros. We will give a basis-free definition of the characteristic polynomial later - but for a square matrix A, it is the determinant $\det(tI - A)$.

Example 45. Consider the operator

$$T : \mathbb{C}^{2,2} \longrightarrow \mathbb{C}^{2,2}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

(i.e. the transpose). With respect to the basis $\{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\}$, it is represented by

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

The eigenvalues are the roots of

$$\det \begin{pmatrix} \lambda - 1 & 0 & 0 & 0 \\ 0 & \lambda & -1 & 0 \\ 0 & -1 & \lambda & 0 \\ 0 & 0 & 0 & \lambda - 1 \end{pmatrix} = (\lambda - 1)^3(\lambda + 1),$$

so they are 1 and -1. To find bases of the corresponding eigenspaces, we look at

$$\operatorname{null}(A-I) = \operatorname{null}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \operatorname{Span}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right)$$

and

$$\operatorname{null}(A+I) = \operatorname{null}\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} = \operatorname{Span}\left(\begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \right).$$

These are coordinates for the following eigenvectors of T:

$$\underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}}_{\text{eigenvalue } 1}, \underbrace{\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}}_{\text{eigenvalue } -1}.$$

Diagonalizability - 7/13

Upper-triangular matrices

Definition 33. A square matrix $A = (a_{ij})_{i,j}$ is **upper-triangular** if $a_{ij} = 0$ for all pairs (i,j) with $i > j$.
The **diagonal** of A is the entries $a_{11}, ..., a_{nn}$. A is a **diagonal matrix** if all entries that are not on the diagonal are 0.

Example 46. The matrix $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 8 & 9 \\ 0 & 0 & 0 & 10 \end{pmatrix}$ is upper-triangular; its diagonal is $1, 5, 8, 10$.

Proposition 52 (LADR 5.26)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional space, and let $\{v_1, ..., v_n\}$ be an ordered basis of V. The following are equivalent:*
(i) The matrix of T with respect to $\{v_1, ..., v_n\}$ is upper-triangular;
(ii) $T(v_j) \in \mathrm{Span}(v_1, ..., v_j)$ for each j;
(iii) $\mathrm{Span}(v_1, ..., v_j)$ is invariant under T for each j.

Proof. (i) $\Rightarrow$ (ii): Let $(a_{ij})_{i,j}$ be the matrix of T. Then

$$T(v_j) = a_{1j}v_1 + ... + a_{nj}v_n = a_{1j}v_1 + ... + a_{jj}v_j + 0v_{j+1} + ... + 0v_n \in \mathrm{Span}(v_1, ..., v_j).$$

(ii) $\Rightarrow$ (iii): Assuming (ii), we see that

$$T(v_i) \in \mathrm{Span}(v_1, ..., v_i) \subseteq \mathrm{Span}(v_1, ..., v_j)$$

for all $i \leq j$; so $T(\mathrm{Span}(v_1, ..., v_j)) \subseteq \mathrm{Span}(v_1, ..., v_j)$.
(iii) $\Rightarrow$ (i): Write $T(v_j) = a_{1j}v_1 + ... + a_{nj}v_n$ for each j. Since $T(v_j) \in \mathrm{Span}(v_1, ..., v_j)$, it follows that $a_{j+1,j}, ..., a_{nj} = 0$; in other words, $a_{ij} = 0$ whenever $i > j$, so the matrix $(a_{ij})_{i,j}$ is upper-triangular. $\square$

Proposition 53 (LADR 5.30, 5.32). *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional space, and assume that T is represented by an upper-triangular matrix with respect to some basis of V. Then the eigenvalues of T are exactly the entries on the diagonal of that matrix.*

Proof. Assume that T is represented by the matrix

$$\mathcal{M}(T) = \begin{pmatrix} \lambda_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_n \end{pmatrix};$$

then, for any $\lambda \in \mathcal{F}$, the matrix of $T - \lambda I$ is

$$\mathcal{M}(T - \lambda I) = \begin{pmatrix} \lambda_1 - \lambda & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_n - \lambda \end{pmatrix}.$$

By induction on n, we prove that this is invertible if and only if all $\lambda \neq \lambda_k$ for all k:
(i) When $n = 1$, this is obvious.
(ii) In general, assume first that $\lambda \neq \lambda_n$. Then the equation

$$\begin{pmatrix} \lambda_1 - \lambda & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_n - \lambda \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

implies $(\lambda_n - \lambda)v_n = 0$ in the lowest row, and therefore $v_n = 0$. The first $(n-1)$ rows give us the equation

$$\begin{pmatrix} \lambda_1 - \lambda & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_{n-1} - \lambda \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_{n-1} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

By the induction assumption, this matrix is injective if and only if $\lambda \neq \lambda_k$ for all $1 \leq k \leq n-1$.
On the other hand, if $\lambda = \lambda_n$, then the matrix

$$\mathcal{M}(T - \lambda I) = \begin{pmatrix} \lambda_1 - \lambda & * & * \\ 0 & \ddots & * \\ 0 & 0 & 0 \end{pmatrix}$$

has its range contained in the proper subspace $\mathrm{Span}(e_1, ..., e_{n-1})$, so it is not surjective. $\square$

Example 47. The derivative $T : \mathcal{P}_3(\mathbb{R}) \to \mathcal{P}_3(\mathbb{R})$ is represented by the matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

so its only eigenvalue is 0.

Proposition 54 (LADR 5.27)**.** *Let V be a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some basis of V.*

Proof. Induction on $\dim(V)$. This is clear when $\dim(V) = 1$.
In general, fix an eigenvector $v \in V$ and eigenvalue λ, and define $U := \mathrm{range}(T - \lambda I)$. Then U is an invariant subspace of V with $\dim(U) < \dim(V)$. By the induction assumption, there is a basis $\{u_1, ..., u_m\}$ of U such that $T|_U$ is represented by an upper-triangular matrix. Extend this to a basis $\{u_1, ..., u_m, v_1, ..., v_n\}$ of V; then

$$Tv_k = \underbrace{(T - \lambda I)v_k}_{\in \mathrm{Span}(u_1,...,u_m)} + \lambda v_k \in \mathrm{Span}(u_1, ..., u_m, v_1, ..., v_k) \text{ for all } k,$$

so the matrix of T with respect to that basis is upper-triangular. □

Diagonalizable maps

Definition 34 (LADR 5.39)**.** Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional vector space. T is **diagonalizable** if there is a basis of V, with respect to which T is represented by a diagonal matrix.

Proposition 55 (LADR 5.41)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional vector space. The following are equivalent:*
(i) T is diagonalizable;
(ii) There is a basis of V consisting of eigenvectors of T;
(iii) $V = \bigoplus_{i=1}^m E(\lambda_i, T)$ is a direct sum of eigenspaces of T;
(iv) $\dim(V) = \sum_{i=1}^m \dim E(\lambda_i, T)$.

The dimension $\dim E(\lambda, T) = \dim \mathrm{null}(T - \lambda I)$ is called the **geometric multiplicity** of λ as an eigenvalue of T. It is 0 whenever λ is not an eigenvalue.

Proof. (i) $\Rightarrow$ (ii): Let $\{v_1, ..., v_n\}$ be a basis of V with respect to which T is represented by the diagonal matrix $\begin{pmatrix} \lambda_1 & ... & 0 \\ 0 & \ddots & 0 \\ 0 & ... & \lambda_n \end{pmatrix}$. Then $T(v_i) = \lambda_i v_i$ for all i, so $\{v_1, ..., v_n\}$ is a basis consisting of eigenvectors of T.
(ii) $\Rightarrow$ (iii): The spaces $E(\lambda_i, T)$, where λ_i are distinct, always form a direct sum: if there were a nonzero element

$$0 \neq v_j \in E(\lambda_j, T) \cap \Big(E(\lambda_1, T) + ... + E(\lambda_{j-1}, T)\Big),$$

then $v_j = v_1 + ... + v_{j-1}$ would be a sum of eigenvectors for distinct, different eigenvalues (where some v_i may be zero and not appear in the sum). This is a contradiction, because eigenvectors to distinct eigenvalues are linearly independent.
The assumption (ii) implies that $E(\lambda_1, v) \oplus + ... + E(\lambda_m, T)$ is all of V, since it contains the basis of eigenvectors $(v_1, ..., v_n)$.
(iii) $\Rightarrow$ (iv): This is the formula for the dimension of a direct sum.
(iv) $\Rightarrow$ (i): Induction on m.
(1) When $m = 1$, i.e. $\dim(V) = \dim E(\lambda_1, T)$, it follows that $V = E(\lambda_1, T)$, so $T = \lambda_1 I$ has a diagonal matrix with respect to *every* basis of V.
(2) The restrictions of T to the invariant subspaces $U = E(\lambda_1, T) + ... + E(\lambda_{m-1}, T)$ and $W = E(\lambda_m, T)$ are diagonalizable (by the induction assumption). Let $\{v_1, ..., v_m\}$ be a basis of U and let $\{w_1, ..., w_k\}$ be a basis of W, with respect to which $T|_U$ and $T|_W$ are represented by diagonal matrices. Since $U \cap W = \{0\}$ (by linear independence of eigenvectors for distinct eigenvalues), $\{v_1, ..., v_m, w_1, ..., w_k\}$ is a basis of V, with respect to which

$$\mathcal{M}(T) = \begin{pmatrix} \mathcal{M}(T|_U) & 0 \\ 0 & \mathcal{M}(T|_W) \end{pmatrix}$$

is diagonal. $\square$

Example 48. Unless $\dim(V) = 0$ or 1, there are always operators that are not diagonalizable. For example, let $\{v_1, ..., v_n\}$ be a basis of V and define

$$T \in \mathcal{L}(V), \;\; T(v_1) = 0, \;\; T(v_k) := v_{k-1} \, (k \geq 2).$$

Then T is represented by the matrix

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

This is upper-triangular with only 0 on the diagonal, so 0 is the only eigenvalue. However, the null space of T is one-dimensional, spanned by v_1; so

$$n = \dim(V) \neq \sum \dim E(\lambda, T) = 1.$$

The following theorem shows that "most" operators are diagonalizable:

Proposition 56 (LADR 5.44)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional space that has $n = \dim(V)$ distinct eigenvalues. Then T is diagonalizable.*

Proof. Choose eigenvectors $v_1, ..., v_n$ for the distinct eigenvalues $\lambda_1, ..., \lambda_n$; then $\{v_1, ..., v_n\}$ is linearly independent. Since $n = \dim(V)$, we know that $\{v_1, ..., v_n\}$ is a basis of V. $\square$

Example 49. Diagonalizability of a matrix depends on the field that matrix is defined over. For example, the matrix $\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$ has distinct eigenvalues $2 \pm \sqrt{5}$, so it is diagonalizable as an operator on $\mathbb{R}^2$; but if we interpret it as an operator on $\mathbb{Q}^2$, then it has no eigenvalues at all and is not diagonalizable.
Similarly, the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is diagonalizable over $\mathbb{C}$ because it has distinct eigenvalues $\pm i$; but it is not diagonalizable over $\mathbb{R}$.

Minimal polynomial - 7/14

Minimal polynomial

Definition 35 (LADR 5.17)**.** Let $T \in \mathcal{L}(V)$ be an operator and let $p \in \mathcal{P}(\mathbb{F})$ be a polynomial. If $p(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_0$, we define

$$p(T) := a_n T^n + a_{n-1} T^{n-1} + ... + a_1 T + a_0 I.$$

When V is finite-dimensional, we know that $\mathcal{L}(V, V)$ is also finite-dimensional with $\dim \mathcal{L}(V, V) = \dim(V)^2$. In particular, the list $\{I, T, T^2, T^3, ...\}$ will eventually become linearly dependent.

Definition 36 (LADR 8.43)**.** Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional space, and let k be the smallest index such that $T^k \in \mathrm{Span}(I, T, T^2, ..., T^{k-1})$. Write

$$T^k = a_{k-1} T^{k-1} + a_{k-2} T^{k-2} + ... + a_1 T + a_0 I.$$

The **minimal polynomial** of T is

$$p(x) := x^k - a_{k-1} x^{k-1} - ... - a_0.$$

Example 50. (i) The minimal polynomial of $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is $x - 1$.

(ii) Let $T = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ as an operator on $\mathbb{C}^2$. It is clear that $\{I, T\}$ is linearly independent; however, we find

$$T^2 = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix} = 4 \cdot \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} + (-3) \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

so the minimal polynomial of T is $x^2 - 4x + 3$.

By construction, the minimal polynomial p of T satisfies $p(T) = 0$. It is minimal in the following sense:

Proposition 57 (LADR 8.46)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional space, and let $q \in \mathcal{P}(\mathbb{F})$ be a polynomial. Then $q(T) = 0$ if and only if q is a polynomial multiple of the minimal polynomial p.*

Proof. If $q = p \cdot r$ is a multiple, then $q(T) = p(T) \cdot r(T) = 0 \cdot r(T) = 0$.
On the other hand, let $q \in \mathcal{P}(\mathbb{F})$ such that $q(T) = 0$. We perform polynomial division with remainder: write

$$q = p \cdot r + s, \quad \text{where } \deg(s) < \deg(p) \text{ or } s = 0.$$

It follows that

$$0 = q(T) = p(T) \cdot r(T) + s(T) = s(T).$$

If $d = \deg(p)$ is the degree of p, then $\{I, T, ..., T^{d-1}\}$ is linearly independent (since d is the smallest index with $T^d \in \mathrm{Span}(I, ..., T^{d-1})$.) The expression $s(T) = 0$ is a linear combination of $\{I, T, ..., T^{d-1}\}$ to zero; it must be the trivial combination, so all coefficients of s are 0. In other words, $q = p \cdot r + 0 = p \cdot r$ is a multiple of p. $\square$

One of the most important properties of the minimal polynomial is that we can read off the eigenvalues of T from its zeros. For example, the minimal polynomial of $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ was $x^2 - 4x + 3 = (x-3)(x-1)$, so the eigenvalues of $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ are 3 and 1.

Proposition 58 (LADR 5.B.10, 8.49)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional space with minimal polynomial p. Then the eigenvalues of T are exactly the zeros of p.*

Proof. Let $\lambda \in \mathbb{F}$ be any eigenvalue of T, with eigenvector $v \in V$. Since $Tv = \lambda v$, we can compute

$$T^2 v = T(Tv) = T(\lambda v) = \lambda^2 v, \ T^3 v = T(T^2 v) = T(\lambda^2 v) = \lambda^3 v, ...$$

and therefore, if $p = x^n + a_{n-1}x^{n-1} + ... + a_0$,

$$0 = p(T)v = T^n v + a_{n-1}T^{n-1}v + ... + a_0 I v = \lambda^n v + a_{n-1}\lambda^{n-1}v + ... + a_0 v = p(\lambda)v,$$

so $p(\lambda) = 0$.

On the other hand, assume that λ is a zero of p, and factor $p(x) = q(x) \cdot (x - \lambda)$. Then

$$0 = p(T) = q(T) \cdot (T - \lambda I).$$

If λ is not an eigenvalue of T, then $T - \lambda I$ is invertible and therefore

$$0 = q(T) \cdot (T - \lambda I) \cdot (T - \lambda I)^{-1} = q(T).$$

This is impossible, because q cannot be a multiple of p due to its lower degree. $\square$

If $\dim(V) = n$, then $\mathcal{L}(V)$ is n^2-dimensional, so it may seem that the minimal polynomial could have a very large degree (all the way to n^2). This is not the case. The following theorem has a short proof at the end of page 263 in [LADR], but it requires quite a bit of background and is only valid when $\mathbb{F} = \mathbb{C}$. Here is a direct argument.

Proposition 59. *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional space with minimal polynomial p. Then* $\deg(p) \leq \dim(V)$.

Proof. Induction on $\dim(V)$. This is clear when $\dim(V) = 0$ or 1, since $\dim(V)^2 = \dim(V)$ in those cases.
In general, let $n = \dim(V)$ and fix an arbitrary nonzero vector $v \in V$. Since $\{v, Tv, ..., T^n v\}$ is a collection of $(n+1)$ vectors, it must be linearly dependent; so there is a polynomial $q \in \mathcal{P}_n(\mathbb{F})$ such that $q(T)v = 0$. The null space $U := \operatorname{null} q(T)$ is an invariant subspace under T, because: if $q(T)v = 0$, then $q(T)Tv = Tq(T)v = 0$.
Case 1: $q(T) = 0$. Then the minimal polynomial of T is a factor of q, so its degree is at most $\deg(q) \leq n$.
Case 2: $q(T) \neq 0$; then $\dim(U) < \dim(V)$. By the induction assumption, the minimal polynomial r of the restriction $T|_U$ has degree at most $\dim(U)$. In particular, we know

$$0 = r(T|_U)u = r(T)u \text{ for all } u \in U.$$

Also, since $U \neq 0$, the space V/U has strictly smaller dimenision than V. By the induction assumption, the minimal polynomial $\overline{p}$ of the quotient operator

$$\overline{T} : V/U \longrightarrow V/U, \;\; \overline{T}(v + U) := T(v) + U$$

has degree less than $\dim(V/U)$. It follows that

$$\overline{p}(T)v + U = \overline{p}(\overline{T})(v + U) = 0 + U$$

for all $v \in V$, so $\overline{p}(T)v \in U$ for all $v \in V$, and therefore

$$r(T)\overline{p}(T)v \in r(T)(U) = 0.$$

Therefore, the minimal polynomial of p is a factor of $r \cdot \overline{p}$, which has degree

$$\deg(r \cdot \overline{p}) = \deg(r) + \deg(\overline{p}) \leq \dim(U) + \dim(V/U) = \dim(V). \qquad \square$$

Example 51. The minimal polynomial can have degree anywhere between 1 and $\dim(V)$. For example, the minimal polynomial of I is always $x-1$, regardless of dimension. On the other hand, the dimension of the operator

$$T = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 \\ 0 & 0 & 0 & ... & 0 \end{pmatrix} \in \mathcal{L}(\mathbb{C}^n)$$

is x^n: there is an obvious pattern in the powers T^k, so we can calculate its minimal polynomial directly.

Finally, knowing the minimal polynomial gives us a powerful test for diagonalizability:

Proposition 60 (LADR 8.C.12)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional space with minimal polynomial p. Then T is diagonalizable if and only if $p = (x-\lambda_1)...(x-\lambda_m)$ splits into distinct linear factors.*

This is notable because, over $\mathbb{C}$, we can test this condition without factoring p - in particular, without ever finding the eigenvalues. From calculus, we know that p has distinct roots if and only if it shares no roots in common with its derivative p'. (The only thing that can go wrong over $\mathbb{R}$ is that the roots of p might not be real.)

Example 52. Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional complex space, such that $T^n = I$ for some number n. Then $T^n - I = 0$, so the minimal polynomial of T is a factor of $x^n - 1$. Here, $x^n - 1$ has no repeated roots because it has no roots in common with its derivative nx^{n-1} (which has no roots other than 0); therefore, T is diagonalizable.

Even when the roots of p' are not as obvious as the example above, it is easy to test by hand whether p and p' share common roots using the Euclidean algorithm (repeated division with remainder).

Proof. (i) Assume that T is diagonalizable, and write $V = E(\lambda_1, T) \oplus ... \oplus E(\lambda_m, T)$. Since $(T - \lambda_i I)$ is 0 on $E(\lambda_i, T)$, it follows that $(T-\lambda_1 I) \cdot ... \cdot (T - \lambda_m I)$ is zero on each $E(\lambda_i, T)$ and therefore it is zero on all of V; so the minimal polynomial of T is $p(x) = (x-\lambda_1) \cdot ... \cdot (x-\lambda_m)$.

(ii) On the other hand, assume that p splits into distinct linear factors. We use induction on $\dim(V)$.
(1) When $\dim(V) = 0$ or 1, every operator is diagonalizable.
(2) Assume that $m > 1$ (if $p(x) = x - \lambda_1$, then $T = \lambda_1 I$ is diagonalizable). Using division with remainder, we find polynomials $q, s \in \mathcal{P}(\mathbb{F})$ with

$$(x-\lambda_2) \cdot ... \cdot (x-\lambda_m) = q(x)(x-\lambda_1) + s(x), \text{ and } \deg(s) < \deg(x-\lambda_1) = 1.$$

In particular, s is constant; it is nonzero, since $s(\lambda_1) = (\lambda_1 - \lambda_2) \cdot ... \cdot (\lambda_1 - \lambda_m) \neq 0$. For any $v \in V$, consider the vector

$$u := \frac{1}{s}(T - \lambda_2 I) \cdot ... \cdot (T - \lambda_m I)v;$$

then $(T - \lambda_1 I)u = \frac{1}{s}p(T) = 0$, so $u \in E(\lambda_1, T)$. Also,

$$\begin{aligned} v - u &= \frac{1}{s}\Big(sv - (T - \lambda_2 I)...(T - \lambda_m I)v\Big) \\ &= -\frac{1}{s}q(T)(T - \lambda_1 I)v \\ &= (T - \lambda_1 I)\Big(-\frac{1}{s}q(T)v\Big) \in \mathrm{range}(T - \lambda_1 I). \end{aligned}$$

So the decomposition $v = u + (v - u)$ for arbitrary v shows that

$$V = \mathrm{null}(T - \lambda_1 I) + \mathrm{range}(T - \lambda_1 I).$$

This must be a direct sum, because the rank-nullity theorem implies that

$$\begin{aligned} \dim\Big(\mathrm{null}(T - \lambda_1 I) \cap \mathrm{range}(T - \lambda_1 I)\Big) &= \dim \mathrm{null}(T - \lambda_1 I) + \dim \mathrm{range}(T - \lambda_1 I) - \dim(V) \\ &= 0. \end{aligned}$$

By induction, the restriction $T|_{\mathrm{range}(T - \lambda_1 I)}$ is diagonalizable (since its minimal polynomial is a factor of p and $\dim \mathrm{range}(T - \lambda_1 I) < \dim(V)$); since

$$V = \mathrm{null}(T - \lambda_1 I) \oplus \mathrm{range}(T - \lambda_1 I),$$

it follows that V also has a basis consisting of eigenvectors of T. $\square$

Example 53. Whenever $T \in \mathcal{L}(V)$ is diagonalizable and $U \subseteq V$ is an invariant subspace, the restriction $T|_U \in \mathcal{L}(U)$ and quotient operator $\overline{T} \in \mathcal{L}(V/U)$ are also diagonalizable because their minimal polynomials are factors of the minimal polynomial of T.

Generalized eigenvectors - 7/18

Generalized eigenvectors

Definition 37 (LADR 8.9, 8.10)**.** Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional vector space.
(i) The **generalized eigenspace** of T for $\lambda \in \mathbb{F}$ is

$$G(\lambda; T) := \bigcup_{k=0}^{\infty} \mathrm{null}\Big((T - \lambda I)^k\Big).$$

(ii) Nonzero vectors $v \in G(\lambda; T)$ are called **generalized eigenvectors** of T for λ.

In other words, $v \in V$ is a generalized eigenvector for λ if and only if $(T - \lambda I)^k v = 0$ for some $k \in \mathbb{N}$.

It is a little suspicious that we are defining the subspace $G(\lambda, T)$ as a union of null spaces - remember that in general, unions are not subspaces at all. The reason that $G(\lambda, T)$ is a vector space will be the following lemma. It is phrased in terms of an operator T: later, we will replace T by $(T - \lambda I)$ so it will apply to $G(\lambda, T)$.

Proposition 61 (LADR 8.2, 8.3)**.** *Let $T \in \mathcal{L}(V)$ be an operator. Then there is a chain of increasing subspaces*

$$\{0\} = \mathrm{null}(T^0) \subseteq \mathrm{null}(T^1) \subseteq \mathrm{null}(T^2) \subseteq \mathrm{null}(T^3) \subseteq ...$$

If $\mathrm{null}(T^m) = \mathrm{null}(T^{m+1})$ *for any m, then the chain stops at m:*

$$\mathrm{null}(T^m) = \mathrm{null}(T^{m+1}) = \mathrm{null}(T^{m+2}) = \mathrm{null}(T^{m+3}) = ...$$

Proof. (i) If $T^k v = 0$ for some k, then $T^{k+1} v = T(T^k v) = T(0) = 0$. This shows that $\mathrm{null}(T^k) \subseteq \mathrm{null}(T^{k+1})$ for every k.
(ii) Assume that $\mathrm{null}(T^m) = \mathrm{null}(T^{m+1})$. Let $v \in \mathrm{null}(T^{m+k+1})$ for some $k \geq 0$; then

$$T^{m+1}(T^k v) = T^{m+k+1} v = 0,$$

so $T^k v \in \text{null}(T^{m+1}) = \text{null}(T^m)$, so $T^{k+m}v = T^m(T^k v) = 0$. This proves the reverse inclusion: $\text{null}(T^{m+k+1}) \subseteq \text{null}(T^{m+k})$. $\square$

Proposition 62 (LADR 8.4, 8.11)**.** *Let $T \in \mathcal{L}(V)$ be an operator and let $n = \dim(V)$. Then $G(\lambda, T) = \text{null}((T - \lambda I)^n)$.*

Proof. It is enough to prove that $\text{null}(T^n) = \bigcup_{k=0}^{\infty} \text{null}(T^k)$; the claim for λ other than 0 follows by replacing T by $T - \lambda I$.

Assume that $\text{null}(T^n) \neq \text{null}(T^{n+1})$. By the previous proposition, the chain

$$0 \subsetneq \text{null}(T^1) \subsetneq \text{null}(T^2) \subsetneq ... \subsetneq \text{null}(T^{n+1})$$

cannot have "=" anywhere: if $\text{null}(T^N)$ is ever $\text{null}(T^{N+1})$, then $\text{null}(T^k)$ stops growing altogether after $k = N$. It follows that

$$0 = \dim(0) < \dim \text{null}(T^1) < \dim \text{null}(T^2) < ... < \dim \text{null}(T^{n+1}) \leq \dim(V) = n.$$

This is impossible, because it implies that $\dim \text{null}(T^1), ..., \dim \text{null}(T^{n+1})$ are $(n+1)$ distinct integers in $\{1, ..., n\}$. $\square$

Example 54. Consider the matrix

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 3 \end{pmatrix} \in \mathbb{Q}^{3,3}.$$

We will find the generalized eigenvectors corresponding to $\lambda = 1$. Here,

$$\text{null}(A - I) = \text{null} \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \end{pmatrix} = \text{Span}\Big(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \Big),$$

and

$$\text{null}((A - I)^2) = \text{null} \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 4 \end{pmatrix} = \text{Span}\Big(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \Big).$$

Finally, since

$$\text{null}((A - I)^3) = \text{null} \begin{pmatrix} 0 & 0 & 6 \\ 0 & 0 & 4 \\ 0 & 0 & 8 \end{pmatrix} = \text{Span}\Big(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \Big),$$

the chain of subspaces $\text{null}((A-I)^k)$ has stopped growing and

$$G(1,A) = \text{Span}\Big(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \Big).$$

Another consequence of this line of reasoning is that there is no need to define "generalized eigenvalues"; any generalized eigenvector of T must correspond to a true eigenvalue.

Proposition 63. *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional vector space. Then $G(\lambda, T) \neq \{0\}$ if and only if λ is an eigenvalue of T.*

Proof. If $\{0\} = \text{null}(T - \lambda I)$, then the chain of subspaces

$$\{0\} \subseteq \text{null}(T-\lambda I) \subseteq \text{null}((T-\lambda I)^2) \subseteq ...$$

has already stopped at the beginning; i.e.

$$\{0\} = \text{null}(T-\lambda I) = \text{null}((T-\lambda I)^2) = ... = \text{null}((T-\lambda I)^{\dim V}) = G(\lambda, T).$$

In other words, we have shown that when λ is not an eigenvalue of T, then $G(\lambda, T) = \{0\}$. □

As with eigenvectors, generalized eigenvectors for distinct eigenvalues are linearly independent:

Proposition 64 (LADR 8.13)**.** *Let $T \in \mathcal{L}(V)$ and let $v_1, ..., v_m$ be generalized eigenvectors for distinct eigenvalues $\lambda_1, ..., \lambda_m$. Then $\{v_1, ..., v_m\}$ is linearly independent.*

Proof. Assume not; let k be minimal such that $\{v_1, ..., v_k\}$ is linearly dependent, and fix a linear combination

$$a_1 v_1 + ... + a_k v_k = 0$$

with $a_k \neq 0$. Choose N to be the largest exponent with

$$w := (T - \lambda_k I)^N v_k \neq 0.$$

Then $(T - \lambda_k I) w = (T - \lambda_k I)^{N+1} v_k = 0$, so w is an eigenvector with eigenvalue λ_k. Therefore, $(T - \lambda I) w = (\lambda_k - \lambda) w$ for all $\lambda \in \mathbb{F}$; in particular,

$$(T - \lambda I)^n w = (\lambda_k - \lambda)^n w.$$

Fix $n = \dim(V)$. Applying $(T - \lambda_1 I)^n \cdot ... \cdot (T - \lambda_{k-1} I)^n (T - \lambda_k I)^N$ to $a_1 v_1 + ... + a_k v_k$ gets rid of $v_1, ..., v_{k-1}$ (since these are generalized eigenvectors for $\lambda_1, ..., \lambda_{k-1}$ and we are left with

$$\begin{aligned} 0 &= (T - \lambda_1 I)^n \cdot ... \cdot (T - \lambda_{k-1} I)^n (T - \lambda_k I)^N a_k v_k \\ &= a_k (T - \lambda_1 I)^n \cdot ... \cdot (T - \lambda_{k-1} I)^n w \\ &= a_k (\lambda - \lambda_1)^n \cdot ... \cdot (\lambda - \lambda_{k-1})^n w, \end{aligned}$$

implying $a_k = 0$; contradiction. □

Over $\mathbb{C}$, every operator is "diagonalizable" by generalized eigenvectors:

Proposition 65 (LADR 8.21, 8.23)**.** *Let V be a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, ..., \lambda_m$ be the distinct eigenvalues of T. Then*

$$V = \bigoplus_{i=1}^m G(\lambda_i, T).$$

In particular, there is a basis of V consisting of generalized eigenvectors of T.

Proof. Induction on $n = \dim(V)$. This is clear when $n = 1$.
In general, fix an eigenvalue λ_1 of T. Then V decomposes as

$$V = \mathrm{null}(T - \lambda_1 I)^n + \mathrm{range}(T - \lambda_1 I)^n = G(\lambda_1, T) \oplus U,$$

with $U = \mathrm{range}(T - \lambda_1 I)^n$, because: by the rank-nullity theorem, it is enough to verify that $G(\lambda_1, T) \cap \mathrm{range}(T - \lambda_1 I)^n = \{0\}$. This is true, because: if $w = (T - \lambda_1 I)^n v \in G(\lambda_1, T)$, then $(T - \lambda_1 I)^n w = (T - \lambda_1 I)^{2n} v = 0$, which shows that $v \in G(\lambda_1, T)$ and $w = 0$.

By induction, there is a basis of U consisting of generalized eigenvectors of the restricted operator $T|_U$; i.e.

$$U = G(\lambda_1, T|_U) \oplus ... \oplus G(\lambda_m, T|_U) = \bigoplus_{i=1}^m G(\lambda_i, T|_U).$$

Here,

$$G(\lambda_1, T|_U) = G(\lambda_1, T) \cap U = \{0\},$$

as we proved in the previous paragraph. If $i \neq 1$, then $G(\lambda_i, T|_U) = G(\lambda_i, T)$ because: let $v \in G(\lambda_i, T)$ and write

$$v = v_1 + u = v_1 + v_2 + ... + v_m, \text{ with } v_1 \in G(\lambda_1, T),\ u \in U,\ v_i \in G(\lambda_i, T|_U).$$

Since generalized eigenvectors for distinct eigenvalues are linearly independent, the equation

$$v_1 + ... + (v_i - v) + ... + v_m = 0$$

implies that all v_k, $k \neq i$ are 0; in particular, $v_1 = 0$ and $v = u \in U$. Therefore,

$$V = G(\lambda_1, T) \oplus U = G(\lambda_1, T) \oplus \Big(G(\lambda_2, T) \oplus ... \oplus G(\lambda_m, T)\Big). \qquad \square$$

Definition 38. The dimension $\dim G(\lambda, T)$ is called the **algebraic multiplicity** $\mu_{alg}(\lambda)$ of λ as an eigenvalue of T.

Compare this to the geometric multiplicity $\mu_{geo}(\lambda) = \dim E(\lambda, T)$. We see that $\mu_{geo}(\lambda) \leq \mu_{alg}(\lambda)$, and T is diagonalizable if and only if $\mu_{geo}(\lambda) = \mu_{alg}(\lambda)$ for all $\lambda \in \mathbb{F}$.

Finally, we define the characteristic polynomial and prove the Cayley-Hamilton theorem. This will only be valid over $\mathbb{C}$.

Proposition 66 (LADR 8.34, 8.37)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional complex vector space. Let $\lambda_1, ..., \lambda_m$ be the distinct eigenvalues of T, with algebraic multiplicites $d_1, ..., d_m$. Define the* ***characteristic polynomial***

$$q(x) := (x - \lambda_1)^{d_1} \cdot ... \cdot (x - \lambda_m)^{d_m}.$$

Then $q(T) = 0$.

Proof. Since $q(T)$ sends every generalized eigenvector of T to 0, and since these generalized eigenvectors form a basis of V, it follows that $q(T) = 0$. $\square$

Remark: The characteristic polynomial of a matrix operator $A \in \mathcal{L}(\mathbb{C}^n)$ is the familiar expression $q(x) = \det(xI - A)$. In practice, this is the only effective way to calculate the characteristic polynomial. If you remember how to work with determinants from Math 54, then you will be able to prove this. (The first step is to represent A by an upper-triangular matrix, so without loss of generality, you will assume that A is upper-triangular.)

Jordan normal form - 7/19

Yesterday, we argued that every complex matrix admits a basis of generalized eigenvectors, so it is "generalized diagonalizable". This is less useful than it sounds. For example, the basis vectors e_1, e_2, e_3 are generalized eigenvectors of the three matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 3 & -1 & -1 \\ 2 & 0 & -1 \\ 2 & -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 3 & -2 & 0 \\ 1 & 1 & -1 \\ 0 & 2 & -1 \end{pmatrix}$$

for the eigenvalue 1, but the right two matrices are not in a form that is useful for computations. Also, it is not obvious at first glance whether the right two matrices are similar (they are not). We will need something better.

Nilpotent operators

Definition 39. Let $N \in \mathcal{L}(V)$ be an operator on a finite-dimensional space V. N is **nilpotent** if $N^k = 0$ for some $k \in \mathbb{N}$.

In other words, $V = G(0, T)$: every nonzero vector is a generalized eigenvector of N for 0.

Proposition 67 (LADR 8.55)**.** *Let $N \in \mathcal{L}(V)$ be a nilpotent operator on a finite-dimensional space. Then N is represented by a matrix $(a_{ij})_{i,j}$ with 0 in every entry except for possible 1s on the* ***superdiagonal****, i.e. the entries $a_{k,k+1}$, $1 \leq k \leq n-1$.*

For example, the matrices

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

are of this form.

Proof. Induction on $n = \dim(V)$. When $n = 1$, N must be 0.
In general, $U := \text{range}(N) \neq V$ (since N is not invertible), and $N|_U$ is also a nilpotent operator. By the induction assumption, $N|_U$ is represented by a superdiagonal matrix of this form.

Note that the basis with respect to which $N|_U$ has this matrix must be of the form

$$\{N^{d_1}v_1, N^{d_1-1}v_1, ..., v_1, N^{d_2}v_2, ..., v_2, ..., N^{d_k}v_k, ..., v_k\}$$

for some vectors $v_1, ..., v_k \in U$, such that $N^{d_1+1}v_1 = ... = N^{d_k+1}v_k = 0$. Choose vectors $u_1, ..., u_k \in V$ with $Nu_j = v_j$. Then

$$\{N^{d_1+1}u_1, N^{d_1}u_1, ..., u_1, N^{d_k+1}u_k, ..., u_k\}$$

is linearly independent, because: applying N to any linear combination

$$\lambda_{1,1}u_1 + ... + \lambda_{1,d_1+1}N^{d_1+1}u_1 + ... + \lambda_{k,1}u_k + ... + \lambda_{k,d_k+1}N^{d_k+1}u_k = 0$$

results in

$$\lambda_{1,1}v_1 + ... + \lambda_{1,d_1}N^{d_1}v_1 + ... + \lambda_{k,1}v_k + ... + \lambda_{k,d_k}N^{d_k}v_k = 0$$

and therefore

$$\lambda_{1,1} = ... = \lambda_{1,d_1} = ... = \lambda_{k,1} = ... = \lambda_{k,d_k} = 0.$$

It follows that

$$\lambda_{1,d_1+1}N^{d_1+1}u_1 + ... + \lambda_{k,d_k+1}N^{d_k+1}u_k = 0;$$

i.e.

$$\lambda_{1,d_1+1}N^{d_1}v_1 + ... + \lambda_{k,d_k+1}N^{d_k}v_k = 0.$$

These vectors are also linearly independent, so $\lambda_{1,d_1+1} = ... = \lambda_{k,d_k+1} = 0$.

Notice that $\{N^{d_1}v_1, ..., N^{d_k}v_k\}$ is a basis of $\text{null}(N) \cap U$ (which you can see by applying N to a combination of the basis vectors of U above and setting it equal to 0). Extending it to a basis $\{N^{d_1}v_1, ..., N^{d_k}v_k, w_1, ..., w_l\}$ of $\text{null}(N)$, it follows that

$$\{N^{d_1+1}u_1, N^{d_1}u_1, ..., u_1, N^{d_k+1}u_k, ..., u_k, w_1, ..., w_l\}$$

is still linearly independent. It spans V, because: if $v \in V$, then

$$Nv = \sum_{k,e} N^e v_k \in U$$

is a linear combination of $v_1, ..., N^{d_k}v_k$, so v differs from $\sum_{k,e} N^e u_k$ by an element of $\text{null}(N)$. $\square$

The matrix we have just constructed is the **Jordan normal form** of N. That was the hardest part of constructing the Jordan normal form of an arbitrary operator.

Jordan normal form

Definition 40. A complex matrix A is in **Jordan normal form** if it consists of **Jordan blocks**

$$A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix},$$

with

$$A_j = \begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix}$$

for some $\lambda_j \in \mathbb{C}$. The λ_j do not have to be distinct!

Here are three examples of matrices in Jordan normal form:

$$\begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

The matrix on the left consists of two Jordan blocks; the matrix in the middle consists of three Jordan blocks; the matrix on the right consists of five Jordan blocks.

Proposition 68 (LADR 8.60)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional complex vector space. Then T is represented by a matrix in Jordan normal form.*

Proof. For any eigenvalue λ_i, the operator $(T - \lambda_i I)$ is nilpotent on the invariant subspace $G(\lambda_i, T)$. Choose a basis $\{v_{\lambda_i,1}, ..., v_{\lambda_i,k_i}\}$ of each $G(\lambda_i, T)$ for which $(T - \lambda_i I)|_{G(\lambda_i,T)}$ is represented in Jordan normal form; then $T|_{G(\lambda_i,T)}$ is also represented in Jordan normal form.

Combining the bases for each i gives a basis of $V = G(\lambda_1, T) \oplus ... \oplus G(\lambda_m, T)$ with respect to which T is represented in Jordan normal form. □

Here is an example of calculating the Jordan normal form.

Example 55. Consider the matrix

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

The only eigenvalue is 1; it turns out that

$$\dim \operatorname{null}(A - I) = 1, \quad \dim \operatorname{null}(A - I)^2 = 2, \quad \dim \operatorname{null}(A - I)^3 = \dim G(1, T) = 3.$$

We start a **Jordan chain** by choosing an element of $\operatorname{null}(A-I)^3$ that is not contained in

$$\operatorname{null}(A - I)^2 = \operatorname{null} \begin{pmatrix} 0 & 0 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \operatorname{Span}\Big(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \Big);$$

for example, $v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ will do. Then we define

$$v_2 = (A - I)v_3 = \begin{pmatrix} 0 & 2 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix}$$

and

$$v_1 = (A - I)v_2 = \begin{pmatrix} 0 & 2 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix}.$$

You can verify that

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 3 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 3 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}.$$

If we had ever found that $\operatorname{null}(A - I)^k$ was not accounted for by the vectors we had found previously, then we would find new vectors in $\operatorname{null}(A - I)^k$ and start Jordan chains at them also. Each Jordan chain corresponds to a single block in the Jordan normal form.

If we are only interested in the Jordan normal form, and not the basis with respect to which T is in Jordan normal form, we only need to look at the dimensions of certain null spaces.

Example 56. Let $T \in \mathcal{L}(\mathbb{C}^8)$ be an operator with eigenvalues 1 and 2, and assume that

$$\dim \operatorname{null}(T - I) = 2, \quad \dim \operatorname{null}(T - I)^2 = 4, \quad \dim \operatorname{null}(T - I)^3 = 4$$

and

$$\dim \operatorname{null}(T-2I)=2,\ \ \dim \operatorname{null}(T-2I)^2=3,\ \ \dim \operatorname{null}(T-2I)^3=4.$$

Each Jordan block contributes exactly one eigenvector (up to scale), so there are 2 Jordan blocks for 1 and 2 Jordan blocks for 2.
The fact that $\operatorname{null}(T-I)^2=\operatorname{null}(T-I)^3$ stops increasing means that both Jordan blocks for 1 become 0 after squaring $(T-I)^2$, so each Jordan block for 1 is (2×2).
On the other hand, since $\dim \operatorname{null}(T-2I)^2=3$, only one of the Jordan blocks for 2 became smaller after squaring $(T-2I)^2$; so one Jordan block was size (1×1) and the other must be (3×3). The Jordan normal form is

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

Here is what happens in general. Try this formula on the previous example.

Proposition 69. *Let $T\in\mathcal{L}(V)$ be an operator on a finite-dimensional complex vector space. For any $\lambda\in\mathbb{F}$ and $k\in\mathbb{N}$, the number of Jordan blocks for λ of size $(k\times k)$ in the Jordan normal form for T is uniquely determined and it is*

$$2\cdot\Big(\dim \operatorname{null}(T-\lambda I)^k\Big)-\dim \operatorname{null}(T-\lambda I)^{k-1}-\dim \operatorname{null}(T-\lambda I)^{k+1}.$$

In particular, the Jordan normal form of T is unique up to reordering of the Jordan blocks.

Proof. Without loss of generality, we may assume that $\lambda=0$ (otherwise, replace $T-\lambda I$ by T) and that 0 is the only eigenvalue of T (since T acts as an invertible operator on all other generalized eigenspaces); so assume that T is nilpotent.

It is enough to prove that $\dim \operatorname{null}(T^k)-\dim \operatorname{null}(T^{k-1})$ is the number of Jordan blocks of size *at least* k: because, assuming that, it follows that the number of Jordan blocks of size exactly k is

$$\begin{aligned}&\Big(\dim \operatorname{null}(T^k)-\dim \operatorname{null}(T^{k-1})\Big)-\Big(\dim \operatorname{null}(T^{k+1})-\dim \operatorname{null}(T^k)\Big)\\ &=2\cdot\Big(\dim \operatorname{null}(T-\lambda I)^k\Big)-\dim \operatorname{null}(T-\lambda I)^{k-1}-\dim \operatorname{null}(T-\lambda I)^{k+1}.\end{aligned}$$

If T is represented by the Jordan normal form

$$\begin{pmatrix} A_{1,1} & & & & & & 0 \\ & \ddots & & & & & \\ & & A_{1,d_1} & & & & \\ & & & \ddots & & & \\ & & & & A_{n,1} & & \\ & & & & & \ddots & \\ 0 & & & & & & A_{n,d_n} \end{pmatrix},$$

where $A_{j,1}, ..., A_{j,d_j}$ are the Jordan blocks of size $(j \times j)$, then

$$\dim \mathrm{null}(T^k) = \sum_{i,j} \dim \mathrm{null}(A_{i,j}^k).$$

The nullity of the powers of a $(j \times j)$-Jordan block follows the pattern

$$\dim \mathrm{null}(A_{j,-}) = 1, \;\; \dim \mathrm{null}(A_{j,-}^2) = \begin{cases} 1: & j = 1; \\ 2: & j \geq 2; \end{cases} \;\; \dim \mathrm{null}(A_{j,-}^3) = \begin{cases} 1: & j = 1; \\ 2: & j = 2; \\ 3: & j \geq 3; \end{cases} \;...$$

and we see that

$$\dim \mathrm{null}(A_{j,-}^k) - \dim \mathrm{null}(A_{j,-}^{k-1}) = \begin{cases} 1: & j \geq k; \\ 0: & j < k; \end{cases}$$

so $\dim \mathrm{null}(T^k) - \dim \mathrm{null}(T^{k-1})$ counts 1 for each Jordan block of size at least k. $\square$

Remark: For calculations, it is useful to know that a Jordan matrix decomposes as $J = D + N$, where D is its diagonal, $N = J - D$ is nilpotent and D and N commute. (You can prove that D and N commute by showing that they commute on each generalized eigenspace: D restricts to a multiple of the identity.) For example, we will calculate the 100-th power

$$J^{100} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}^{100}.$$

Write $J = D + N$ where $D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. Since D and N

commute, the binomial theorem is valid and

$$
\begin{aligned}
J^{100} &= (D+N)^{100} \\
&= D^{100} + 100D^{99}N + \binom{100}{2} D^{98}N^2 + \binom{100}{3} D^{97} \underbrace{N^3}_{=0} + 0 + ... + 0 \\
&= \begin{pmatrix} 1 & 100 & 4950 \\ 0 & 1 & 100 \\ 0 & 0 & 1 \end{pmatrix}.
\end{aligned}
$$

The decomposition $J = D + N$ is also called the (additive) Jordan-Chevalley decomposition. (There is also a "multiplicative" Jordan-Chevalley decomposition, when J is invertible: it is $J = D(I + D^{-1}N) = DU$, where D and U also commute.)

Inner product - 7/20

Inner product and norm

Today, $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. When $\mathbb{F} = \mathbb{R}$, all definitions are the same, but the complex conjugations should be ignored.

Definition 41 (LADR 6.3, 6.7)**.** Let V be a vector space over $\mathbb{F}$. An **inner product** on V is a function $\langle -, - \rangle : V \times V \to \mathbb{F}$ with the following properties:
(i) $\langle -, - \rangle$ is positive definite: $\langle v, v \rangle > 0$ for all $v \neq 0$.
(ii) $\langle -, - \rangle$ is **sesquilinear** (linear in the first entry and conjugate-linear in the second):

$$\langle \lambda u + v, w \rangle = \lambda \langle u, w \rangle + \langle v, w \rangle, \quad \langle u, \lambda v + w \rangle = \overline{\lambda} \langle u, v \rangle + \langle u, w \rangle;$$

(iii) $\langle -, - \rangle$ is **conjugate symmetric**:

$$\langle u, v \rangle = \overline{\langle v, u \rangle}.$$

Example 57. On $\mathbb{C}^n$, the **dot product** is defined by

$$\langle (w_1, ..., w_n), (z_1, ..., z_n) \rangle := w_1 \overline{z_1} + ... + w_n \overline{z_n}.$$

It is straightforward to verify conditions (i)-(iii) above.

We needed to include complex conjugation to make the inner product positive definite. Trying to carry over the definition of inner product on $\mathbb{R}^n$ to $\mathbb{C}^n$ directly would result in equations like

$$\begin{pmatrix} i & 1 \end{pmatrix} \begin{pmatrix} i \\ 1 \end{pmatrix} = i^2 + 1^2 = 0.$$

Example 58. Let V be the space of continuous complex-valued functions on the interval $[0, 1]$. Then

$$\langle f, g \rangle := \int_0^1 f(x) \overline{g(x)} \, \mathrm{d}x$$

defines an inner product, called the L^2-product.
More generally, we can define the L^2-product on any bounded interval $[a, b]$. On unbounded intervals such as $\mathbb{R}$, we should modify V to make sure that the indefinite integral $\int_{-\infty}^{\infty} f(x) \overline{g(x)} \, \mathrm{d}x$ exists for any $f, g \in V$.

Definition 42. Let V be a vector space over $\mathbb{F}$. A **norm** on V is a function $\|\cdot\| : V \to \mathbb{R}$ with the following properties:
(i) $\|\cdot\|$ is positive: $\|v\| > 0$ for all $v \neq 0$;
(ii) $\|\cdot\|$ is homogeneous: $\|\lambda v\| = |\lambda| \cdot \|v\|$ for any $v \in V$ and $\lambda \in \mathbb{F}$, where $|\lambda|$ denotes the absolute value;
(iii) $\|\cdot\|$ satisfies the **triangle inequality**

$$\|u+v\| \leq \|u\| + \|v\|, \quad u, v \in V.$$

$\|\cdot\|$ is an abstraction of the concept of "length" of a vector. Although for general vector spaces V, this has no geometric meaning (what is the "length" of a polynomial?'), we can apply geometric intuition to prove algebraic results in V.

Proposition 70 (LADR 6.10, 6.18)**.** *Let V be a vector space with inner product $\langle -, - \rangle$. Then*

$$\|v\| := \sqrt{\langle v, v \rangle}$$

is a norm on V.

Proving the triangle inequality takes a little work. We will use another important inequality to prove it - the **Cauchy-Schwarz inequality**:

Proposition 71 (LADR 6.15)**.** *Let V be a vector space with inner product $\langle -, - \rangle$ and induced norm $\|\cdot\|$. Then*

$$|\langle u, v \rangle| \leq \|u\| \cdot \|v\|, \quad u, v \in V,$$

and equality holds if and only if $\{u, v\}$ is linearly dependent.

Proof. Case 1: $v = 0$. Then both sides of the inequality are 0.
Case 2: $v \neq 0$. Define the vector

$$z := \frac{\langle u, v \rangle}{\langle v, v \rangle} v - u$$

(geometrically, z is the distance between u and its projection onto the line through v). Then:

$$
\begin{aligned}
0 &\leq \langle z, z\rangle \\
&= \frac{\langle u, v\rangle}{\langle v, v\rangle}\langle v, z\rangle - \langle u, z\rangle \\
&= \frac{\langle u, v\rangle}{\langle v, v\rangle}\left\langle v, \frac{\langle u, v\rangle}{\langle v, v\rangle}v - u\right\rangle - \left\langle u, \frac{\langle u, v\rangle}{\langle v, v\rangle}v - u\right\rangle \\
&= \frac{\langle u, v\rangle \cdot \overline{\langle u, v\rangle}}{\langle v, v\rangle \cdot \overline{\langle v, v\rangle}}\langle v, v\rangle - \frac{\langle u, v\rangle \cdot \langle v, u\rangle}{\langle v, v\rangle} - \frac{\langle u, v\rangle \cdot \overline{\langle u, v\rangle}}{\overline{\langle v, v\rangle}} + \langle u, u\rangle \\
&= \frac{|\langle u, v\rangle|^2}{\langle v, v\rangle} - \frac{|\langle u, v\rangle|^2}{\langle v, v\rangle} - \frac{|\langle u, v\rangle|^2}{\langle v, v\rangle} + \langle u, u\rangle \\
&= -\frac{|\langle u, v\rangle|^2}{\langle v, v\rangle} + \langle u, u\rangle.
\end{aligned}
$$

Rearranging this inequality gives us $|\langle u, v\rangle|^2 \leq \langle u, u\rangle \cdot \langle v, v\rangle$, and the claim follows after taking square roots. Equality occurs only when $z = 0$, in which case u is a multiple of v. $\square$

Example 59. It is difficult to calculate the integral $\int_0^\pi \sqrt{\sin(x)}\,\mathrm{d}x \approx 2.396$ directly. However, we can get a reasonably good bound by letting $f = \sqrt{\sin(x)}$ and $g = 1$ and calculating

$$
\int_0^\pi \sqrt{\sin(x)}\,\mathrm{d}x = \langle f, g\rangle \leq \|f\| \cdot \|g\| = \sqrt{\int_0^\pi |\sin(x)|\,\mathrm{d}x} \cdot \sqrt{\int_0^\pi 1\,\mathrm{d}x} = \sqrt{2\pi} \approx 2.507,
$$

letting $\langle -, -\rangle$ denote the L^2-product on $[0, \pi]$.

Proof. [Proof of Proposition 70] (i) Since $\langle -, -\rangle$ is positive definite, $\|\cdot\|$ is also positive definite.
(ii) For any $\lambda \in \mathbb{F}$ and $v \in V$,

$$
\|\lambda v\|^2 = \langle \lambda v, \lambda v\rangle = \lambda\overline{\lambda}\langle v, v\rangle = |\lambda|^2\|v\|^2,
$$

so $\|\lambda v\| = |\lambda| \cdot \|v\|$.
(iii) For any $u, v \in V$,

$$
\begin{aligned}
\|u + v\|^2 &= \langle u + v, u + v\rangle \\
&= \langle u, u\rangle + \langle v, u\rangle + \langle u, v\rangle + \langle v, v\rangle \\
&= \|u\|^2 + 2\mathrm{Re}\Big[\langle u, v\rangle\Big] + \|v\|^2 \\
&\leq \|u\|^2 + 2|\langle u, v\rangle| + \|v\|^2 \\
&\leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \\
&= (\|u\| + \|v\|)^2,
\end{aligned}
$$

where the last inequality was the Cauchy-Schwarz inequality. Therefore,

$$\|u+v\| \leq \|u\| + \|v\|. \qquad \square$$

Conversely, we can ask when a norm comes from an inner product. It turns out that there are many norms that do not; for example, on $\mathbb{C}^n$, the **maximum norm**

$$\|(z_1, ..., z_n)\|_\infty := \max\{|z_1|, ..., |z_n|\}$$

is a norm but it does not come from a scalar product (you will verify this on the problem set). There is a simple criterion - but proving it would take a little too long. This is known as the **Jordan-von Neumann theorem**.

Proposition 72 (LADR 6.22, 6.A.19, 6.A.20)**.** *Let $\|\cdot\|$ be a norm on a vector space V. The following are equivalent:*
(i) $\|\cdot\|$ is induced by a scalar product $\langle -, - \rangle$ on V;
(ii) The ***parallelogram law*** *holds:*

$$\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2, \;\; u, v \in V.$$

In this case, the scalar product is given by the ***polarization identity****:*

$$\langle u, v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2}{4}$$

if V is a real vector space, and

$$\langle u, v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + i\|u+iv\|^2 - i\|u-iv\|^2}{4}$$

if V is a complex vector space.

Proof. We will only prove the easy direction (i) $\Rightarrow$ (ii). In this case, the squared norms are inner products, and we can verify that

$$\begin{aligned} &\langle u+v, u+v\rangle + \langle u-v, u-v\rangle \\ &= \Big(\langle u,u\rangle + \langle v,u\rangle + \langle u,v\rangle + \langle v,v\rangle\Big) + \Big(\langle u,u\rangle - \langle v,u\rangle - \langle u,v\rangle + \langle v,v\rangle\Big) \\ &= 2\cdot\langle u,u\rangle + 2\cdot\langle v,v\rangle. \end{aligned}$$

In other words, $\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2$. $\square$

Geometrically, the parallelogram law states that the sum of the squares of side lengths of a parallelogram equals the sum of the squares of its diagonal lengths.

Angle

Definition 43 (LADR 6.A.13)**.** Let V be a vector space with inner product $\langle -, - \rangle$. The **angle** $\theta \in [0, \pi]$ between two nonzero vectors $u, v \in V$ is defined by

$$\cos(\theta) = \frac{\langle u, v \rangle}{\|u\| \cdot \|v\|}.$$

Example 60. The angle between the vectors $(1, 0)$ and $(1, 1)$, in the usual sense of an angle between lines in the plane $\mathbb{R}^2$, is

$$\cos(\theta) = \frac{(1,0) \cdot (1,1)}{\|(1,0)\| \cdot \|(1,1)\|} = \frac{1}{1 \cdot \sqrt{2}} = \frac{\sqrt{2}}{2},$$

i.e. $\theta = \frac{\pi}{4}$.

Definition 44. Two vectors $u, v \in V$ are **orthogonal** if $\langle u, v \rangle = 0$.

Orthogonal vectors are also called **perpendicular**. We'll talk about that more tomorrow.

Orthogonality - 7/21

The field $\mathbb{F}$ will always be either $\mathbb{R}$ or $\mathbb{C}$ in this section.

Orthogonal basis

Proposition 73 (LADR 6.25)**.** *Let V be a vector space and let $\{e_1, ..., e_m\}$ be an **orthonormal list** of vectors of V: i.e. $\langle e_i, e_j \rangle = \delta_{ij}$ for all i, j. Then, for any $a_1, ..., a_m \in \mathbb{F}$,*

$$\|a_1 e_1 + ... + a_m e_m\|^2 = |a_1|^2 + ... + |a_m|^2.$$

You may have heard of Parseval's theorem in the study of Fourier series: given a Fourier series $f(x) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x}$,

$$\sum_{n=-\infty}^{\infty} |c_n|^2 = \|f\|^2 = \int_0^1 |f(x)|^2 \, \mathrm{d}x.$$

For example, $f(x) = x$ has the following Fourier series on $[0, 1]$:

$$x = \frac{1}{2} + \frac{1}{2\pi i} \sum_{n \neq 0} \frac{1}{n} e^{2\pi i n x}, \quad 0 < x < 1,$$

and therefore

$$\frac{1}{4} + \frac{1}{4\pi^2} \sum_{n \neq 0} \frac{1}{n^2} = \int_0^1 x^2 \, \mathrm{d}x = \frac{1}{3},$$

and we conclude that $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$. This is an infinite-dimensional case of the theorem above, and the proof is essentially the same.

Proof. Induction on m. When $m = 1$, $\|a_1 e_1\|^2 = |a_1|^2 \langle e_1, e_1 \rangle = |a_1|^2$.
In general, we use the **Pythagorean theorem**: for any orthogonal vectors $u, v \in V$,

$$\|u + v\|^2 = \langle u + v, u + v \rangle = \langle u, u \rangle + \underbrace{\langle v, u \rangle}_{=0} + \underbrace{\langle u, v \rangle}_{=0} + \langle v, v \rangle = \|u\|^2 + \|v\|^2.$$

Since e_m is orthogonal to $a_1e_1 + ... + a_{m-1}e_{m-1}$, it follows that

$$\begin{aligned}\|a_1e_1 + ... + a_me_m\|^2 &= \|a_1e_1 + ... + a_{m-1}e_{m-1}\|^2 + \|a_me_m\|^2 \\ &= |a_1|^2 + ... + |a_{m-1}|^2 + \|a_me_m\|^2 \\ &= |a_1|^2 + ... + |a_m|^2. \end{aligned}$$ □

We immediately get the corollary:

Proposition 74 (LADR 6.26). *Let V be a vector space and let $\{e_1, ..., e_m\}$ be an orthonormal list of vectors. Then $\{e_1, ..., e_m\}$ is linearly independent.*

Proof. If

$$a_1e_1 + ... + a_me_m = 0,$$

then taking the norm squared shows that $0 = |a_1|^2 + ... + |a_m|^2$; so $|a_i|^2 = 0$ for all i; so $a_i = 0$ for all i. □

If V has an orthonormal *basis*, then it is very easy to find the coefficients of $v \in V$:

Proposition 75 (LADR 6.30). *Let $e_1, ..., e_n$ be an orthonormal basis of V. For any $v \in V$,*

$$v = \langle v, e_1\rangle e_1 + ... + \langle v, e_n\rangle e_n.$$

Proof. If we write $v = a_1e_1 + ... + a_ne_n$, then

$$\langle v, e_k\rangle = \langle a_1e_1 + ... + a_ne_n, e_k\rangle = a_1\langle e_1, e_k\rangle + ... + a_n\langle e_n, e_k\rangle = a_k.$$ □

Now we prove that every finite-dimensional space has an orthonormal basis. This proof is algorithmic: the algorithm is called the Gram-Schmidt procedure.

Proposition 76 (LADR 6.31). *Let V be a finite-dimensional vector space with basis $\{v_1, ..., v_n\}$. Define vectors $e_1, ..., e_n$ inductively by*

$$e_1 = \frac{v_1}{\|v_1\|}$$

and

$$e_j = \frac{v_j - \langle v_j, e_1\rangle e_1 - ... - \langle v_j, e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j, e_1\rangle e_1 - ... - \langle v_j, e_{j-1}\rangle e_{j-1}\|}.$$

Then $e_1, ..., e_n$ is an orthonormal basis of V with the property that $\mathrm{Span}(e_1, ..., e_j) = \mathrm{Span}(v_1, ..., v_j)$ for all $1 \le j \le n$.

Proof. We prove by induction on j that $\{e_1, ..., e_j\}$ is orthonormal and $\mathrm{Span}(e_1, ..., e_j) = \mathrm{Span}(v_1, ..., v_j)$. When j reaches n, it follows that $\mathrm{Span}(e_1, ..., e_n) = V$, so $\{e_1, ..., e_n\}$ is an orthonormal basis.
$j = 1$: This is because $\langle e_1, e_1\rangle = \frac{1}{\|v_1\|^2}\langle v_1, v_1\rangle = 1$.
For general j, defining e_j as above, it follows that

$$\begin{aligned}\langle e_j, e_k\rangle &= \frac{1}{\|v_j - \langle v_j, e_1\rangle e_1 - ... - \langle v_j, e_{j-1}\rangle e_{j-1}\|}\langle v_j - \langle v_j, e_1\rangle e_1 - ... - \langle v_j, e_{j-1}\rangle e_{j-1}, e_k\rangle \\ &= \frac{1}{\|v_j - \langle v_j, e_1\rangle e_1 - ... - \langle v_j, e_{j-1}\rangle e_{j-1}\|}\Big(\langle v_j, e_k\rangle - \langle v_j, e_k\rangle\Big) \\ &= 0,\end{aligned}$$

for any $k < j$; and $\langle e_j, e_j\rangle = 1$ was guaranteed by dividing the expression $v_j - ... - \langle v_j, e_{j-1}\rangle e_{j-1}$ by its norm.
Also, it is clear by definition that $e_j \in \mathrm{Span}(e_1, ..., e_{j-1}) + \mathrm{Span}(v_j) = \mathrm{Span}(v_1, ..., v_{j-1}, v_j)$; i.e.

$$\mathrm{Span}(e_1, ..., e_j) \subseteq \mathrm{Span}(v_1, ..., v_j).$$

Since $\{e_1, ..., e_j\}$ and $\{v_1, ..., v_j\}$ are both linearly independent, it follows that both spaces are j-dimensional, so $\mathrm{Span}(e_1, ..., e_j) = \mathrm{Span}(v_1, ..., v_j)$. □

Example 61. We will find an orthonormal basis of $\mathcal{P}_2(\mathbb{C})$, with its L^2 product

$$\langle p, q\rangle = \int_0^1 p(x)\overline{q(x)}\,\mathrm{d}x.$$

Start with the basis $\{1, x, x^2\}$.
(i) $\|1\|^2 = \int_0^1 1^2\,\mathrm{d}x = 1$, so $e_1 = 1$.
(ii) e_2 is the expression $x - \langle x, 1\rangle 1 = x - \int_0^1 x\,\mathrm{d}x = x - \frac{1}{2}$ divided by its norm

$$\|x - 1/2\| = \sqrt{\int_0^1 (x - 1/2)^2\,\mathrm{d}x} = \sqrt{\int_{-1/2}^{1/2} x^2\,\mathrm{d}x} = \sqrt{1/12},$$

i.e. $e_2 = \sqrt{12}x - \sqrt{3} = \sqrt{3}(2x-1)$.
(iii) e_3 is the expression

$$\begin{aligned} &x^2 - \langle x^2, 1\rangle 1 - \langle x^2, \sqrt{3}(2x-1)\rangle \sqrt{3}(2x-1) \\ &= x^2 - \int_0^1 x^2 \, \mathrm{d}x - 3(2x-1)\int_0^1 2x^3 - x^2 \, \mathrm{d}x \\ &= x^2 - \frac{1}{3} - \frac{1}{2}(2x-1) \\ &= x^2 - x + \frac{1}{6} \end{aligned}$$

divided by its norm

$$\|x^2 - x + 1/6\| = \sqrt{\int_0^1 (x^2 - x + 1/6)^2 \, \mathrm{d}x} = \frac{1}{6\sqrt{5}},$$

i.e. $e_3 = 6\sqrt{5}x^2 - 6\sqrt{5}x + \sqrt{5} = \sqrt{5}(6x^2 - 6x + 1)$.

Remark: In general, infinite-dimensional spaces do *not* have orthogonal bases. The reason is that, if $\{e_i\}_{i\in I}$ is an orthogonal basis of V, then every $v \in V$ must be a finite sum $v = a_1e_1 + ... + a_ne_n$ where $e_1, ..., e_n \in \{e_i\}_{i\in I}$, and therefore $\langle v, e_i\rangle = 0$ for all e_i other than $e_1, ..., e_n$. (In particular, all but finitely many.)
In certain sequence spaces or function spaces, where infinite sums can be defined, an infinite combination of the e_i's will typically have nonzero scalar products with infinitely many e_i's, and therefore could not have been a finite sum.

A concrete example of an inner product space with no orthonormal basis is the space

$$\ell^2 = \{(a_1, a_2, a_3, ...) : \sum_{i=1}^{\infty} |a_i|^2 < \infty\}$$

of sequences whose sum of squares converges, with the inner product

$$\langle (a_1, a_2, a_3, ...), (b_1, b_2, b_3, ...)\rangle = \sum_{i=1}^{\infty} a_i \overline{b_i}.$$

Orthogonal complement

Proposition 77 (LADR 6.42, 6.B.17)**.** *Let V be a finite-dimensional vector space with inner product $\langle -, -\rangle$. Then there is a* real-linear *isomorphism*

$$\Gamma : V \longrightarrow V', \quad \Gamma(v)(u) := \langle u, v\rangle.$$

When V is a complex vector space, Γ is not complex-linear: because $\Gamma(iv) = -i\Gamma(v)$ instead of $\Gamma(iv) = i\Gamma(v)$.

Proof. Γ is real-linear, since $\langle -, - \rangle$ is real-linear in both components.
It is injective, because: if $\Gamma(v) = 0$, then $\|v\|^2 = \langle v, v \rangle = \Gamma(v)(v) = 0$.
Since $\dim_{\mathbb{R}}(V) = \dim_{\mathbb{R}}(V')$, it follows that Γ is a real-linear isomorphism. $\square$

In other words, for any functional $\psi \in V'$, there is a unique vector $v \in V$ with $\psi(u) = \langle u, v \rangle$. If we have an orthonormal basis $e_1, ..., e_n$, then it is easy to find v: writing

$$v = a_1 e_1 + ... + a_n e_n,$$

it follows that $\psi(e_i) = \langle e_i, a_1 e_1 + ... + a_n e_n \rangle = \overline{a_i}$; so

$$v = \overline{\psi(e_1)} e_1 + ... + \overline{\psi(e_n)} e_n.$$

Definition 45 (LADR 6.45)**.** Let $U \subseteq V$ be a subspace. The **orthogonal complement** of U is the set

$$U^\perp = \{v \in V : \langle v, u \rangle = 0 \text{ for all } u \in U\}.$$

$U^\perp$ itself is a vector subspace: for any $v, w \in U^\perp$ and $\lambda \in \mathbb{F}$, and any $u \in U$,

$$\langle \lambda v + w, u \rangle = \lambda \langle v, u \rangle + \langle w, u \rangle = \lambda \cdot 0 + 0 = 0.$$

When V is finite-dimensional, the identification $\Gamma : V \xrightarrow{\sim} V'$ identifies $U^\perp$ with U^0, since the functional $\langle -, v \rangle$ annihilates U if and only if $v \in U^\perp$.

Proposition 78 (LADR 6.47)**.** *Let $U \subseteq V$ be a finite-dimensional subspace. Then $V = U \oplus U^\perp$.*

Proof. (i) The intersection $U \cap U^\perp$ is $\{0\}$, because: if $u \in U \cap U^\perp$, then u is orthogonal to itself, so

$$\|u\|^2 = \langle u, u \rangle = 0,$$

so $u = 0$.
(ii) The sum $U + U^\perp$ is V, because: fix $v \in V$ and fix an orthogonal basis $\{e_1, ..., e_n\}$ of U. Then $v - (\langle v, e_1 \rangle e_1 + ... + \langle v, e_n \rangle e_n)$ is orthogonal to U, since

$$\Big\langle v - (\langle v, e_1 \rangle e_1 + ... + \langle v, e_n \rangle e_n), e_k \Big\rangle = \langle v, e_k \rangle - \langle v, e_k \rangle = 0$$

for all k, and

$$v = \Big(v - (\langle v, e_1 \rangle e_1 + ... + \langle v, e_n \rangle e_n)\Big) + (\langle v, e_1 \rangle e_1 + ... + \langle v, e_n \rangle e_n) \in U^\perp + U. \quad \square$$

In particular, when V is finite-dimensional, it follows that

$$\dim(U^{\perp}) = \dim(U^0) = \dim(V) - \dim(U).$$

Proposition 79 (LADR 6.50)**.** *Let $U \subseteq V$ be a finite-dimensional subspace. Then $U = (U^{\perp})^{\perp}$.*

Proof. (i) Any element $u \in U$ is orthogonal to any element in $U^{\perp}$ by definition of $U^{\perp}$; so $u \in (U^{\perp})^{\perp}$.
(ii) Let $v \in (U^{\perp})^{\perp}$, and write $v = u + w$ with $u \in U$ and $w \in U^{\perp}$. Then $w = v - u \in (U^{\perp})^{\perp}$, so w is orthogonal to itself; therefore, $w = 0$ and $v = u \in U$. □

Definition 46 (LADR 6.53)**.** Let $U \subseteq V$ be a finite-dimensional subspace (so $V = U \oplus U^{\perp}$). The **orthogonal projection** onto U is the map

$$P_U : V \longrightarrow V, \;\; P_U(u+w) = u, \;\; u \in U, \; w \in U^{\perp}.$$

P_U is also called the projection onto U along $U^{\perp}$. It is the unique **projector** (an operator P such that $P^2 = P$) with $\mathrm{range}(P) = U$ and $\mathrm{null}(P) = U^{\perp}$.

If we have an orthonormal basis of U, then it is easier to calculate the orthogonal projection onto U:

Proposition 80 (LADR 6.55 (i))**.** *Let U be a finite-dimensional subspace of V. Let $e_1, ..., e_m$ be an orthonormal basis of U. Then*

$$P_U(v) = \langle v, e_1 \rangle e_1 + ... + \langle v, e_m \rangle e_m.$$

Proof. Since $V = U \oplus U^{\perp}$, we can write $v \in V$ uniquely in the form $v = u + w$ with $u \in U$ and $w \in U^{\perp}$. Since

$$\langle v, e_k \rangle = \langle u + w, e_k \rangle = \langle u, e_k \rangle + \langle w, e_k \rangle = \langle u, e_k \rangle + 0$$

for all k, it follows that

$$P_U(v) = u = \sum_{k=1}^{m} \langle u, e_k \rangle e_k = \sum_{k=1}^{m} \langle v, e_k \rangle e_k.$$ □

Example 62. We will calculate the orthogonal projection of $x^3 \in \mathcal{P}_3(\mathbb{R})$ onto $U := \mathrm{Span}(1, x)$ with respect to the inner product $\langle p, q \rangle = \int_0^1 p(x)q(x)\, \mathrm{d}x$. We saw earlier that $\{1, 2\sqrt{3}x - \sqrt{3}\}$

is an orthonormal basis of U. Therefore,

$$\begin{aligned}
P_U(x^3) &= \langle x^3, 1\rangle + \langle x^3, 2\sqrt{3}x - \sqrt{3}\rangle(2\sqrt{3}x - \sqrt{3}) \\
&= \frac{1}{4} + \frac{3\sqrt{3}}{20}(2\sqrt{3}x - \sqrt{3}) \\
&= \frac{9}{10}x - \frac{1}{5}.
\end{aligned}$$

You can verify that

$$\left\langle \frac{9}{10}x - \frac{1}{5},\, x^3 - \left(\frac{9}{10}x - \frac{1}{5}\right)\right\rangle = 0.$$

Self-adjoint and normal operators - 7/25

Adjoint

In the previous lecture, we showed that an inner product on a finite-dimensional space V almost allows you to identify V with its dual V' via the *real*-linear isomorphism

$$\Gamma_V : V \longrightarrow V', \;\; \Gamma(v) := \langle -, v \rangle$$

(although multiplication by i did not work out correctly); and under this identification, the orthogonal complement of a subspace corresponds to its annihilator. Today, we will study what happens to the dual map.

Definition 47. Let $T \in \mathcal{L}(V, W)$ be a linear map between finite-dimensional inner product spaces. The **adjoint** of T is the map

$$T^* : W \longrightarrow V$$

defined by

$$\langle Tv, w \rangle = \langle v, T^* w \rangle, \;\; v \in V, \; w \in W.$$

Other texts use $T^\dagger$ ("T dagger") to denote the adjoint.

We are requiring $\langle -, T^* w \rangle = \langle -, w \rangle \circ T = T'(\langle -, w \rangle)$; in other words, the diagram

$$\begin{tikzcd} V \arrow[d, "\Gamma_V"] & W \arrow[l, "T^*"'] \arrow[d, "\Gamma_W"] \\ V' & W' \arrow[l, "T'"'] \end{tikzcd}$$

commutes. In particular, T^* exists and is unique: it is $T^* = \Gamma_V^{-1} \circ T' \circ \Gamma_W$. Compare this with LADR 7.A.20.

Example 63. Consider the space $V = \mathcal{P}_1(\mathbb{R})$ with its L^2 inner product

$$\langle f, g \rangle = \int_0^1 f(x) g(x) \, \mathrm{d}x,$$

and the differentiation operator $D \in \mathcal{L}(V)$. The adjoint D^* is defined by

$$\langle 1, D^*(1) \rangle = \langle D(1), 1 \rangle = 0, \quad \langle x, D^*(1) = \langle D(x), 1 \rangle = \langle 1, 1 \rangle = 1,$$

so if $D^*(1) = ax + b$, then

$$\langle 1, ax + b \rangle = \frac{1}{2}a + b = 0 \text{ and } \langle x, ax + b \rangle = \frac{1}{3}a + \frac{1}{2}b = 1,$$

so $D^*(1) = 12x - 6$. Also,

$$\langle 1, D^*(x) \rangle = \langle D(1), x \rangle = 0, \quad \langle x, D^*(x) \rangle = \langle D(x), x \rangle = \langle 1, x \rangle = \frac{1}{2},$$

so if $D^*(x) = ax + b$, then

$$\langle 1, ax + b \rangle = \frac{1}{2}a + b = 0 \text{ and } \langle x, ax + b \rangle = \frac{1}{3}a + \frac{1}{2}b = \frac{1}{2},$$

so $D^*(x) = 6x - 3$. In other words,

$$D^*(ax + b) = a(6x - 3) + b(12x - 6) = (3a + 6b)(2x - 1).$$

The following theorem proves that the adjoint T^* is also linear, and gives a way to calculate it quickly if we have an orthonormal basis given:

Proposition 81 (LADR 7.5, 7.10)**.** *Let $T \in \mathcal{L}(V, W)$ be a linear map between finite-dimensional inner product spaces, where V has orthonormal basis $e_1, ..., e_n$ and W has orthonormal basis $f_1, ..., f_m$. Let $\mathcal{M}(T) = (a_{ij})_{i,j}$ be the matrix of T. Then T^* is linear and the matrix of T^* is*

$$\mathcal{M}(T^*) = (\overline{a_{ji}})_{i,j};$$

i.e. $\mathcal{M}(T^*)$ is the ***conjugate transpose*** *of $\mathcal{M}(T)$.*

Proof. (i) T^* is linear, because: for any $\lambda \in \mathbb{F}$, $w_1, w_2 \in W$ and $v \in V$,

$$\langle v, T^*(\lambda w_1 + w_2) \rangle = \langle Tv, \lambda w_1 + w_2 \rangle = \overline{\lambda} \langle Tv, w_1 \rangle + \langle Tv, w_2 \rangle = \langle v, \lambda T^* w_1 + T^* w_2 \rangle,$$

i.e. $T^*(\lambda w_1 + w_2) = \lambda T^* w_1 + T^* w_2$.

(ii) Let $(b_{ij})_{i,j}$ denote the matrix of T^*, i.e.

$$T^* f_j = \sum_{i=1}^n b_{ij} e_i.$$

Then

$$\overline{b_{kj}} = \langle e_k, \sum_{i=1}^n b_{ij} e_i \rangle = \langle e_k, T^* f_j \rangle = \langle Te_k, f_j \rangle = \langle \sum_{i=1}^n a_{ik} f_i, f_j \rangle = a_{jk},$$

so $\overline{b_{kj}} = a_{jk}$. In other words, $b_{ij} = \overline{a_{ji}}$. $\square$

Example 64. We will work out the previous example, but with less effort. Letting $\{e_1, e_2\} = \{f_1, f_2\} = \{1, \sqrt{3}(2x-1)\}$ be the ONB from Thursday, we see that D is represented by the matrix $\begin{pmatrix} 0 & 2\sqrt{3} \\ 0 & 0 \end{pmatrix}$. Therefore, D^* is represented by its conjugate transpose $\begin{pmatrix} 0 & 0 \\ 2\sqrt{3} & 0 \end{pmatrix}$, i.e.

$$D^*(1) = 2\sqrt{3}(\sqrt{3}(2x-1)) = 12x - 6$$

and $D^*(\sqrt{3}(2x-1)) = 0$, so

$$D^*(x) = D^*(x - 1/2) + D^*(1/2) = 1/2 D^*(1) = 6x - 3.$$

Proposition 82 (LADR 7.7)**.** *Let $T \in \mathcal{L}(V, W)$ be a linear map between finite-dimensional inner product spaces. Then:*
(i) $\mathrm{null}(T^*) = \mathrm{range}(T)^\perp$*;*
(ii) $\mathrm{range}(T^*) = \mathrm{null}(T)^\perp$.

Proof. (i) This is because

$$w \in \mathrm{null}(T^*) \;\Leftrightarrow\; \langle v, T^*w\rangle = 0 \;\forall\, v \in V \;\Leftrightarrow\; \langle Tv, w\rangle = 0 \;\forall\, v \in V \;\Leftrightarrow\; v \in \mathrm{range}(T)^\perp.$$

(ii) Taking orthogonal complements implies that $\mathrm{null}(T^*)^\perp = \mathrm{range}(T)$. Since this equation is also valid for T^* instead of T, we see that

$$\mathrm{null}(T)^\perp = \mathrm{range}(T^*).$$ □

Self-adjoint operators

Definition 48 (LADR 7.11)**.** Let $(V, \langle -, - \rangle)$ be a finite-dimensional inner product space. An operator $T \in \mathcal{L}(V)$ is **self-adjoint** if $T = T^*$.

When V is a real vector space, self-adjoint operators are also called **symmetric**; when V is a complex vector space, self-adjoint operators are also called **Hermitian**. It is more precise to say that the matrix of a self-adjoint operator with respect to an orthonormal basis of V is Hermitian; that means, it equals its own conjugate transpose.

Example 65. The matrix $\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$ is Hermitian; on the other hand, the matrix $\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$ is not Hermitian.

Proposition 83 (LADR 7.13)**.** *Let $T \in \mathcal{L}(V)$ be a self-adjoint operator. Then every eigenvalue of T is real.*

This is true by definition when V is a real vector space, but it is an interesting statement over $\mathbb{C}$.

Proof. Let $Tv = \lambda v$, where $v \neq 0$ and $\lambda \in \mathbb{F}$. Then

$$\lambda\langle v, v\rangle = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle,$$

so $\lambda = \overline{\lambda}$. □

Of course, the converse is false: for example, the matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not Hermitian (i.e. self-adjoint with respect to the dot product) although it has real eigenvalues. Here is a sort of converse:

Proposition 84 (LADR 7.15)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional complex inner product space. Then T is self-adjoint if and only if $\langle Tv, v\rangle \in \mathbb{R}$ for all $v \in V$.*

Proof. (i) Assume that T is self-adjoint. Then

$$\langle Tv, v\rangle = \langle v, Tv\rangle = \overline{\langle Tv, v\rangle} \text{ for all } v \in V;$$

here, the first equality uses self-adjointness and the second is the conjugate symmetry of $\langle -, -\rangle$.
(ii) Assume that $\langle Tv, v\rangle$ is real for all vectors $v \in V$. Using

$$\langle T(v + w), v + w\rangle - \langle T(v - w), v - w\rangle = 2\langle Tv, w\rangle + 2\langle Tw, v\rangle$$

(since the $\langle Tv, v\rangle$ and $\langle Tw, w\rangle$ in the above expression cancel), we replace w by iw and get

$$\langle T(v+iw), v+iw\rangle - \langle T(v-iw), v-iw\rangle = 2\langle Tv, iw\rangle + 2\langle T(iw), v\rangle = -2i\langle Tv, w\rangle + 2i\langle Tw, v\rangle.$$

Combining these shows that

$$\langle Tv, w\rangle = \frac{1}{4}\Big(\langle T(v+w), v+w\rangle + \langle T(v-w), v-w\rangle + i\langle T(v+iw), v+iw\rangle - i\langle T(v-iw), v-iw\rangle\Big).$$

Swapping v and w and considering that

$$\langle T(v + iw), v + iw\rangle = \langle -iT(v + iw), -i(v + iw)\rangle = \langle T(w - iv), w - iv\rangle$$

and $\langle T(v - iw), v - iw\rangle = \langle T(w + iv), w + iv\rangle$ shows that

$$\langle Tv, w\rangle = \overline{\langle Tw, v\rangle} = \langle v, Tw\rangle, \ v, w \in V.$$ □

In particular, the self-adjoint operators as a subset of $\mathcal{L}(V)$ are a kind of generalization of the real numbers $\mathbb{R} \subseteq \mathcal{L}(\mathbb{C})$.

Proposition 85 (LADR 7.16)**.** *Let $T \in \mathcal{L}(V)$ be a self-adjoint operator such that $\langle Tv, v\rangle = 0$ for all $v \in V$. Then $T = 0$.*

Proof. For any $v, w \in V$, since

$$\langle T(v+w), v+w\rangle - \langle T(v-w), v-w\rangle = 2\langle Tw, v\rangle + 2\langle Tv, w\rangle = 4\langle Tv, w\rangle$$

(using the fact that T is self-adjoint), it follows that

$$\langle Tv, w\rangle = \frac{\langle T(v+w), v+w\rangle - \langle T(v-w), v-w\rangle}{4} = 0$$

for all $w \in V$. Therefore, $Tv = 0$. □

Normal operators

Definition 49. An operator on a finite-dimensional inner product space $T \in \mathcal{L}(V)$ is **normal** if it commutes with its adjoint: $TT^* = T^*T$.

Remark: Over $\mathbb{C}$, any operator T can be decomposed into its "real" and "imaginary" parts:

$$T = \frac{T+T^*}{2} + i\frac{T-T^*}{2i} = R + iS,$$

where R, S are self-adjoint. T is normal if and only if R and S commute. Be careful that R and S are not the real and imaginary parts in the usual sense: for example, the decomposition of the normal matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is

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

Proposition 86 (LADR 7.20)**.** *An operator $T \in \mathcal{L}(V)$ is normal if and only if*

$$\|Tv\| = \|T^*v\|$$

for all $v \in V$.

Proof. The operator $N := T^*T - TT^*$ is self-adjoint, and T is normal if and only if $N = 0$. By LADR 7.16,

$$N = 0 \;\Leftrightarrow\; \langle Nv, v\rangle = 0 \;\forall\, v \in V.$$

Writing

$$\langle Nv, v\rangle = \langle T^*Tv, v\rangle - \langle TT^*v, v\rangle = \langle Tv, Tv\rangle - \langle T^*v, T^*v\rangle = \|Tv\|^2 - \|T^*v\|^2$$

makes it clear that

$$\langle Nv, v\rangle = 0 \;\Leftrightarrow\; \|Tv\| = \|T^*v\|$$

for all $v \in V$. □

Proposition 87. *Let $T \in \mathcal{L}(V)$ be a normal operator. Then*

$$\mathrm{null}(T^*) = \mathrm{null}(T) \;\text{ and }\; \mathrm{range}(T^*) = \mathrm{range}(T).$$

In particular, $\mathrm{null}(T)^\perp = \mathrm{null}(T^*)^\perp = \mathrm{range}(T)$.

Proof. (i) By the previous proposition, $\|T^*v\| = 0$ if and only if $\|Tv\| = 0$.
(ii) This is because $\mathrm{range}(T^*) = \mathrm{null}(T)^\perp = \mathrm{null}(T^*)^\perp = \mathrm{range}(T)$. □

In particular, $v \in V$ is an eigenvector of T for λ if and only if

$$v \in \mathrm{null}(T - \lambda I) = \mathrm{null}(T - \lambda I)^* = \mathrm{null}(T^* - \overline{\lambda} I),$$

i.e. if v is an eigenvector of T^* for $\overline{\lambda}$.

Spectral theorem - 7/26

Real spectral theorem

V will denote a nonzero, finite-dimensional inner product space over $\mathbb{R}$ or $\mathbb{C}$.

The spectral theorem is arguably the most important criterion for diagonalizability: the conditions (self-adjoint resp. normal) are often straightforward to verify, and apply to a large number of operators that occur in "real life". For example, the Hessian matrix of a smooth function is symmetric. Also, all rotations and reflections of $\mathbb{R}^n$ are normal.

Proposition 88 (LADR 7.26, 7.27)**.** *Let $T \in \mathcal{L}(V)$ be a self-adjoint operator. Then T has an eigenvalue.*

Proof. Assume that $\mathbb{F} = \mathbb{R}$.
Let $p(x)$ denote the minimal polynomial of T, and assume that p has no real roots. Then p can be factored in the form

$$p(x) = (x^2 + b_1 x + c_1) \cdot ... \cdot (x^2 + b_m x + c_m),$$

where each $x^2 + b_k x + c_k$ has a pair of complex conjugate roots; in particular, its discriminant $b_k^2 - 4c_k < 0$ is negative. Then $T^2 + b_k T + c_k I$ is an operator with the property that, for any nonzero $v \in V$,

$$\begin{aligned}\langle (T^2 + b_k T + c_k I)v, v\rangle &= \langle T^2 v, v\rangle + b_k \langle Tv, v\rangle + c_k \langle v, v\rangle \\ &= \|Tv\|^2 + b_k \langle Tv, v\rangle + c_k \|v\|^2,\end{aligned}$$

since $\langle T^2 v, v\rangle = \langle Tv, Tv\rangle = \|Tv\|^2$ by self-adjointness. Using the Cauchy-Schwarz inequality,

$$\begin{aligned}\|Tv\|^2 + b_k \langle Tv, v\rangle + c_k \|v\|^2 &\geq \|Tv\|^2 - |b_k| \|Tv\| \|v\| + c_k \|v\|^2 \\ &= \underbrace{\Big(\|Tv\| - \frac{1}{2}|b_k| \|v\|\Big)^2}_{\geq 0} + \underbrace{\Big(c_k - \frac{b_k^2}{4}\Big)}_{>0} \underbrace{\|v\|^2}_{>0} \\ &> 0.\end{aligned}$$

Therefore, $(T^2 + b_k T + c_k I)v$ must have been nonzero, so $T^2 + b_k T + c_k I$ is injective and therefore invertible.

But this implies that $p(T) = (T^2 + b_1 T + c_1 I) \cdot ... \cdot (T^2 + b_m T + c_m I)$ is a product of invertible operators and therefore $p(T)$ is invertible. Contradiction, because $p(T) = 0$. $\square$

Now we prove that self-adjoint operators are *semisimple*: every invariant subspace has an invariant complementary subspace. It turns out that over $\mathbb{C}$, this is equivalent to diagonalizability. It is also the key to the spectral theorem in this situation.

Proposition 89 (LADR 7.28)**.** *Let $T \in \mathcal{L}(V)$ be a self-adjoint operator and let $U \subseteq V$ be an invariant subspace. Then $U^\perp$ is also invariant, and the restrictions $T|_U$ and $T|_{U^\perp}$ are self-adjoint.*

Note that $\langle -, - \rangle$ restricts to a scalar product on U.

Proof. (i) $U^\perp$ is invariant, because: for any $v \in U^\perp$ and $w \in U$,

$$\langle Tv, w \rangle = \langle v, Tw \rangle \in \langle v, U \rangle = \{0\},$$

so Tv is also orthogonal to U.
(ii) $T|_U$ is self-adjoint, because

$$\langle T|_U v, w \rangle = \langle Tv, w \rangle = \langle v, Tw \rangle = \langle v, T|_U w \rangle$$

for any $v, w \in U$. Similarly, $T|_{U^\perp}$ is also self-adjoint. $\square$

Proposition 90 (LADR 7.29)**.** *Let $T \in \mathcal{L}(V)$ be an operator.*
(i) Any self-adjoint operator T is ***orthogonally diagonalizable****: there is an orthonormal basis of V consisting of eigenvectors of T.*
(ii) If V is a real vector space, then any orthogonally diagonalizable operator is self-adjoint.

Proof. (i) Induction on $n = \dim(V)$. This is clear when $n = 1$.
In general, fix an eigenvector $v_1 \in V$ and assume without loss of generality that $\|v_1\| = 1$. Define $U := \mathrm{Span}(v_1)$; then U is an invariant subspace, so $U^\perp$ is also invariant. By induction, $T|_{U^\perp}$ is orthogonally diagonalizable: there is an orthonormal basis $\{v_2, ..., v_n\}$ of $U^\perp$ consisting of eigenvectors of $T|_{U^\perp}$. Then $v_1, ..., v_n$ is an orthonormal basis of V consisting of eigenvectors of T.

(ii) Let $v_1, ..., v_n$ be an orthonormal basis of V consisting of eigenvectors of T, and let $\lambda_1, ..., \lambda_n$ be the corresponding eigenvalues. Then, for any i, j,

$$\langle Tv_i, v_j \rangle = \langle \lambda_i v_i, v_j \rangle = \lambda_i \delta_{ij} = \lambda_j \delta_{ij} = \langle v_i, \lambda_j v_j \rangle = \langle v_i, Tv_j \rangle.$$

This implies that $\langle Tv, w \rangle = \langle v, Tw \rangle$ for any vectors $v, w \in V$. Here, we are using the fact that λ_j is real in order to know that $\langle v_i, \lambda_j v_j \rangle = \lambda_j \langle v_i, v_j \rangle$. $\square$

Complex spectral theorem

V will denote a nonzero, finite-dimensional inner product space over $\mathbb{C}$.

The following theorem comprises the main part of the complex spectral theorem. We will give several independent proofs to make the theorem extra convincing.

Proposition 91. *Let $T \in \mathcal{L}(V)$ be a normal operator. Then T is diagonalizable.*

Proof. Let $p(x)$ be the minimal polynomial of T, and assume that p has a double root: i.e. $p(x) = (x-\lambda)^2 r(x)$ for some $\lambda \in \mathbb{F}$ and $r \in \mathcal{P}(\mathbb{C})$. Define $q(x) := (x-\lambda)r(x) = \frac{p(x)}{x-\lambda}$. Then p is a factor of q^2, so $q(T)$ is a normal operator with $q(T)^2 = 0$. Using LADR 7.20 (proposition 86 from yesterday), it follows that

$$0 = \|q(T)q(T)v\|^2 = \|q(T)^* q(T)v\|^2 \text{ for all } v \in V,$$

i.e. $q(T)^* q(T) = 0$. Therefore,

$$0 = \langle q(T)^* q(T)v, v \rangle = \langle q(T)v, q(T)v \rangle = \|q(T)v\|^2$$

for all $v \in V$, so $q(T) = 0$. This is a contradiction, because $\deg(q) < \deg(p)$. $\square$

Proof. Consider the decomposition $T = R + iS$, where R, S are self-adjoint and $RS = SR$. By the real spectral theorem, R and S are both diagonalizable; since they commute, the sum $R + iS$ is also diagonalizable.

Explicitly: each eigenspace $E(\lambda, S)$ of S is R-invariant, because $Sv = \lambda v$ implies $S(Rv) = RSv = R(\lambda v) = \lambda Rv$. By diagonalizing $R|_{E(\lambda,S)}$ for each eigenvalue λ of S, we get a basis of V consisting of vectors that are simultaneously eigenvectors for R and for S. Any such vector will also be an eigenvector of $R + iS$. $\square$

Proof. Induction on $n = \dim(V)$. When $n = 1$, this is clear.
In general, let λ be any eigenvalue of T, and define the T-invariant subspace $U = \text{range}(T - \lambda I)$. Then U is also T^*-invariant, because:

$$T^*(T - \lambda I)v = T^*Tv - \lambda T^*v = (T - \lambda I)T^*v \in U$$

for any $v \in V$. Therefore, $T|_U$ and $T^*|_U$ are well-defined and $T|_U$ is normal, with $(T|_U)^* = T^*|_U$. By induction, U admits a basis of eigenvectors of T. Also, since T is normal, the orthogonal complement of U is

$$U^\perp = \text{range}(T - \lambda I)^\perp = \text{null}(T - \lambda I),$$

which admits a basis of eigenvectors of T by definition. Therefore, V admits a basis of eigenvectors of T. □

Proof. Let $v_1, ..., v_k$ be any Jordan chain for T for some eigenvalue $\lambda \in \mathbb{C}$, and assume that $k \geq 2$. In other words, $Tv_j = \lambda v_j + v_{j-1}$ for all j. Then

$$v_1 = (T - \lambda I)v_2,$$

so

$$v_1 \in \text{range}(T - \lambda I) \cap \text{null}(T - \lambda I).$$

Since $T - \lambda I$ is normal, it follows that $\text{null}(T - \lambda I) = \text{range}(T - \lambda I)^\perp$ and therefore $v_1 = 0$; contradiction.
Therefore, the Jordan normal form of T consists only of (1×1)-blocks; i.e. T is diagonalizable. □

Proof. **Schur's theorem** states that, for any operator T, there is an orthonormal basis $\{e_1, ..., e_n\}$ of V, with respect to which T is represented by an upper triangular matrix. This is not hard to prove: if $\{v_1, ..., v_n\}$ is any basis of V such that all $\text{Span}(v_1, ..., v_k)$ are T-invariant, then the Gram-Schmidt process gives us an orthonormal basis $\{e_1, ..., e_n\}$ such that all $\text{Span}(e_1, ..., e_k) = \text{Span}(v_1, ..., v_k)$ are T-invariant.

Assume that

$$\mathcal{M}(T) = \begin{pmatrix} a_{1,1} & ... & a_{1,n} \\ & \ddots & \vdots \\ 0 & & a_{n,n} \end{pmatrix}.$$

Since T is normal,

$$|a_{1,1}|^2 = \|Te_1\|^2 = \|T^*e_1\|^2 = |a_{1,1}|^2 + |a_{1,2}|^2 + ... + |a_{1,n}|^2,$$

so $a_{1,2} = ... = a_{1,n} = 0$. It follows that

$$|a_{2,2}|^2 = |a_{1,2}|^2 + |a_{2,2}|^2 = \|Te_2\|^2 = \|T^* e_2\|^2 = |a_{2,2}|^2 + |a_{2,3}|^2 + ... + |a_{2,n}|^2,$$

so $a_{2,3} = ... = a_{2,n} = 0$.
This argument shows that all nondiagonal entries of $\mathcal{M}(T)$ are 0, so $\{e_1, ..., e_n\}$ is a basis of eigenvectors of T. □

Now we prove the spectral theorem:

Proposition 92 (LADR 7.24)**.** *Let $T \in \mathcal{L}(V)$ be an operator. The following are equivalent:*
(i) T is normal;
(ii) T is orthogonally diagonalizable.

Proof. (i) ⇒ (ii): We have already seen that T is diagonalizable. If v, w are eigenvectors for distinct eigenvalues of T, with $Tv = \lambda v$ and $Tw = \mu w$, then

$$w = (T - \lambda I)(\mu - \lambda)^{-1} w \in \mathrm{range}(T - \lambda I) = \mathrm{null}(T - \lambda I)^{\perp}$$

and therefore $\langle v, w \rangle = 0$.

(ii) ⇒ (i): Let $v_1, ..., v_n$ be an orthonormal basis of eigenvectors of T with eigenvalues $\lambda_1, ..., \lambda_n$. Then

$$\langle T^* T v_i, v_j \rangle = \langle T v_i, T v_j \rangle = \langle \lambda_i v_i, \lambda_j v_j \rangle = \lambda_i \overline{\lambda_j} \delta_{ij} = |\lambda_i|^2 \delta_{ij},$$

and

$$\langle T T^* v_i, v_j \rangle = \langle T^* v_i, T^* v_j \rangle = \langle \overline{\lambda_i} v_i, \overline{\lambda_j} v_j \rangle = \overline{\lambda_i} \lambda_j \delta_{ij} = |\lambda_i|^2 \delta_{ij};$$

these are equal, so $T^* T = T T^*$. □

Positive operators and isometries - 8/1

Positive operators

Definition 50. Let V be a finite-dimensional inner product space. A self-adjoint operator $T \in \mathcal{L}(V)$ is **positive** if

$$\langle Tv, v \rangle \geq 0$$

for all $v \in V$.

Positive operators are often called **positive semidefinite**. This is distinguished from **positive definite** operators, which have the strict inequality: $\langle Tv, v \rangle > 0$ for all $v \neq 0$.

Example 66. The orthogonal projection P_U onto a subspace $U \subseteq V$ is positive semidefinite, because: for any $v \in V$,

$$\langle P_U(v), v \rangle = \langle P_U(v), P_U(v) + \underbrace{v - P_U(v)}_{\in U^\perp} \rangle = \langle P_U(v), P_U(v) \rangle \geq 0.$$

Proposition 93 (LADR 7.35)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional inner product space. The following are equivalent:*
(i) T is positive semidefinite;
(ii) T is self-adjoint, and all eigenvalues of T are nonnegative;
(iii) T is the square of a self-adjoint operator;
(v) There is an operator $R \in \mathcal{L}(V)$ such that $T = R^* R$.

Proof. (i) $\Rightarrow$ (ii): Let v be an eigenvector of T with eigenvalue λ; then $\lambda \|v\|^2 = \langle \lambda v, v \rangle = \langle Tv, v \rangle \geq 0$, so $\lambda \geq 0$.
(ii) $\Rightarrow$ (iii): Let $e_1, ..., e_n$ be an orthonormal basis of eigenvectors of T, with real nonnegative eigenvalues $\lambda_1, ..., \lambda_n$ (which exists by the spectral theorem). Define an operator R by $Re_k = \sqrt{\lambda_k} e_k$. Then R is orthogonally diagonalizable with real eigenvalues, so it is self-adjoint; and $R^2 e_k = \sqrt{\lambda_k}^2 e_k = \lambda_k e_k = Te_k$ for all k, so $R^2 = T$.

(iii) $\Rightarrow$ (iv): If $T = R^2$ where R is self-adjoint, then T is also R^*R.
(iv) $\Rightarrow$ (i): For any $v \in V$,

$$\langle Tv, v\rangle = \langle R^*Rv, v\rangle = \langle Rv, Rv\rangle \geq 0. \qquad \square$$

In particular, knowing that (ii) $\Rightarrow$ (i) holds, we see that the operator R constructed in the proof of (ii) $\Rightarrow$ (iii) is positive semidefinite; so every positive semidefinite operator has a positive semidefinite square root.

There is a similar characterization of positive definite operators:

Proposition 94. *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional inner product space. The following are equivalent:*
(i) T is positive definite;
(ii) T is self-adjoint, and all eigenvalues of T are strictly positive;
(iii) There is an invertible operator $R \in \mathcal{L}(V)$ such that $T = R^*R$.

The proof is almost exactly the same. You should work out the details!

Finally, we will show that the positive square root is unique:

Proposition 95 (LADR 7.36, 7.44)**.** *Let $T \in \mathcal{L}(V)$ be a positive semidefinite operator. Then the positive square root is unique. It is denoted $\sqrt{T}$.*

Example 67. In general, an operator that has a square root (not necessarily positive semidefinite) will have far more than two of them. For example, the square roots of $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ on $\mathbb{C}$ consist of $\pm\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\pm\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, and every matrix of the form

$$\begin{pmatrix} a & b \\ (1-a^2)/b & -a \end{pmatrix}, \; a, b \in \mathbb{C}, \; b \neq 0.$$

Proof. (i) The only positive square root of a positive multiple of the identity λI is $R = \sqrt{\lambda} I$, because: let R be positive semidefinite with $R^2 = \lambda I$. Then R is diagonalizable, and for every eigenvalue μ of R, μ^2 is an eigenvalue of I (i.e. $\mu^2 = \lambda$). Since μ is nonnegative, it must be $\sqrt{\lambda}$.
Since R is diagonalizable with no eigenvalues other than $\sqrt{\lambda}$, its minimal polynomial must be $x - \sqrt{\lambda}$; therefore, $R - \sqrt{\lambda} I = 0$ and $R = \sqrt{\lambda} I$.

(ii) Let T be an arbitrary positive semidefinite operator. On each eigenspace $E(\lambda, T)$, the only positive square root of $T|_{E(\lambda,T)} = \lambda I|_{E(\lambda,T)}$ is $\sqrt{\lambda} I|_{E(\lambda,T)}$. Therefore, $Rv = \sqrt{\lambda} v$ for any eigenvector $v \in E(\lambda, T)$, so R is uniquely determined. $\square$

One of the important applications of positive operators is in classifying the inner products on $\mathbb{R}^n$ (or $\mathbb{C}^n$):

Proposition 96. *Every inner product on $\mathbb{C}^n$ has the form*

$$\langle v, w \rangle = v^T G \overline{w}$$

*for a unique, positive definite (with respect to the dot product) matrix G (called the **Gram matrix** of the inner product). Conversely, if G is a positive definite matrix, then $\langle v, w \rangle = v^T G \overline{w}$ defines a scalar product on $\mathbb{C}^n$.*

Proof. Let $e_1, ..., e_n$ be the standard basis of $\mathbb{C}^n$, and define G by

$$G_{ij} = \langle e_i, e_j \rangle.$$

Then, for any $v = (v_1, ..., v_n)$ and $w = (w_1, ..., w_n)$,

$$\langle v, w \rangle = \sum_{i,j} \langle v_i e_i, w_j e_j \rangle = \sum_{i,j} v_i \overline{w_j} \langle e_i, e_j \rangle = v^T G \overline{w}.$$

The matrix G is positive definite, because:
(i) it is self-adjoint, i.e. Hermitian:

$$G_{ji} = \langle e_j, e_i \rangle = \overline{\langle e_i, e_j \rangle} = \overline{G_{ij}};$$

(ii) for any $v \neq 0$, $(Gv) \cdot v = (Gv)^T \overline{v} = v^T G \overline{v} = \langle v, v \rangle > 0$.

Conversely, if G is positive definite, then $\langle v, w \rangle := v^T G \overline{w}$ is sesquilinear;

$$\langle w, v \rangle = w^T G \overline{v} = (w^T G \overline{v})^T = \overline{v}^T G^T w = \overline{v^T G \overline{w}} = \overline{\langle v, w \rangle},$$

and for any nonzero vector v,

$$\langle v, v \rangle = v^T G \overline{v} = (Gv) \cdot v > 0.$$

□

Isometries

Definition 51 (LADR 7.37)**.** A linear map $S \in \mathcal{L}(V, W)$ between inner product spaces is an **isometry** if $\|Sv\| = \|v\|$ for all $v \in V$.

This differs from the definition in the textbook, which only refers to operators on a single space as isometries.

Isometric operators on real inner product spaces are usually called **orthogonal** operators; isometric operators on complex inner product spaces are usually called **unitary** operators.

Proposition 97 (LADR 7.42 (a),(e)). *Let $S \in \mathcal{L}(V, W)$ be a linear map between finite-dimensional inner product spaces. The following are equivalent:*
(i) S is an isometry;
(ii) $S^* S = I$.

Proof. (i) $\Rightarrow$ (ii): Assume that $\|Sv\| = \|v\|$ for all v. Then

$$\langle S^* S v, v \rangle = \langle Sv, Sv \rangle = \langle v, v \rangle \text{ for all } v \in V,$$

so $S^* S - I$ is a self-adjoint operator with the property that

$$\langle (S^* S - I)v, v \rangle = 0$$

for all $v \in V$. This implies that $S^* S - I = 0$, so $S^* S = I$.
(ii) $\Rightarrow$ (i): If $S^* = S^{-1}$, then

$$\langle Sv, Sv \rangle = \langle S^* S v, v \rangle = \langle v, v \rangle$$

for all $v \in V$; taking square roots shows that $\|Sv\| = \|v\|$ for all $v \in V$. □

In particular, every isometric operator is normal: its adjoint is its inverse.

Remark: The adjoint of any isometric operator is also isometric: this is because $S^* S = I$ implies that $S^* = S^{-1}$, so $(S^*)^* S^* = SS^* = SS^{-1} = I$. On $\mathbb{R}^n$ with the dot product, this reduces to the statement that, if the columns of a square matrix are orthonormal, then the rows of that matrix are also orthonormal.
The adjoint of an isometry between two different spaces does not need to be isometric.

Proposition 98 (LADR 7.42 (b), (c), (d)). *Let $S \in \mathcal{L}(V, W)$ be a linear map. The following are equivalent:*
(i) S is an isometry;
(ii) $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$;
(iii) For any orthonormal basis $e_1, ..., e_n$ of V, $S(e_1), ..., S(e_n)$ is an orthonormal list in W;
(iv) There is an orthonormal basis $e_1, ..., e_n$ of V such that $S(e_1), ..., S(e_n)$ is an orthonormal list in W.

Proof. (i) $\Rightarrow$ (ii) Since $S^* S = I$, it follows that $\langle Su, Sv \rangle = \langle S^* S u, v \rangle = \langle u, v \rangle$ for all $v \in V$.

(ii) $\Rightarrow$ (iii): This is because $\langle S(e_i), S(e_j)\rangle = \langle e_i, e_j\rangle = \delta_{ij}$ for all i, j.
(iii) $\Rightarrow$ (iv): This is clear.
(iv) $\Rightarrow$ (i): For any $v = \lambda_1 e_1 + ... + \lambda_n e_n$, by the generalized Pythagorean theorem,

$$\|Sv\|^2 = \|\lambda_1 S(e_1) + ... + \lambda_n S(e_n)\|^2 = |\lambda_1|^2 + ... + |\lambda_n|^2 = \|\lambda_1 e_1 + ... + \lambda_n e_n\|^2 = \|v\|^2,$$

so $\|Sv\| = \|v\|$. □

Proposition 99 (LADR 7.43). *Let $S \in \mathcal{L}(V)$ be an operator on a finite-dimensional complex inner product space. The following are equivalent:*
(i) S is an isometry;
(ii) S is normal, and the absolute value of any eigenvalue of S is 1.

Proof. (i) $\Rightarrow$ (ii): Every isometry S commutes with its adjoint $S^* = S^{-1}$. Also, if λ is an eigenvalue of S with eigenvector v, then $\|v\| = \|Sv\| = \|\lambda v\| = |\lambda| \cdot \|v\|$, so $|\lambda| = 1$.

(ii) $\Rightarrow$ (i): Since S is normal, if λ is any eigenvalue of S with eigenvector v,

$$S^*Sv = S^*(\lambda v) = \lambda S^* v = \lambda\overline{\lambda} v = |\lambda|^2 v = v.$$

Therefore, S^*S agrees with I on a basis of eigenvectors of S, so $S^*S = I$. □

Polar and singular value decomposition - 8/2

Polar decomposition

Proposition 100 (LADR 7.45)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional inner product space. Then there is an isometry S such that*

$$T = S\sqrt{T^*T}.$$

Here, $\sqrt{T^*T}$ is the positive semidefinite square root of T^*T.

Proof. We define the linear map

$$S_1 : \mathrm{range}(\sqrt{T^*T}) \longrightarrow \mathrm{range}(T), \quad S_1(\sqrt{T^*T}v) := Tv.$$

This is well-defined, because: if $\sqrt{T^*T}v_1 = \sqrt{T^*T}v_2$, then

$$T^*Tv_1 = \sqrt{T^*T}\sqrt{T^*T}v_1 = \sqrt{T^*T}\sqrt{T^*T}v_2 = T^*Tv_2,$$

so

$$\|Tv_1 - Tv_2\|^2 = \langle T(v_1 - v_2), T(v_1 - v_2)\rangle = \langle \underbrace{T^*T(v_1 - v_2)}_{=0}, v_1 - v_2\rangle = 0$$

and $T(v_1) = T(v_2)$.

S_1 is surjective by definition: every vector $Tv \in \mathrm{range}(T)$ is the image of $\sqrt{T^*T}v$. Also, S_1 preserves norms: for any $v \in V$,

$$\|\sqrt{T^*T}v\|^2 = \langle \sqrt{T^*T}v, \sqrt{T^*T}v\rangle = \langle T^*Tv, v\rangle = \langle Tv, Tv\rangle = \|Tv\|^2.$$

This implies in particular that it is injective: if $Tv = 0$, then $\|\sqrt{T^*T}v\| = \|Tv\| = 0$, so $\sqrt{T^*T}v$ was 0.

Since $\mathrm{range}(\sqrt{T^*T})$ and $\mathrm{range}(T)$ are isomorphic subspaces of V (via the isomorphism S_1), they have the same dimension. Taking orthogonal complements, it follows that

$$\dim \mathrm{range}(\sqrt{T^*T})^\perp = \dim \mathrm{range}(T)^\perp,$$

so we can find an orthonormal basis $e_1, ..., e_m$ of range$(\sqrt{T^*T})^\perp$ and a linear map

$$S_2 : \text{range}(\sqrt{T^*T})^\perp \longrightarrow \text{range}(T)^\perp$$

such that $S_2(e_1), ..., S_2(e_m)$ is also an orthonormal basis. In particular, $\|S_2(v)\| = \|v\|$ for all $v \in \text{range}(\sqrt{T^*T})^\perp$.

Let $S : V \to V$ be the linear function with $S|_{\text{range}(\sqrt{T^*T})} = S_1$ and $S|_{\text{range}(\sqrt{T^*T})^\perp} = S_2$. Then S is an isometry, because: for any vector

$$v = u + w \in V, \;\text{ with } u \in \text{range}(\sqrt{T^*T}),\; w \in \text{range}(\sqrt{T^*T})^\perp,$$

we know that $S(u) \in \text{range}(T)$ and $S(w) \in \text{range}(T)^\perp$ are orthogonal, so the Pythagorean theorem implies that

$$\|Sv\|^2 = \|Su + Sw\|^2 = \|Su\|^2 + \|Sw\|^2 = \|u\|^2 + \|w\|^2 = \|v\|^2.$$

By construction, for any $v \in V$,

$$S\sqrt{T^*T}v = S_1\sqrt{T^*T}v = Tv;$$

so $S\sqrt{T^*T} = T$. □

Remark: The proof can be made much shorter if we assume that T is invertible: in this case, $\sqrt{T^*T}$ is also invertible, since its square T^*T is positive definite. Define S by

$$S := T(\sqrt{T^*T})^{-1};$$

then since $T^* = (S\sqrt{T^*T})^* = \sqrt{T^*T}S^*$, it follows that

$$S^*S = \sqrt{T^*T}^{-1}\sqrt{T^*T} \cdot S^*S \cdot \sqrt{T^*T}\sqrt{T^*T}^{-1} = \sqrt{T^*T}^{-1}T^*T\sqrt{T^*T}^{-1} = I,$$

so S is unitary.

Example 68. The polar decomposition of complex numbers is a special case. For any nonzero $\lambda \in \mathbb{C}\backslash\{0\}$, interpreted as an operator on $\mathbb{C}$, the adjoint is the complex conjugate: $\lambda^* = \overline{\lambda}$. The polar decomposition becomes

$$\lambda = re^{i\theta},$$

where $r = \sqrt{\lambda^*\lambda} = |\lambda|$ and $e^{i\theta} = \frac{\lambda}{|\lambda|}$.

Remark[LADR 7.D.8, 7.D.9] Given any other factorization $T = SP$, where S is isometric and P is positive (semidefinite), we see that

$$T^*T = P^*S^*SP = P^*P = P^2,$$

and therefore $P = \sqrt{T^*T}$ is the positive square root of T^*T. In this sense, the polar decomposition is unique: the operator P is uniquely determined and if T is invertible, then the isometry $S = TP^{-1}$ is uniquely determined.

Singular value decomposition

Singular value decompositions of a matrix (or operator) are used in applied mathematics and statistics. Instead of studying the operator T itself, we study the eigenvalues of the positive semidefinite operator $\sqrt{T^*T}$; the spectral theorem guarantees that $\sqrt{T^*T}$ will have nice properties (for example, diagonalizable with real eigenvalues), and we use this to approximate properties of T.
Although we will not go into the applications here, you may remember a similar idea from Math 54: to approximate solutions to an unsolvable system of equations $Ax = b$, we studied the normal equations $(A^T A)x = A^T b$ instead.

Definition 52 (7.49)**.** Let $T \in \mathcal{L}(V)$ be an operator. A **singular value** of T is an eigenvalue of $\sqrt{T^*T}$.

Since $\sqrt{T^*T}$ is positive, the singular values $s_1, ..., s_n$ of T are all real and nonnegative.

Remark: The greatest singular value of T is often called the **spectral norm** of T, denoted $\|T\|_2$.
This is definite, because: if all eigenvalues of $\sqrt{T^*T}$ are 0, then it follows that $\sqrt{T^*T} = 0$; using the polar decomposition, we conclude that $T = 0$.
It is homogeneous, because: for any $\lambda \in \mathbb{C}$, the eigenvalues of $\sqrt{(\lambda T)^*(\lambda T)} = |\lambda|\sqrt{T^*T}$ are exactly $|\lambda|$ times the eigenvalues of $\sqrt{T^*T}$.
The triangle inequality is harder to verify.
Question for the reader: is the spectral norm induced by an inner product on $\mathcal{L}(V)$?

Proposition 101 (LADR 7.51)**.** *Let $T \in \mathcal{L}(V)$ have singular values $s_1, ..., s_n$. Then there are orthonormal bases $e_1, ..., e_n$ and $f_1, ..., f_n$ of V such that*

$$Tv = s_1\langle v, e_1\rangle f_1 + ... + s_n\langle v, e_n\rangle f_n$$

for all $v \in V$.

Proof. Let $e_1, ..., e_n$ be an orthonormal basis of eigenvectors of $\sqrt{T^*T}$ with eigenvalues $s_1, ..., s_n$. Let $T = S\sqrt{T^*T}$ be the polar decomposition of T, and define $f_1 = Se_1, ..., f_n = Se_n$. Since

$$v = \langle v, e_1\rangle e_1 + ... + \langle v, e_n\rangle e_n,$$

it follows that

$$Tv = S\sqrt{T^*T}v = S\Big(s_1\langle v, e_1\rangle e_1 + ... + s_n\langle v, e_n\rangle\Big) = s_1\langle v, e_1\rangle f_1 + ... + s_n\langle v, e_n\rangle f_n. \quad \square$$

This can be formulated as follows: the matrix of T with respect to the orthonormal bases $\mathcal{B} = \{e_1, ..., e_n\}$ and $\mathcal{C} = \{f_1, ..., f_n\}$ is

$$\mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(T) = \begin{pmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_n \end{pmatrix}.$$

This property determines the singular values:

Proposition 102. *Assume that $e_1, ..., e_n$ and $f_1, ..., f_n$ are orthonormal bases of V with the property that $Te_k = s_k f_k$ for nonnegative real numbers s_k. Then $s_1, ..., s_n$ are the singular values of T.*

Proof. Since

$$\langle e_j, T^* f_k \rangle = \langle Te_j, f_k \rangle = \langle s_j f_j, f_k \rangle = \delta_{jk} s_j = \delta_{jk} s_k = \langle e_j, s_k e_k \rangle$$

for all $1 \le j, k \le n$, it follows that $T^* f_k = s_k e_k$. Therefore, $T^* T e_k = s_k T^* f_k = s_k^2 e_k$, so s_k^2 is an eigenvalue of $T^* T$. Therefore, its nonnegative square root s_k is an eigenvalue of $\sqrt{T^* T}$, i.e. a singular value of T. □

The proofs above are deceptively short. Calculating the singular value decomposition is a lot of work, even for very simple operators.

Example 69. We will work through the singular value decomposition of the matrix $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ (with respect to the dot product on $\mathbb{R}^2$). In other words, we will find orthogonal matrices P and Q and a diagonal matrix Σ with nonnegative entries such that

$$A = P \Sigma Q^{-1} = P \Sigma Q^T.$$

In the notation above, the columns of Q will be e_1, e_2 and the columns of P will be f_1, f_2.

The eigenvalues of $A^T A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ are $\frac{3 \pm \sqrt{5}}{2}$, so the singular values of A are

$$s_1 = \sqrt{\frac{3 + \sqrt{5}}{2}} \approx 1.618, \quad s_2 = \sqrt{\frac{3 - \sqrt{5}}{2}} \approx 0.618.$$

The corresponding orthonormal basis of eigenvectors of $A^T A$ is (up to a choice of ± 1)

$$e_1 = \begin{pmatrix} \sqrt{\frac{5+\sqrt{5}}{10}} \\ \sqrt{\frac{5-\sqrt{5}}{10}} \end{pmatrix}, \; e_2 = \begin{pmatrix} -\sqrt{\frac{5-\sqrt{5}}{10}} \\ \sqrt{\frac{5+\sqrt{5}}{10}} \end{pmatrix}.$$

This allows us to compute the positive square root

$$
\begin{aligned}
\sqrt{A^T A} &= Q\Sigma Q^T \\
&= \begin{pmatrix} \sqrt{\frac{5+\sqrt{5}}{10}} & -\sqrt{\frac{5-\sqrt{5}}{10}} \\ \sqrt{\frac{5-\sqrt{5}}{10}} & \sqrt{\frac{5+\sqrt{5}}{10}} \end{pmatrix} \begin{pmatrix} \sqrt{\frac{3+\sqrt{5}}{2}} & 0 \\ 0 & \sqrt{\frac{3-\sqrt{5}}{2}} \end{pmatrix} \begin{pmatrix} \sqrt{\frac{5+\sqrt{5}}{10}} & \sqrt{\frac{5-\sqrt{5}}{10}} \\ -\sqrt{\frac{5-\sqrt{5}}{10}} & \sqrt{\frac{5+\sqrt{5}}{10}} \end{pmatrix} \\
&= \begin{pmatrix} 3/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix},
\end{aligned}
$$

and therefore the polar decomposition

$$
A = S\sqrt{A^T A}, \quad S = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix}^{-1} = \begin{pmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix}.
$$

Finally, the matrix P will be

$$
P = \begin{pmatrix} Se_1 & Se_2 \end{pmatrix} = \begin{pmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix} \begin{pmatrix} \sqrt{\frac{5+\sqrt{5}}{10}} & -\sqrt{\frac{5-\sqrt{5}}{10}} \\ \sqrt{\frac{5-\sqrt{5}}{10}} & \sqrt{\frac{5+\sqrt{5}}{10}} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{5-\sqrt{5}}{10}} & -\sqrt{\frac{5+\sqrt{5}}{10}} \\ \sqrt{\frac{5+\sqrt{5}}{10}} & \sqrt{\frac{5-\sqrt{5}}{10}} \end{pmatrix}.
$$

The singular value decomposition is now

$$
A = P\Sigma Q^T = \begin{pmatrix} \sqrt{\frac{5-\sqrt{5}}{10}} & -\sqrt{\frac{5+\sqrt{5}}{10}} \\ \sqrt{\frac{5+\sqrt{5}}{10}} & \sqrt{\frac{5-\sqrt{5}}{10}} \end{pmatrix} \begin{pmatrix} \sqrt{\frac{3+\sqrt{5}}{2}} & 0 \\ 0 & \sqrt{\frac{3-\sqrt{5}}{2}} \end{pmatrix} \begin{pmatrix} \sqrt{\frac{5+\sqrt{5}}{10}} & \sqrt{\frac{5-\sqrt{5}}{10}} \\ -\sqrt{\frac{5-\sqrt{5}}{10}} & \sqrt{\frac{5+\sqrt{5}}{10}} \end{pmatrix}.
$$

The first practical algorithm for computing the SVD (and essentially the same algorithm still used today) was found by Stanford professor Gene Golub and Berkeley professor William Kahan in 1965. Golub was so proud of this that he referenced it on his license plate: `https://upload.wikimedia.org/wikipedia/commons/9/90/Profsvd.JPG`

Complexification - 8/3

Complexification

Complexifying a real vector space is an abstraction of creating $\mathbb{C}^n$ from $\mathbb{R}^n$ by attaching imaginary parts to real vectors.

Definition 53 (LADR 9.2)**.** Let V be a real vector space. The complexification of V is the set

$$V_{\mathbb{C}} = V \times V,$$

with its componentwise addition. Scalar multiplication by complex numbers is defined by

$$(a + bi) \cdot (v, w) := (av - bw, bv + aw).$$

This makes $V_{\mathbb{C}}$ a $\mathbb{C}$-vector space. Elements $(v, w) \in V_{\mathbb{C}}$ are usually denoted $v + iw$.

Proposition 103 (LADR 9.4)**.** *Let V be a real vector space with basis $v_1, ..., v_n$. Then $v_1, ..., v_n$ are a basis of $V_{\mathbb{C}}$.*

Here, v_k denotes $v_k + 0i \in V_{\mathbb{C}}$.

Proof. (i) $v_1, ..., v_n$ spans $V_{\mathbb{C}}$, because: let $v + iw \in V_{\mathbb{C}}$ be any element, and write

$$v = \lambda_1 v_1 + ... + \lambda_n v_n, \quad w = \mu_1 v_1 + ... + \mu_n v_n, \quad \lambda_k, \mu_k \in \mathbb{R}.$$

Then $v + iw = (\lambda_1 + i\mu_1)v_1 + ... + (\lambda_n + i\mu_n)v_n$.
(ii) $v_1, ..., v_n$ is linearly independent, because: assume that

$$(\lambda_1 + i\mu_1)v_1 + ... + (\lambda_n + i\mu_n)v_n = 0.$$

Comparing the real and imaginary parts shows that

$$\lambda_1 v_1 + ... + \lambda_n v_n = \mu_1 v_1 + ... + \mu_n v_n = 0$$

in the real vector space V, so $\lambda_1 = \mu_1 = ... = \lambda_n = \mu_n = 0$. $\square$

Definition 54 (LADR 9.5)**.** Let V be a real vector space and let $T \in \mathcal{L}(V)$ be an operator. The **complexification** of T is the map

$$T_{\mathbb{C}} : V_{\mathbb{C}} \longrightarrow V_{\mathbb{C}}, \;\; T_{\mathbb{C}}(v + iw) = T(v) + iT(w).$$

$T_{\mathbb{C}}$ is $\mathbb{C}$-linear, because: it is $\mathbb{R}$-linear, and

$$T_{\mathbb{C}}(i(v + iw)) = T_{\mathbb{C}}(iv - w) = T(-w) + iT(v) = i(T(v) + iT(w)) = iT_{\mathbb{C}}(v + iw)$$

for all $v, w \in V$.

Proposition 104 (LADR 9.7)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a real vector space, and fix a basis $v_1, ..., v_n$. Then the matrix $\mathcal{M}(T)$ of T equals the matrix $\mathcal{M}(T_{\mathbb{C}})$ of $T_{\mathbb{C}}$.*

In particular, the complexification of a matrix map $A \in \mathbb{R}^{n,n}$ is just the same matrix, where the entries are interpreted as complex numbers: $A \in \mathbb{C}^{n,n}$.

Proof. If $\mathcal{M}(T) = (a_{ij})_{i,j}$, then $Tv_j = \sum_{i=1}^n a_{ij} v_i$; therefore,

$$T_{\mathbb{C}}(v_j) = T(v_j) = \sum_{i=1}^n a_{ij} v_i,$$

so $\mathcal{M}(T_{\mathbb{C}}) = (a_{ij})_{i,j}$. □

$T_{\mathbb{C}}$ inherits many of the properties of T. We will list a few here.

Proposition 105 (LADR 9.10)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a real vector space. Then the minimal polynomial of T equals the minimal polynomial of $T_{\mathbb{C}}$.*

Proof. Using the definition $T_{\mathbb{C}}(u + iv) = T(u) + iT(v)$, it follows that

$$T_{\mathbb{C}}^2(u + iv) = T_{\mathbb{C}}(T(u) + iT(v)) = T^2(u) + iT^2(v),$$

and repeating this argument shows that

$$T_{\mathbb{C}}^n(u + iv) = T^n(u) + iT^n(v)$$

for all $u, v \in V$. Therefore,

$$p(T_{\mathbb{C}})(u + iv) = p(T)u + ip(T)v, \;\; u, v \in V$$

for any real polynomial $p \in \mathcal{P}(\mathbb{R})$. In particular, if p is the minimal polynomial of T, then we see that

$$p(T_{\mathbb{C}})(u + iv) = p(T)u + ip(T)v = 0 + 0i.$$

We need to check that allowing complex coefficients does not let us find a polynomial $q \in \mathcal{P}(\mathbb{C})$ of smaller degree such that $q(T_{\mathbb{C}}) = 0$. Assume that $q(T_{\mathbb{C}}) = 0$ and write

$$q(x) = \sum_{k=0}^{n} (a_k + ib_k)x^k.$$

Then

$$0 = q(T_{\mathbb{C}})v = \sum_{k=0}^{n} (a_k + ib_k)T_{\mathbb{C}}^k v = \Big(\sum_{k=0}^{n} a_k T^k v\Big) + i\Big(\sum_{k=0}^{n} b_k T^k v\Big)$$

for all $v \in V$, so $\sum_{k=0}^{n} a_k T^k = \sum_{k=0}^{n} b_k T^k = 0$. Therefore, both $\sum_{k=0}^{n} a_k x^k$ and $\sum_{k=0}^{n} b_k x^k$ are polynomial multiples of p; so $\sum_{k=0}^{n} (a_k + ib_k)x^k$ is a (complex) polynomial multiple of p. $\square$

Proposition 106 (LADR 9.11)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a real vector space. A real number $\lambda \in \mathbb{R}$ is an eigenvalue of T if and only if it is an eigenvalue of $T_{\mathbb{C}}$, and*

$$E(\lambda, T_{\mathbb{C}}) = E(\lambda, T)_{\mathbb{C}}, \;\; G(\lambda, T_{\mathbb{C}}) = G(\lambda, T)_{\mathbb{C}}.$$

Proof. We first show that $\mathrm{null}(T_{\mathbb{C}}) = \mathrm{null}(T)_{\mathbb{C}}$; i.e. the null space of $T_{\mathbb{C}}$ is the complexification of the null space of T. This is because

$$u + iv \in \mathrm{null}(T_{\mathbb{C}}) \;\Leftrightarrow\; Tu + iTv = 0 + 0i \;\Leftrightarrow\; u, v \in \mathrm{null}(T).$$

Applying this to $T - \lambda I$ and $(T - \lambda I)^k$, it follows that

$$E(\lambda, T)_{\mathbb{C}} = \mathrm{null}(T - \lambda I)_{\mathbb{C}} = \mathrm{null}(T_{\mathbb{C}} - \lambda I) = E(\lambda, T_{\mathbb{C}})$$

and

$$G(\lambda, T)_{\mathbb{C}} = \bigcup_{k=1}^{\infty} \mathrm{null}(T - \lambda I)_{\mathbb{C}}^k = \bigcup_{k=1}^{\infty} \mathrm{null}(T_{\mathbb{C}} - \lambda I)^k = G(\lambda, T_{\mathbb{C}}). \qquad \square$$

It will be useful to have a notion of taking the complex conjugate of an operator. The idea is that, for matrices, the complex conjugate of a matrix should just consist of the complex conjugate in each entry.

For any $v + iw \in V_{\mathbb{C}}$, define $\overline{v + iw} := v - iw$; and for any $S \in \mathcal{L}(V_{\mathbb{C}})$, define $\overline{S}$ by

$$\overline{S}(u + iv) = \overline{S(u - iv)}.$$

In other words, for $v \in V_{\mathbb{C}}$, we define $\overline{S}v := \overline{S(\overline{v})}$. Then

$$\overline{S}(i(u + iv)) = \overline{S}(-v + ui) = \overline{S(-v - ui)} = \overline{i}\overline{S(-v - ui)} = i\overline{S(u - vi)} = i\overline{S}(u + iv),$$

so $\overline{S}$ is also $\mathbb{C}$-linear. We can check the following properties:
(i) $\overline{S}\overline{v} = \overline{Sv}$ for all $v \in V_{\mathbb{C}}$, by definition;
(ii) $\overline{S+T} = \overline{S} + \overline{T}$ for all $S, T \in \mathcal{L}(V_{\mathbb{C}})$;
(iii) for any $S, T \in \mathcal{L}(V_{\mathbb{C}})$ and $v \in V_{\mathbb{C}}$,

$$\overline{S} \cdot \overline{T}v = \overline{ST \cdot \overline{v}} = \overline{S\overline{\overline{T\overline{v}}}} = \overline{(ST)\overline{v}} = \overline{ST}v,$$

so $\overline{ST} = \overline{S} \cdot \overline{T}$.

Proposition 107. *Let $S \in \mathcal{L}(V_{\mathbb{C}})$ be an operator on the complexification of a real vector space V. Then $S = T_{\mathbb{C}}$ for some $T \in \mathcal{L}(V)$ if and only if $S = \overline{S}$.*

Proof. Assume that $S = \overline{S}$. Then, for any $v \in V$,

$$\overline{Sv} = \overline{S}\overline{v} = S\overline{v} = Sv.$$

Therefore, S defines an operator $T : V \to V$, $Tv + 0i := S(v + 0i)$. Then $S = T_{\mathbb{C}}$, because: for any $v + iw \in V_{\mathbb{C}}$,

$$S(v + iw) = Sv + iSw = Tv + iTw = T_{\mathbb{C}}(v + iw).$$

On the other hand, the conjugate of any operator $T_{\mathbb{C}}$ is

$$\overline{T_{\mathbb{C}}}(v + iw) = \overline{T_{\mathbb{C}}(v - iw)} = \overline{T(v) - iT(w)} = T(v) + iT(w) = T_{\mathbb{C}}(v + iw),$$

i.e. $\overline{T_{\mathbb{C}}} = T_{\mathbb{C}}$. □

Passing to $T_{\mathbb{C}}$ can create new (nonreal) eigenvalues. The eigenspace of any nonreal eigenvalue and its complex conjugate are closely related:

Proposition 108 (LADR 9.12)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a real vector space and $\lambda \in \mathbb{C}$. For any $k \in \mathbb{N}$ and $u, v \in V$,*

$$(T_{\mathbb{C}} - \lambda I)^k (u + iv) = 0 \;\Leftrightarrow\; (T_{\mathbb{C}} - \overline{\lambda} I)^k (u - iv) = 0.$$

Proof. This is because

$$(T_{\mathbb{C}} - \overline{\lambda} I)^k (u - iv) = \overline{(T_{\mathbb{C}} - \lambda I)^k (u + iv)}.$$ □

In particular, the complex conjugation is an $\mathbb{R}$-linear map between $G(\lambda, T_{\mathbb{C}})$ and $G(\overline{\lambda}, T_{\mathbb{C}})$, so the algebraic multiplicities of λ and $\overline{\lambda}$ are equal.

Definition 55. Let $T \in \mathcal{L}(V)$ be an operator on a real vector space. The **characteristic polynomial** of T is defined as the characteristic polynomial of its complexification $T_{\mathbb{C}}$.

As before, this can also be computed as the determinant $\det(xI - T)$ (once the determinant has been defined). The Cayley-Hamilton theorem holds: any operator T satisfies its characteristic equation, because $T_{\mathbb{C}}$ satisfies its characteristic equation.

Normal operators on a real space - 8/4

Real normal operators

Given a real inner product space V, the complexification becomes an inner product space via

$$\langle v_1 + iw_1, v_2 + iw_2 \rangle := \langle v_1, v_2 \rangle + \langle w_1, w_2 \rangle + i\Big(\langle w_1, v_2 \rangle - \langle v_1, w_2 \rangle\Big).$$

(You will show this on the problem set.)

Proposition 109 (LADR 9.B.4)**.** *Let $T \in \mathcal{L}(V)$ be an operator. Then:*
(i) T is self-adjoint if and only if $T_{\mathbb{C}}$ is self-adjoint;
(ii) T is normal if and only if $T_{\mathbb{C}}$ is normal.

Proof. This will be clear when we show that the complexification of T^* is always $(T_{\mathbb{C}})^*$. For any $v_1, v_2, w_1, w_2 \in V$,

$$\begin{aligned}
\langle T_{\mathbb{C}}(v_1 + iw_1), v_2 + iw_2 \rangle &= \langle T(v_1) + iT(w_1), v_2 + iw_2 \rangle \\
&= \langle T(v_1), v_2 \rangle + \langle T(w_1), w_2 \rangle + i\Big(\langle T(w_1), v_2 \rangle - \langle T(v_1), w_2 \rangle\Big) \\
&= \langle v_1, T^*(v_2) \rangle + \langle w_1, T^*(w_2) \rangle + i\Big(\langle w_1, T^*(v_2) \rangle - \langle v_1, T^* w_2 \rangle\Big) \\
&= \langle v_1 + iw_1, T^*(v_2) + iT^*(w_2) \rangle \\
&= \langle v_1 + iw_1, (T^*)_{\mathbb{C}}(v_2 + iw_2) \rangle,
\end{aligned}$$

so $(T^*)_{\mathbb{C}} = (T_{\mathbb{C}})^*$. $\square$

The eigenvalues of the complexified operator $T_{\mathbb{C}}$ are either real or come in complex conjugate pairs. First, we will study the complex conjugate pairs by themselves.

Proposition 110 (LADR 9.27)**.** *Let $T \in \mathcal{L}(V)$ be a normal operator on a 2-dimensional real inner product space. Assume that T is not self-adjoint. Then the matrix of T with respect to every orthonormal basis of V has the form*

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad b \neq 0.$$

Conversely, any matrix of this form is normal and not symmetric, which implies that T is normal and not self-adjoint.

The matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ is unique up to the sign of b, because its eigenvalues are $a \pm ib$, which must be the two eigenvalues of $T_{\mathbb{C}}$.

Proof. (i) If the matrix of T with respect to an orthonormal basis $\{e_1, e_2\}$ is $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then the matrix of T^* is $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$. Therefore,

$$a^2 + b^2 = \|T^* e_1\|^2 = \|T e_1\|^2 = a^2 + c^2,$$

so $b = \pm c$. Since T is not self-adjoint, $b = -c$ and b is nonzero. Also,

$$(a+b)^2 + (-b+d)^2 = \|T(e_1+e_2)\|^2 = \|T^*(e_1+e_2)\|^2 = (a-b)^2 + (b+d)^2,$$

which implies that $2ab - 2bd = -2ab + 2bd$, so $2ab - 2bd = 0$ and therefore $a = d$.

(ii) It is straightforward to check that

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} a & b \\ -b & a \end{pmatrix} = \begin{pmatrix} a^2+b^2 & 0 \\ 0 & a^2+b^2 \end{pmatrix} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \begin{pmatrix} a & -b \\ b & a \end{pmatrix}. \qquad \square$$

We can always assume that $b > 0$, because: if T is represented by $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ with respect to the basis e_1, e_2, then it is represented by $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ with respect to the basis e_2, e_1.

Proposition 111 (LADR 9.30)**.** *Let V be an inner product space and let $T \in \mathcal{L}(V)$ be normal. Let U be a T-invariant subspace of V. Then $U^\perp$ is also T-invariant; U is also T^*-invariant; and the adjoint of $T|_U$ is $(T^*)|_U$.*

Compare the third proof of the complex spectral theorem in the notes from 7/26.

Proof. Since $T_{\mathbb{C}}$ is normal, we have seen on midterm 2 that there is a polynomial $p(x) = \sum_{k=0}^{n}(a_k + ib_k)x^k \in \mathcal{P}(\mathbb{C})$ such that $p(T_{\mathbb{C}}) = (T_{\mathbb{C}})^* = (T^*)_{\mathbb{C}}$. It follows that

$$(T^*)_{\mathbb{C}} = \frac{1}{2}((T^*)_{\mathbb{C}} + \overline{(T^*)_{\mathbb{C}}}) = \frac{1}{2}(p(T_{\mathbb{C}}) + \overline{p(T_{\mathbb{C}})}) = \sum_{k=0}^{n} a_k T_{\mathbb{C}}^k,$$

so $T^* = \sum_{k=0}^{n} a_k T^k =: q(T)$ is a *real* polynomial expression in T. Therefore:

(i) If U is T-invariant, then U is T^k-invariant for all exponents k; so U is also $q(T) = T^*$-invariant;

(ii) Let P_U denote the orthogonal projection onto U (which is self-adjoint). The statement that U is T^*-invariant means that $P_{U^\perp} T^* P_U = 0$: i.e. the $U^\perp$-component of any element $T(u)$, $u \in U$ is 0. Taking adjoints shows that

$$0 = (P_{U^\perp} T^* P_U)^* = P_U T P_{U^\perp},$$

which means that $U^\perp$ is also T-invariant.

(iii) This is because $\langle T|_U(u), v\rangle = \langle Tu, v\rangle = \langle u, T^* v\rangle = \langle u, T^*|_U(v)\rangle$ for all $u, v \in U$.

□

Proposition 112 (LADR 9.34)**.** *Let V be a real inner product space and $T \in \mathcal{L}(V)$. Then T is normal if and only if there is an orthonormal basis of V, with respect to which T is represented by a block diagonal matrix of* (1×1)*-blocks and* (2×2)*-blocks of the form* $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ *with* $b > 0$.

Proof. In view of 9.27, it is clear that any matrix of this form is normal.

Assume that T is normal. We use induction on $\dim(V)$.

(i) If $\dim(V) = 1$, then T is represented by a (1×1)-block.

(ii) If $\dim(V) = 2$, then T is either self-adjoint (in which case it is orthogonally diagonalizable; i.e. we get two (1×1)-blocks) or it is represented by a matrix of the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ with $b > 0$.

(iii) If $\dim(V) \geq 3$, then we can find a proper invariant subspace U: if T has a real eigenvalue, then we can let U be the span of an eigenvector. Otherwise, if $T_{\mathbb{C}}$ has a complex eigenvalue $(a+ib)$ with complex eigenvector $u+iv$, then

$$T(u)+iT(v)=T_{\mathbb{C}}(u+iv)=(a+ib)(u+iv)=(au-bv)+i(bu+av)$$

implies that $U=\mathrm{Span}(u,v)$ is an invariant subspace. Since $T|_U$ and $T|_{U^\perp}$ are normal, they can both be represented by matrices of this form; therefore, T can be represented by a matrix of this form. $\square$

Real isometries

We proved earlier that $(T^*)_{\mathbb{C}}=(T_{\mathbb{C}})^*$ for any operator $T\in\mathcal{L}(V)$ on a real inner product space. In particular, T is an isometry if and only if $T_{\mathbb{C}}$ is an isometry.

Proposition 113 (LADR 9.36)**.** *Let $S\in\mathcal{L}(V)$ be an operator on a finite-dimensional real inner product space. The following are equivalent:*
(i) S is an isometry;
(ii) There is an orthonormal basis of V, with respect to which S is represented by a block-diagonal matrix consisting of (1×1)-blocks ± 1 and (2×2)-blocks of the form

$$\begin{pmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{pmatrix},\quad \theta\in(0,\pi).$$

In other words, every isometry is a combination of rotations (the blocks $\begin{pmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{pmatrix}$) and reflections (the blocks (-1)) performed in sequence.

Proof. Any isometry S is normal (it commutes with its adjoint $S^*=S^{-1}$), so it has a representation with respect to an orthonormal basis by a matrix consisting of (1×1)-blocks and (2×2)-blocks of the form $\begin{pmatrix}a & -b\\ b & a\end{pmatrix}$, with $b>0$. Also, the complexification $S_{\mathbb{C}}$ is an isometry on $V_{\mathbb{C}}$, so its eigenvalues all have absolute value 1. This forces the (1×1)-blocks to be ± 1, and the (2×2)-blocks $\begin{pmatrix}a & -b\\ b & a\end{pmatrix}$ have eigenvalues $a\pm ib$, so a^2+b^2 must be 1. This means that $a=\cos(\theta)$ and $b=\sin(\theta)$ for some $\theta\in(0,\pi)$.

On the other hand, the blocks (± 1) and $\begin{pmatrix}a & -b\\ b & a\end{pmatrix}$ have inverse equal to their transpose, so this is true for the entire matrix of S; therefore, S is represented by an isometry with respect to an orthonormal basis of V, so S itself is an isometry. $\square$

Trace - 8/8

Alternating forms

The **trace** of a matrix is the sum of its diagonal: for example,

$$\mathrm{tr}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = 1 + 5 + 9 = 15.$$

It turns out that the trace is invariant under similarity: $\mathrm{tr}(PAP^{-1}) = \mathrm{tr}(A)$ for any invertible matrix P. This implies that the trace is the same under change of basis, so we expect that we can define it without needing to choose a basis and work in coordinates.

On a complex vector space, the trace turns out to be the sum with multiplicites of the eigenvalues; but this definition is difficult to work with and is not valid over other fields. We will give a basis-free definition using the concept of alternating forms, which will also be useful tomorrow to talk about the determinant. Alternating forms will not be tested on the exam.

Definition 56. Let V be a vector space over $\mathbb{F}$ and let $k \in \mathbb{N}$ be a natural number. An **alternating k-form** on V is a map

$$\omega : \underbrace{V \times ... \times V}_{k \,\text{times}} \longrightarrow \mathbb{F}$$

with the following properties:
(i) ω is linear in every component.
(ii) ω is *alternating*: if the list $(v_1, ..., v_k)$ contains two copies of the same vector, then $\omega(v_1, ..., v_k) = 0$.

Example 70. The determinant

$$\omega\left(\begin{pmatrix} a_1 \\ b_1 \end{pmatrix}, \begin{pmatrix} a_2 \\ b_2 \end{pmatrix} \right) := \det \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} = a_1 b_2 - a_2 b_1$$

is an alternating 2-form on $\mathbb{R}^2$: it is linear in each column vector, and if the column vectors are equal, then the determinant is 0. The determinant is the example of alternating form that you should always have in mind.

These are called alternating because it alternates between positive and negative whenever we swap the order of two vectors. You are probably familiar with this property of the determinant. It follows from the calculation

$$\begin{aligned}
0 &= \omega(v_1, ..., v + w, ..., v + w, ..., v_k) \\
&= \omega(v_1, ..., v, ..., v, ..., v_k) + \omega(v_1, ..., v, ..., w, ..., v_k) \\
&\quad + \omega(v_1, ..., w, ..., v, ..., v_k) + \omega(v_1, ..., w, ..., w, ..., v_k) \\
&= \omega(v_1, ..., v, ..., w, ..., v_k) + \omega(v_1, ..., w, ..., v, ..., v_k),
\end{aligned}$$

so

$$\omega(v_1, ..., v, ..., w, ..., v_k) = -\omega(v_1, ..., w, ..., v, ..., v_k).$$

Proposition 114. *Let V be a vector space over $\mathbb{F}$ and $k \in \mathbb{N}$.*
(i) The alternating k-forms on V form a vector space, denoted $\Omega^k(V)$.
(ii) If V is finite-dimensional with $n = \dim(V)$, then

$$\dim \Omega^k(V) = \binom{n}{k}$$

is the binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \#\{\text{subsets of } \{1, ..., n\} \text{ of size } k\}$.

In particular, $\Omega^n(V)$ is 1-dimensional.

Proof. Calculating the dimension of $\Omega^k(V)$ rigorously will take too much time. The idea is that, if $v_1, ..., v_n$ is a basis of V, then we can define an alternating form $\omega \in \Omega^k(V)$ uniquely by specifying the values $\omega(v_{i_1}, ..., v_{i_k})$, where $\{i_1, ..., i_k\} \subseteq \{1, ..., n\}$ runs through the subsets of size k. □

Trace

The trace of an operator is defined similarly to the dual map.

Proposition 115. *Let V be an n-dimensional vector space. Let $T \in \mathcal{L}(V)$ be an operator. Then*

$$\mathrm{tr}(T) : \Omega^n(V) \longrightarrow \Omega^n(V),$$

$$\begin{aligned}
\mathrm{tr}(T)\omega(v_1, ..., v_n) &:= \omega(Tv_1, v_2, ..., v_n) + \omega(v_1, Tv_2, ..., v_n) + ... + \omega(v_1, v_2, ..., Tv_n) \\
&= \sum_{k=1}^{n} \omega(v_1, ..., Tv_k, ..., v_n)
\end{aligned}$$

is a well-defined linear map, called the ***trace*** *of T.*

Proof. It is not hard to see that

$$\omega(Tv_1, v_2, ..., v_n) + \omega(v_1, Tv_2, ..., v_n) + ... + \omega(v_1, v_2, ..., Tv_n)$$

is linear in every component. If $v_j = v_k$, then most of the terms above are 0, since they contain two equal vectors; we are left with

$$\begin{aligned} &\omega(v_1, ..., Tv_j, ..., v_k, ..., v_n) + \omega(v_1, ..., v_j, ..., Tv_k, ..., v_n) \\ = &\omega(v_1, ..., Tv_j, ..., v_j, ..., v_n) + \omega(v_1, ..., v_j, ..., Tv_j, ..., v_n) \end{aligned}$$

which is also 0 because $\omega(v_1, ..., v_j, ..., Tv_j, ..., v_n)$ results from $\omega(v_1, ..., Tv_j, ..., v_j, ..., v_n)$ by swapping the vectors in the jth and kth positions. □

Since $\Omega^n(V)$ is 1-dimensional, the map $\mathrm{tr}(T)$ is actually multiplication by a scalar. The trace of T will usually refer to that scalar.

Example 71. Let $T = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \in \mathcal{L}(\mathbb{R}^2)$. We consider the alternating 2-form

$$\omega\Big(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \Big) := ad - bc$$

from earlier. If e_1, e_2 is the standard basis of $\mathbb{R}^2$, then

$$\mathrm{tr}(T)\omega(e_1, e_2) = \omega(Te_1, e_2) + \omega(e_1, T_2) = \det \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} + \det \begin{pmatrix} 1 & 2 \\ 0 & 4 \end{pmatrix} = 1 + 4 = 5,$$

and $\omega(e_1, e_2) = 1$; so $\mathrm{tr}(T) = 5$.

Proposition 116 (LADR 10.13, 10.16)**.** *Let $T \in \mathcal{L}(V)$ and let $v_1, ..., v_n$ be a basis of V. Let $\mathcal{M}(T) = (a_{ij})_{i,j}$ be the matrix of T. Then $\mathrm{tr}(T) = a_{11} + a_{22} + ... + a_{nn}$ is the sum of the diagonal of that matrix.*

This means we have recovered the definition at the beginning of these notes.

Proof. After writing $Tv_1 = a_{11}v_1 + ... + a_{n1}v_n$, we see that

$$\omega(Tv_1, v_2, ..., v_n) = a_{11}\omega(v_1, v_2, ..., v_n) + ... + a_{n1}\omega(v_n, v_2, ..., v_n) = a_{11}\omega(v_1, v_2, ..., v_n)$$

for any $\omega \in \Omega^n(V)$, since all but the first term in this sum contain ω evaluated at two copies of the same vector. A similar argument shows that

$$\omega(v_1, ..., Tv_k, ..., v_n) = a_{kk}\omega(v_1, ..., v_n)$$

for all indices k, so

$$\mathrm{tr}(T)\omega(v_1, ..., v_n) = (a_{11} + ... + a_{nn})\omega(v_1, ..., v_n),$$

i.e. $\mathrm{tr}(T) = a_{11} + ... + a_{nn}$. □

In particular, the trace is also the sum of the eigenvalues of T, if V is a complex vector space:

Proposition 117 (LADR 10.16). *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional complex vector space. Then $\mathrm{tr}(T)$ is the sum of the eigenvalues of T, with algebraic multiplicities.*

Proof. Choose a basis of T with respect to which T is represented by its Jordan normal form. Then the diagonal of that matrix consists of exactly the eigenvalues of T, with algebraic multiplicities counted, and the previous proposition shows that $\mathrm{tr}(T)$ is the sum of that diagonal. $\square$

Example 72. The complex matrix $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}$ has eigenvalues

$$\lambda_k = 4 + 2\sqrt{11} \cos\Big(\frac{\arctan(\sqrt{106}/35) + 2\pi k}{3}\Big),\ k = 0, 1, 2.$$

In particular, $\lambda_0 = 10.603...$, $\lambda_1 = 0.151...$, $\lambda_2 = 1.245...$ It is difficult to calculate this by hand. However, we easily calculate that their sum is

$$\lambda_0 + \lambda_1 + \lambda_2 = \mathrm{tr}(A) = 1 + 2 + 9 = 12.$$

The following theorem is much easier to prove using the definition as the sum of the diagonal than the definition as the map on alternating forms, or the sum of eigenvalues:

Proposition 118 (LADR 10.14). *Let $S, T \in \mathcal{L}(V)$ be operators. Then*

$$\mathrm{tr}(ST) = \mathrm{tr}(TS).$$

Proof. Assume that S is represented by $A = (a_{ij})_{i,j}$ and T is represented by $B = (b_{ij})_{i,j}$ with respect to some basis of V. Then

$$\mathrm{tr}(ST) = \sum_{k=1}^n (AB)_{kk} = \sum_{k=1}^n \sum_{l=1}^n a_{kl} b_{lk} = \sum_{l=1}^n \sum_{k=1}^n b_{lk} a_{kl} = \sum_{l=1}^n (BA)_{ll} = \mathrm{tr}(TS). \qquad \square$$

Be careful: this only means that the trace is invariant under *cyclic* permutations of operators! For example, if $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ and $C = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, then

$$\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB) = 4,$$

while

$$\mathrm{tr}(ACB) = \mathrm{tr}(CBA) = \mathrm{tr}(BAC) = 3.$$

An important corollary: on finite-dim. vector spaces over $\mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$, there are no operators S, T such that $ST - TS = I$. This is because $\mathrm{tr}(ST - TS) = 0$ but $\mathrm{tr}(I) \neq 0$.

Trace on an inner product space

On an inner product space, the trace is easier to write:

Proposition 119. *Let V be a finite-dimensional inner product space with orthonormal basis $e_1, ..., e_n$. Let $T \in \mathcal{L}(V)$. Then*

$$\mathrm{tr}(T) = \sum_{k=1}^{n} \langle Te_k, e_k \rangle.$$

Proof. Since

$$Te_j = \langle Te_j, e_1 \rangle e_1 + ... + \langle Te_j, e_n \rangle e_n, \quad j \in \{1, ..., n\},$$

it follows that the representation matrix of T with respect to the basis $\{e_1, ..., e_n\}$ is $(\langle Te_j, e_i \rangle)_{i,j}$. Therefore, the trace is the sum of the diagonal of this matrix:

$$\mathrm{tr}(T) = \sum_{k=1}^{n} \langle Te_k, e_k \rangle. \qquad \square$$

This has several useful applications. One is an easy estimate of the largest eigenvalue of a matrix (a special case of *Schur's inequality*):

Proposition 120. *Let $A = (a_{ij})_{i,j}$ be a square complex matrix and let λ be an eigenvalue of A. Then $|\lambda| \leq \sqrt{\sum_{i,j} |a_{ij}|^2}$.*

Proof. Let e_1 be an eigenvector of A for λ, such that $\|e_1\| = 1$. Extend e_1 to an orthonormal basis $e_1, ..., e_n$ of $\mathbb{F}^n$. Then:

$$
\begin{aligned}
\sum_{i,j} |a_{ij}|^2 &= \sum_{i,j} a_{ij}\overline{a_{ji}} \\
&= \operatorname{tr}(AA^*) \\
&= \operatorname{tr}(A^*A) \\
&= \sum_{k=1}^{n} \langle A^*Ae_k, e_k \rangle \\
&= \sum_{k=1}^{n} \langle Ae_k, Ae_k \rangle \\
&= |\lambda|^2 + \sum_{k\neq 1} \|Ae_k\|^2 \\
&\geq |\lambda|^2.
\end{aligned}
$$

□

Example 73. The largest eigenvalue of

$$
\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}
$$

is

$$
4 + 2\sqrt{11}\cos\left(\frac{\arctan(\sqrt{106}/35)}{3}\right) = 10.603...,
$$

while the bound we have found above is

$$
\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 2^2 + 4^2 + 1^2 + 3^2 + 9^2} = \sqrt{115} = 10.724...
$$

Not bad. (Of course, it won't always be this close.)

Determinant - 8/9

Determinant

Like yesterday, we will first give a basis-free definition of the determinant using alternating forms, and then discuss the practical aspects of the determinant by studying it on matrices.

Proposition 121. *Let V be an n-dimensional vector space and let $T \in \mathcal{L}(V)$ be an operator. Let $\omega \in \Omega^n(V)$ be an alternating form. Then*

$$\psi(v_1, ..., v_n) := \omega(Tv_1, ..., Tv_n)$$

defines an alternating form.

Proof. Since T is linear and ω is linear in each component, it follows that ψ is linear in each component. Also, if $v_j = v_k$ for any indices $j \neq k$, then $Tv_j = Tv_k$, and therefore

$$\omega(Tv_1, ..., Tv_n) = 0;$$

so ψ is alternating. □

Since $\Omega^n(V)$ is 1-dimensional, it follows that there is a scalar, called the **determinant** $\det(T)$, such that

$$\det(T)\omega(v_1, ..., v_n) = \omega(Tv_1, ..., Tv_n), \;\; \omega \in \Omega^n(V).$$

Example 74. Recall that on $\mathbb{R}^2$, there is a nonzero alternating 2-form ω defined by

$$\omega\Big(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \Big) = x_1 y_2 - x_2 y_1.$$

In particular, $\omega(e_1, e_2) = 1$. If $T = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, then

$$\det(T) = \det(T)\omega(e_1, e_2) = \omega(Te_1, Te_2) = \omega\Big(\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \end{pmatrix} \Big) = -2.$$

Proposition 122 (LADR 10.40)**.** *Let $S, T \in \mathcal{L}(V)$ be operators. Then*

$$\det(ST) = \det(S) \cdot \det(T).$$

Proof. For any alternating form $\omega \in \Omega^n(V)$ and $v_1, ..., v_n \in V$,

$$\begin{aligned} \det(ST)\omega(v_1, ..., v_n) &= \omega(STv_1, ..., STv_n) \\ &= \det(S)\omega(Tv_1, ..., Tv_n) \\ &= \det(S)\det(T)\omega(v_1, ..., v_n), \end{aligned}$$

so $\det(ST) = \det(S)\det(T)$. □

Permutations and the Leibniz formula

Definition 57 (LADR 10.27)**.** (i) A **permutation** σ on n numbers is a bijective function $\sigma : \{1, ..., n\} \to \{1, ..., n\}$.
(ii) The **sign** $\mathrm{sgn}(\sigma)$ of a permutation σ is $(-1)^e$, where e number of pairs of indices (j, k) such that $j < k$ but $\sigma(j) > \sigma(k)$.

The permutation σ can be written out as the list $(\sigma(1), ..., \sigma(n))$.

Example 75. The sign of the permutation $(5, 3, 2, 4, 1)$ is 1, because: the pairs of elements that are not in order in the list $(5, 3, 2, 4, 1)$ are

$$(5, 3), (5, 2), (5, 4), (5, 1), (3, 2), (3, 1), (2, 1), (4, 1).$$

There are 8 of these.

Every permutation is made up of cyclic permutations. For example, $(5, 3, 2, 4, 1)$ consists of the cycles

$$1 \to 5 \to 1 \to 5 \to 1 \to ...$$

and

$$2 \to 3 \to 2 \to 3 \to 2 \to ...$$

and

$$4 \to 4 \to 4 \to 4 \to 4 \to ...$$

The *length* of each cycle is the number of distinct numbers it contains. We write the cycles as tuples without commas: here, they are $(1\ 5)$, $(2\ 3)$ and (4).

A faster way to calculate the sign of the permutation is as follows: we multiply -1 for every cycle of even length. For example, $(5, 3, 2, 4, 1)$ consists of two cycles $(1\,5)$ and $(2\,3)$ of length two and one cycle (4) of length 1; the sign is $(-1) \cdot (-1) = 1$.

Proposition 123. *Let $\omega \in \Omega^n(V)$ be an alternating form. For any permutation σ and $v_1, ..., v_n \in V$,*

$$\omega(v_{\sigma(1)}, ..., v_{\sigma(n)}) = \mathrm{sgn}(\sigma) \cdot \omega(v_1, ..., v_n).$$

Compare to LADR 10.38.

Proof. ω is alternating, so every time we swap two vectors v_j and v_k, the sign changes:

$$\omega(v_1, ..., v_j, ..., v_k, ..., v_n) = -\omega(v_1, ..., v_k, ..., v_j, ..., v_n).$$

More generally, applying a cyclic permutation of length ℓ can be understood as a sequence of $\ell - 1$ swaps: for example, the cyclic permutation

$$1 \longrightarrow 2 \longrightarrow 3 \longrightarrow 4 \longrightarrow 1 \longrightarrow 2 \longrightarrow 3 \longrightarrow 4 \longrightarrow 1 \longrightarrow ...$$

i.e. the permutation $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$, is equivalent to swapping $(3,4)$, then $(2,3)$, then $(1,2)$ in that order.

This implies that, if σ is a cyclic permutation of length ℓ, then

$$\omega(v_{\sigma(1)}, ..., v_{\sigma(n)}) = (-1)^{\ell-1}\omega(v_1, ..., v_n).$$

Here, $(-1)^{\ell-1}$ is 1 if σ has odd length and it is 1 if σ has even length. The claim follows by splitting σ into its cycles, since $\mathrm{sgn}(\sigma)$ is the product of $(-1)^{\ell-1}$ as ℓ runs through the lengths of the cycles of σ. □

Using this notation, we can write down an explicit formula for the determinant. In practice, it is extremely slow, and therefore rarely used; the most important consequence is probably that the determinant is a polynomial expression in the entries of a matrix and it is therefore continuous, differentiable, etc.

Proposition 124 (LADR 10.33)**.** *Let $T \in \mathcal{L}(V)$. Let $v_1, ..., v_n$ be a basis of V and assume that the representation matrix of T is $(a_{ij})_{i,j}$. Then*

$$\det(T) = \sum_{\sigma} \mathrm{sgn}(\sigma) \cdot a_{\sigma(1),1}a_{\sigma(2),2}...a_{\sigma(n),n}.$$

Proof. Each Tv_j is the linear combination $Tv_j = \sum_{i=1}^n a_{ij}v_i$. Using linearity in each component, for any $\omega \in \Omega^n(V)$, we can write

$$\omega(Tv_1, ..., Tv_n) = \omega\Big(\sum_{i_1=1}^n a_{i_1 1}v_{i_1}, ..., \sum_{i_n=1}^n a_{i_n n}v_{i_n}\Big) = \sum_{i_1=1}^n ... \sum_{i_n=1}^n a_{i_1 1}...a_{i_n n}\omega(v_{i_1}, ..., v_{i_n}).$$

Note that ω vanishes whenever we plug in two copies of the same vector; so this sum is actually only over those lists $(i_1, ..., i_n)$ that are permutations: i.e.

$$\det(T)\omega(v_1, ..., v_n) = \sum_\sigma a_{\sigma(1)1}...a_{\sigma(n)n}\omega(v_{\sigma(1)}, ..., v_{\sigma(n)}).$$

Using the previous result, we can rewrite this as

$$\det(T)\omega(v_1, ..., v_n) = \sum_\sigma \mathrm{sgn}(\sigma) \cdot a_{\sigma(1),1}a_{\sigma(2),2}...a_{\sigma(n),n}\omega(v_1, ..., v_n),$$

so $\det(T) = \sum_\sigma \mathrm{sgn}(\sigma) \cdot a_{\sigma(1),1}a_{\sigma(2),2}...a_{\sigma(n),n}$. $\square$

Example 76. You may be familiar with this formula in the case of (3×3)-matrices under the name *Sarrus' rule*: the determinant of a (3×3)-matrix is calculated by copying the left two columns to the right of the matrix and adding diagonally as

$$\begin{array}{ccc|cc} + & + & + & & \\ a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \\ - & - & - & & \end{array}$$

or written out,

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

Here, the upper left - lower right diagonals correspond to the permutations

$$(1, 2, 3),\ (3, 1, 2),\ (2, 3, 1)$$

with positive sign $+1$ and the lower left - upper right diagonals correspond to the permutations

$$(3, 2, 1),\ (1, 3, 2),\ (2, 1, 3)$$

with negative sign -1.

The most practical way to calculate determinants is via Gaussian elimination (row reduction). Some matrices (those with a lot of zero entries) can be calculated quickly by expanding along a row or column. You have probably seen this in Math 54.

Determinant and eigenvalues

Proposition 125 (LADR 10.42)**.** *Let $T \in \mathcal{L}(V)$ be an operator on a complex vector space. Then* $\det(T)$ *is the product of the eigenvalues of T, counted by algebraic multiplicity.*

Proof. Let $A = (a_{ij})_{i,j}$ be the Jordan normal form of T. Since $a_{ij} = 0$ whenever $i > j$, the only nonzero terms in the Leibniz formula

$$\det(T) = \sum_\sigma \mathrm{sgn}(\sigma) \cdot a_{\sigma(1),1} ... a_{\sigma(n),n}$$

must have $\sigma(1) = 1$; and therefore $\sigma(2) = 2$, and so on until $\sigma(n) = n$. In other words, the only σ resulting in a nonzero term is the identity, and the determinant is just the product along the diagonal:

$$\det(T) = a_{11} \cdot ... \cdot a_{nn}.$$

Finally, note that the Jordan normal form has the eigenvalues along the diagonal, each counted as often as their algebraic multiplicity. □

The same argument shows that the determinant of any upper-triangular matrix is the product along the diagonal.

Example 77. Recall that the eigenvalues of $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}$ were

$$\lambda_0 = 4 + 2\sqrt{11} \cos\Big(\frac{\arctan(\sqrt{106}/35)}{3}\Big) = 10.603...$$

$$\lambda_1 = 4 + 2\sqrt{11} \cos\Big(\frac{\arctan(\sqrt{106}/35) + 2\pi}{3}\Big) = 0.151...$$

$$\lambda_2 = 4 + 2\sqrt{11} \cos\Big(\frac{\arctan(\sqrt{106}/35) + 4\pi}{3}\Big) = 1.245...$$

To multiply these together directly would probably involve some mysterious and complicated trig identities. However, we know that the result will be

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