MATH251 - Honours Algebra 2

Vector spaces, linear (in)dependence, span, bases; linear transformations, kernel, image, isomorphisms, nilpotent operators; elementary matrices; diagonalization, eigenthings, Cayley-Hamilton; inner product spaces.

Based on lectures from Winter, 2024 by Prof. Anush Tserunyan
Notes by Louis Meunier

Contents

1 Introduction

Remark 1.1. *This course is about vector spaces and linear transformations between them; a vector space involves multiplication by* scalars*, where the scalars come from some field. We recall first examples of fields, then vector spaces, as a motivation, before presenting a formal definition.*

1.1 Vector Spaces

Remark 1.2. *Much of this is recall from Algebra 1.*

⊛ **Example 1.1: Examples of Fields**

1. $\mathbb{Q}$; the field of rational numbers.
2. $\mathbb{R}$; the field of real numbers; $\mathbb{Q} \subseteq \mathbb{R}$.
3. $\mathbb{C}$; the field of complex numbers; $\mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$.
4. $\mathbb{F}_p \equiv \mathbb{Z}/p\mathbb{Z} \equiv \{0, 1, \dots, p-1\}; the(unique)fieldofpelements, wherepprime.$[a]
   (a) $p = 2; \mathbb{F}_2 \equiv \{0, 1\}$.
   (b) $p = 3; \mathbb{F}_3 \equiv \{0, 1, 2\}$.
   (c) $\cdots$

[a]where $a +_p b :=$ remainder of $\frac{a+b}{p}$, $a \cdot_p b :=$ remainder of $\frac{a \cdot b}{p}$.

Remark 1.3. *Throughout the course, we will denote an abstract field as* $\mathbb{F}$.

⊛ **Example 1.2: Examples of Vector Spaces**

1. $\mathbb{R}^3 := \{(x, y, z) : x, y, z \in \mathbb{R}\}$. We can add elements in $\mathbb{R}^3$, and multiply them by real scalars.
2. $\mathbb{F}^n := \underbrace{\mathbb{F} \times \mathbb{F} \times \cdots \mathbb{F}}_{n \text{ times}} := \{(a_1, a_2, \dots, a_n) : a_i \in \mathbb{F}\}$, where $n \in \mathbb{N}$[1]; this is a generalization of the previous example, where we took $n = 3$, $\mathbb{F} = \mathbb{R}$. Operations follow identically; addition:

$$(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) := (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n)$$

and, taking a scalar $\lambda \in \mathbb{F}$, multiplication:

$$\lambda \cdot (a_1, a_2, \dots, a_n) := (\lambda \cdot a_1, \lambda \cdot a_2, \dots, \lambda \cdot a_n).$$

We refer to these elements $(a_1, \cdots, a_n)$ as *vectors* in $\mathbb{F}^n$; the vector for which $a_i = 0 \, \forall \, i$ is the 0 *vector*, and is the additive identity, making $\mathbb{F}^n$ an abelian group under addition, that admits multiplication by scalars from $\mathbb{F}$.

3. $C(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} : f \text{ continuous}\}$. Here, we have the constant zero function as our additive identity ($x \mapsto 0 \,\forall\, x$), and addition/scalar multiplication of two continuous real functions are continuous.

4. $\mathbb{F}[t] := \{a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n : a_i \in \mathbb{F} \,\forall\, i, n \in \mathbb{N}\}$, ie, the set of all polynomials in t with coefficients from $\mathbb{F}$. Here, we can add two polynomials;

$$(a_0 + a_1 t + \cdots + a_n t^n) + (b_0 + b_1 t + \cdots + b_m t^m) := \sum_{i=0}^{\max\{n,m\}} (a_i + b_i) t^i,$$

(where we "take" undefined a_i/b_i's as 0; that is, if $m > n$, then $a_{m-n}, a_{m-n+1}, \ldots, a_m$ are taken to be 0). Scalar multiplication is defined

$$\lambda \cdot (a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n) := \lambda a_0 + \lambda a_1 t + \lambda a_2 t^2 + \cdots + \lambda a_n t^n.$$

Here, the zero polynomial is simply 0 (that is, $a_i = 0 \,\forall\, i$).

↪ Definition 1.1: Vector Space

A *vector space* V over a field $\mathbb{F}$ is an *abelian group* with an operation denoted $+$ (or $+_V$) and identity element[2] denoted 0_V, equipped with *scalar multiplication* for each scalar $\lambda \in \mathbb{F}$ satisfying the following axioms:

1. $1 \cdot v = v$ for $1 \in \mathbb{F}$, $\forall\, v \in V$.
2. $\alpha \cdot (\beta \cdot v) = (\alpha \cdot \beta) v$, $\forall\, \alpha, \beta \in \mathbb{F}, v \in V$.
3. $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$, $\forall\, \alpha, \beta \in \mathbb{F}, v \in V$.
4. $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$, $\forall\, \alpha \in \mathbb{F}, u, v \in V$.

We refer to elements $v \in V$ as *vectors*.

[1] Where we take $0 \in \mathbb{N}$, for sake of consistency. Moreover, by convention, we define $\mathbb{F}^0$ (that is, when $n = 0$) to be $\{0\}$; the trivial vector space.

[2] The "zero vector".

↪ **Proposition 1.1**

For a vector space V over a field $\mathbb{F}$, the following holds:

1. $0 \cdot v = 0_V, \forall v \in V$ (where $0 := 0_{\mathbb{F}}$)
2. $-1 \cdot v = -v, \forall v \in V$ (where $1 := 1_{\mathbb{F}}$)[3]
3. $\alpha \cdot 0_V = 0_V, \forall \alpha \in \mathbb{F}$

Proof.

1. $0 \cdot v = (0+0) \cdot v = 0 \cdot v + 0 \cdot v \implies 0 \cdot v = 0_V$ (by "cancelling" one of the $0 \cdot v$ terms on each side).
2. $v + (-1 \cdot v) = (1 \cdot v + (-1) \cdot v) = (1-1) \cdot v = 0 \cdot v = 0_V \implies (-1 \cdot v) = -v$.
3. $\alpha \cdot 0_V = \alpha \cdot (0_V + 0_V) = \alpha \cdot 0_V + \alpha \cdot 0_V \implies \alpha \cdot 0_V = 0_V$ (by, again, cancelling a term on each side).

■

↪ Lecture 01; Last Updated: Sat Apr 6 10:19:07 EDT 2024

1.2 Creating Spaces from Other Spaces

↪ **Definition 1.2: Product/Direct Sum of Vector Spaces**

For vector spaces U, V over the same field $\mathbb{F}$, we define their *product* (or *direct sum*) as the set

$$U \times V = \{(u, v) : u \in U, v \in V\},$$

with the operations:

$$(u_1, v_1) + (u_2, v_2) := (u_1 + u_2, v_1 + v_2)$$
$$\lambda \cdot (u, v) := (\lambda \cdot u, \lambda \cdot v)$$

⊛ **Example 1.3:** $\mathbb{F}$

$\mathbb{F}^2 = \mathbb{F} \times \mathbb{F}$, where $\mathbb{F}$ is considered as the vector space over $\mathbb{F}$ (itself).

[3] NB: "additive inverse"

↪ Definition 1.3: Subspace

For a vector space V over a field $\mathbb{F}$, a *subspace* of V is a subset $W \subseteq V$ s.t.

1. $0_V \in W$[4]
2. $u + v \in W \,\forall\, u, v \in W$ (closed under addition)
3. $\alpha \cdot u \in W \,\forall\, u \in W, \alpha \in \mathbb{F}$[5]

Then, W is a vector space in its own right.

⊛ Example 1.4: Examples of Subspaces

1. Let $V := \mathbb{F}^n$.
   - $W := \{(x_1, x_2, \dots, x_n) \in \mathbb{F}^n : x_1 = 0\} = \{(0, x_2, x_3, \dots, x_n) : x_i \in \mathbb{F}\}$.
   - $W := \{(x_1, x_2, \dots, x_n) \in \mathbb{F}^n : x_1 + 2 \cdot x_2 = 0\}$

     Proof. Let $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in W$. Then, $x + y = (x_1 + y_1, \dots, x_n + y_n)$, and $x_1 + y_1 + 2 \cdot (x_2 + y_2) = x_1 + 2 \cdot x_2 + y_1 + 2 \cdot y_2 = 0 + 0 = 0 \implies x + y \in W$. Similar logic follows for axioms 2., 3. ■

   - (More generally)

     $$W := \{(x_1, \dots, x_n) \in \mathbb{F}^n : \begin{matrix} a_{11}x_1 + \cdots a_{1n}x_n = 0 \\ a_{21}x_1 + \cdots + a_{2n}x_n = 0 \\ \ddots \\ a_{k1}x_1 + \cdots a_{kn}x_n = 0 \end{matrix}\},$$

     that is, a linear combination of homogenous "conditions" on each term.
   - $W^* := \{(x_1, \dots, x_n) : x_1 + x_2 = 1\}$ is *not* a subspace; it is not closed under addition, nor under scalar multiplication.
2. Let $\mathbb{F}[t]_n := \{a_0 + a_1 t + \cdots + a_n t^n : a_i \in \mathbb{F}\}$. Then, $\mathbb{F}[t]_n$ is a subspace of $\mathbb{F}[t]$, the more general polynomial space. *However*, the set of all polynomials of degree *exactly* n (all axioms fail, in fact) is not a subspace of $\mathbb{F}[t]_n$.
   - $W := \{p(t) \in \mathbb{F}[t]_n : p(1) = 0\}$.
   - $W := \{p(t) \in \mathbb{F}[t]_n : p''(t) + p'(t) + 2p(t) = 0\}$.

[4]This is equivalent to requiring that $W \neq \emptyset$; stated this way, axiom 3. would necessitate that $0 \cdot w = 0_V \in W$.

[5]Note that these axioms are equivalent to saying that W is a subgroup of V with respect to vector addition; 2. ensures closed under addition, and 3. ensures the existence of additive inverses (as per $-1 \cdot v = -v$).

3. Let $V := C(\mathbb{R})$ be the space of continuous function $\mathbb{R} \to \mathbb{R}$.
   - $W := \{f \in C(\mathbb{R}) : f(\pi) + 7f(\sqrt{2}) = 0\}$.
   - $W := C^1(\mathbb{R}) :=$ everywhere differentiable functions.
   - $W := \{f \in C(\mathbb{R}) : \int_0^1 f \, \mathrm{d}x = 0\}$.

↪ **Proposition 1.2**

Let W_1, W_2 be subspaces of a vector space V over $\mathbb{F}$. Then, define the following:

1. $W_1 + W_2 := \{w_1 + w_2 : w_1 \in W_1, w_2 \in W_2\}$
2. $W_1 \cap W_2 := \{w \in V : w \in W_1 \wedge w \in W_2\}$

These are both subspaces of V.

Proof.

1. (a) $0_V \in W_1$ and $0_V \in W_2 \implies 0_V = 0_V + 0_V \in W_1 + W_2$.
   (b) $(u_1 + u_2) + (v_1 + v_2) = (u_1 + v_1) + (u_2 + v_2) \in W_1 + W_2$.
   (c) $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v \in W_1 + W_2$
2. (a) $0_V \in W_1$ and $0_V \in W_2 \implies 0_V = 0_V + 0_V \in W_1 \cap W_2$.
   (b) $u, v \in W_1 \cap W_2 \implies u + v \in W_1 \wedge u + v \in W_2 \implies u + v \in W_1 \cap W_2$.
   (c) $\alpha \cdot u \in W_1 \wedge \alpha \cdot u \in W_2 \implies \alpha \cdot u \in W_1 \cap W_2$.

■

1.3 Linear Combinations and Span

↪ **Definition 1.4: Linear Combination**

Let V be a vector space over a field $\mathbb{F}$. For finitely many vectors $v_1, v_2, \ldots, v_n$, their *linear combination* is a sum of the form

$$\sum_{i=1}^{n} a_i v_i = a_1 \cdot v_1 + \cdots + a_n \cdot v_n,$$

where $a_i \in \mathbb{F} \, \forall \, i$.

A linear combination is called *trivial* if $a_i = 0 \, \forall \, i$, that is, all coefficients are 0.

If $n = 0$ (ie, we are "summing up" 0 vectors), we define the sum as the zero vector; $\sum_{i=1}^{0} a_i v_i := 0_V$.

↪ Lecture 02; Last Updated: Mon Mar 25 13:48:03 EDT 2024

↪ **Definition 1.5: A More General Definition of Linear Combination**

For a (possibly infinite) set S of vectors from V, a *linear combination* of vectors in S is a linear combination of $a_1 v_1 + \cdots a_n v_n$ for some finite subset $\{v_1, \ldots, v_n\} \subseteq S$.[6]

↪ **Definition 1.6: Span**

For a subset $S \subseteq V$, we define its *span* as

$$\text{Span}(S) := \text{ set of all linear combinations of } S := \{a_1 v_1 + \cdots a_n v_n : a_i \in \mathbb{F}, v_i \in S\}.$$

By convention, we set $\text{Span}(\varnothing) = \{0_V\}$.

⊛ **Example 1.5**

Let $S := \{(1,0,-1), (0,1,-1), (1,1,-2)\} \subseteq \mathbb{R}^3$. Then,

$$0_{\mathbb{R}^3} = (0,0,0) = 1 \cdot (1,0,-1) + 1 \cdot (0,1,-1) + -1 \cdot (1,1,-2).$$

We claim, moreover, that $\text{Span}(S) = U := \{(x,y,z) \in \mathbb{R}^3 : x + y + z = 0\}$ (a plane through the origin).

Proof. Note that $S \subseteq U$, hence $S \subseteq \text{Span}\, S \subseteq U$. OTOH, if $(x,y,z) \in U$, we have $z = -x - y$, and so

$$(x,y,z) = (x,y,-x-y) = x \cdot (1,0,-1) + y \cdot (0,1,-1) \in \text{Span}(S)$$

hence $U \subseteq \text{Span}(S)$ and thus $\text{Span}(S) = U$. ∎

Remark 1.4. *We implicitly used the following claim in the proof above; we prove it more generally.*

↪ **Proposition 1.3**

Let V be a vector space over $\mathbb{F}$ and let $S \subseteq V$. Then, $\text{Span}(S)$ is always a subspace. Moreover, it is the smallest (minimal) subspace containing S (that is, for any subspace $U \supseteq S$, we have that $U \supseteq \text{Span}\, S$).

Proof. Because adding/scalar multiplying linear combinations of elements of S again results in a linear combination of elements of S, and $0_V \in \text{Span}(S)$ by definition, we have that $\text{Span}(S)$ is indeed a subspace.

If $U \supset S$ is a subspace of V containing S, then by definition U is closed under addition, that is, taking linear combinations of its elements (in particular, of elements of S); hence, $U \supset \text{Span}(S)$. ∎

↪ **Lemma 1.1**

For $S \subseteq V$ and $v \in V$, $v \in \text{Span}(S) \iff \text{Span}(S \cup \{v\}) = \text{Span}(S)$.

[6] That is, we do not allow infinite sums.

Proof. ($\Longrightarrow$) Let $v \in \text{Span}(S) \implies v = a_1v_1 + \cdots a_nv_n, a_i \in \mathbb{F}, v_i \in V$. Then, for any linear combination

$$b_1u_1 + \cdots b_mu_m + b \cdot v = b_1u_1 + \cdots b_mu_m + b(a_1v_1 + \cdots + a_nv_n)$$

is a linear combination of vectors in $S \cup \{v\}$ (first equality) or equivalently, a combination of vectors in S (second equality) and thus $\text{Span}(S \cup \{v\}) \subseteq \text{Span}\, S$. The reverse inclusion follows trivially.

($\Longleftarrow$) $\text{Span}(S \cup \{v\}) = \text{Span}\, S \implies v \in \text{Span}(S)$. ■

⊛ **Example 1.6**

(From the above example) We have

$$\text{Span}(\{(1,0,-1),(0,1,-1)\} \cup \{(1,1,-2)\}) = \text{Span}(\{(1,0,-1),(0,1,-1)\}),$$

since $(1,1,-2) \in \text{Span}(\{(1,0,-1),(0,1,-1)\})$ (it was redundant, as it could be generated by the other two vectors).

↪ **Definition 1.7: Spanning Set**

Let V be a vector space over a field $\mathbb{F}$. We call $S \subseteq V$ a *spanning set* for V if $\text{Span}(S) = V$. We call such a spanning set *minimal* if no proper subset of S is a spanning set ($\nexists v \in S$ s.t. $S \setminus \{v\}$ spanning).

Remark 1.5. *Note that any $S \subseteq V$ is spanning for* Span(S). *But, S may not be minimal; indeed, consider the previous example. We were able to remove a vector from S while having the same span.*

⊛ **Example 1.7**

For $\mathbb{F}^n$ as a vector space over $\mathbb{F}$, the *standard spanning set*

$$\underset{n}{\text{St}} := \{\underbrace{(1,\ldots,0)}_{:=e_1}, \underbrace{(0,1,0,\ldots,0)}_{:=e_2}, \ldots, \underbrace{(0,\ldots,1)}_{e_n}\}.$$

Given any $x := (x_1,\ldots,x_n) \in \mathbb{F}^n$, we can write

$$x = x_1 \cdot e_1 + \cdots x_n \cdot e_n.$$

This is clearly minimal; removing any e_i would then result in a 0 in the ith "coordinate" of a vector, hence $\text{St} \setminus \{e_i\}$ would span only vectors whose ith coordinate is 0.

↪ **Definition 1.8: Linear Dependence**

Let V be a vector space over a field $\mathbb{F}$. A set $S \subseteq V$ is said to be *linearly dependent* if there is a nontrivial linear combination of vectors in S that is equal to 0_V.

Conversely, S is called *linearly independent* if there is no nontrivial linear combination of vectors in S that is equal to 0_V; all linear combinations of vectors in S that equal 0_V are trivial.

↪ Lecture 03; Last Updated: Mon Mar 25 13:48:23 EDT 2024

⊛ **Example 1.8**

1. The empty set $\varnothing$ is linearly independent; there are no non-trivial linear combinations that equal 0_V (there are no linear combinations at all).
2. For $v \in V$, the set $\{v\}$ is linearly dependent iff $v = 0_V$.
3. $S := \{(1,0,-1),(0,1,-1),(1,1,-2)\} := \{v_1, v_2, v_3\}$; S is linearly dependent ($v_1+v_2-v_3 = (0,0,0)$).
4. $V := \mathbb{F}^3$; $S := \{(1,0,-1),(0,1,-1),(0,0,1)\} = \{v_1, v_2, v_3\}$ is linearly independent.

   Proof. Suppose

   $$\begin{aligned} &a_1v_1 + a_2v_2 + a_3v_3 = 0_V \\ &\qquad\implies a_1 = 0 \wedge a_2 = 0 \wedge -a_1 - a_2 + a_3 = 0 \implies a_3 = 0 \\ &\qquad\implies a_1 = a_2 = a_3 = 0 \end{aligned}$$

   Hence only a trivial linear combination is possible. ■

5. St_n is linearly independent.

   Proof.

   $$\sum_{i=1}^{n} a_i e_i = 0_{\mathbb{F}^n} \implies a_i = 0 \,\forall\, i$$

   ■

↪ **Lemma 1.2**

Let V be a vector space over a field $\mathbb{F}$, and $S \subseteq V$ (possibly infinite).

1. S is linearly dependent $\iff$ there is a finite subset $S_0 \subseteq S$ that is linearly dependent.
2. S is linearly independent $\iff$ all finite subsets of S are linearly independent.

Proof. 2. follows from the negation of 1.

($\Longleftarrow$) Trivial.

($\Longrightarrow$) Suppose S linearly dependent. Then, 0_V = some nontrivial linear combination of vectors $v_1, \dots, v_n$ in S. Let $S_0 = \{v_1, \dots, v_n\}$, then, S_0 is linearly dependent itself. ■

1.4 Linear Dependence and Span

↪ Proposition 1.4

Let V be a vector space over a field $\mathbb{F}$ and $S \subseteq V$.

1. S linearly dependent $\iff$ $\exists v \in \text{Span}(S \setminus \{v\})$.
2. S linearly independent $\iff$ there is no $v \in \text{Span}(S \setminus \{v\})$.

Proof. 2. follows from the negation of 1.

($\Longrightarrow$) Suppose S linearly dependent. Then, $0_V = \sum_{i=1}^n a_i v_i$ for some nontrivial linear combination of distinct vectors S. At least one of $a_i \neq 0$; we can assume wlog (reindexing) $a_1 \neq 0$. Then,

$$a_1 v_1 = -\sum_{i=2}^n a_i v_i \implies v_1 = (-a_1^{-1}) \sum_{i=2}^n a_i v_i = \sum_{i=2}^n (-a_1^{-1} a_i) v_i,$$

hence, $v_1 \in \text{Span}(\{v_2, \dots, v_n\}) \subseteq \text{Span}(S \setminus \{v\})$

($\Longleftarrow$) Suppose $v \in \text{Span}(S \setminus \{v\})$, then $v = a_1 v_1 + \cdots + a_n v_n$, with $v_1, \dots, v_n \in S \setminus \{v\}$, thus

$$0_V = a_1 v_1 + \cdots a_n v_n - v,$$

which is not a trivial combination (-1 on the v; v cannot "merge" with the other vectors), hence S is linearly dependent. ■

↪ Corollary 1.1

$S \subseteq V$ is linearly independent $\iff$ S a minimal spanning set of Span S.

Proof. Follows from proposition 1.4, 2. ■

↪ Definition 1.9: Maximally Independent

Let V be a vector space over a field $\mathbb{F}$. A set $S \subseteq V$ is called *maximally independent* if S is linearly independent and $\nexists v \in V \setminus S$ s.t. $S \cup \{v\}$ is still linearly independent.

In other words, there is no proper supset $\tilde{S} \supsetneq S$ that is still independent.

↪ **Lemma 1.3**

If $S \subseteq V$ maximally independent, then S is spanning for V.

Proof. Let $S \subseteq V$ be maximally independent. Let $v \in V$; supposing $v \notin S$ (in the case that $v \in S$, then $v \in \text{Span}(S)$ trivially). By maximality, $S \cup \{v\}$ is linearly dependent, hence there exists a nontrivial linear combination that equals 0_V. Since S independent, this combination must include v, with a nonzero coefficient. We can write

$$\begin{aligned} av + \sum_{i=1}^{n} a_i v_i = 0_V \quad a \neq 0, v_i \in S \\ \implies v = \sum_{i=1}^{n} (-a^{-1} a_i) v_i \in \text{Span}\, S. \end{aligned}$$

■

↪ **Theorem 1.1**

Let V be a vector space over a field $\mathbb{F}$ and let $S \subseteq V$. TFAE:

1. S is a minimal spanning set;
2. S is linearly independent and spanning;
3. S is a maximally linearly independent set;
4. Every vector in V is equal to *unique* linear combination of vectors in S.

↪ *Lecture 04; Last Updated: Mon Mar 25 13:48:03 EDT 2024*

Proof. (1. $\implies$ 2.) Suppose S is spanning for V and is minimal. Then, by corollary 1.1, we have that S is linearly independent, and is thus both linearly independent and spanning.

(2. $\implies$ 3.) Suppose S is linearly independent and spanning. Let $v \in V \setminus S$; S is spanning, hence $v \in \text{Span}\, S$, that is, there exists a linear combination of vectors in S that is equal to v:

$$v = a_1 v_1 + \cdots + a_n v_n, a_i \in \mathbb{F}, v_i \in S.$$

Thus, $0_V = a_1 v_1 + \cdots + a_n v_n - v$, thus $S \cup \{v\}$ is linearly dependent, and so S is maximally linearly independent.

(3. $\implies$ 1.) Suppose S is maximally linearly independent. By lemma 1.3, S is spanning, and since S is linearly independent, by corollary 1.1, S is minimally spanning for $\text{Span}\, S$.

(2. $\implies$ 4.) Suppose S is linearly independent and spans V, and let $v \in V$. We have that $v \in \text{Span}\, S$ and hence is equal to a linear combination of vectors in S. This gives existence; we now need to prove uniqueness.

Suppose there exist two linear combinations that equal v,

$$v = a_1 v_1 + \cdots + a_n v_n = b_1 u_1 + \cdots + b_m u_m,$$

$a_i, b_j \in \mathbb{F}$, $v_i, u_j \in S$. With appropriate reindexing/relabelling and allowing certain scalars to equal 0, we can assume that the combinations use the same vectors (with potentially different coefficients), that is,

$$v = a_1 w_1 + \cdots + a_k w_k = b_1 w_1 + \cdots + a_k w_k.$$

This implies, then,

$$(a_1 - b_1)w_1 + \cdots + (a_k - b_k)w_k = 0_V,$$

and by the assumed linear independent of S, each coefficient $(a_i - b_i) = 0 \,\forall\, i \implies a_i = b_i \,\forall\, i$, hence, these are indeed the same representations, and thus this representation is unique.

($4. \implies 2.$) Suppose every vector in V admits a unique linear combination of vectors in S. Clearly, then, S is spanning. It remains to show S is linearly independent. Suppose

$$0_V = a_1 v_1 + \cdots + a_n v_n$$

for $v_i \in S$. But we have that every vector has a unique representation, and we know that $a_i = 0 \,\forall\, i$ is a (valid) linear combination that gives 0_V; hence, this must be the unique combination, $a_i = 0 \,\forall\, i$, and the linear combination above is trivial. Hence, S is linearly independent and spanning. ■

$\hookrightarrow$ **Definition 1.10: Basis**

If any (hence all) of the above statements hold, we call S a *basis* for V.

In the words of 4., we call the unique linear combination of vectors in S that is equal to v the *unique representation of v in S*. Its coefficients are called the *Fourier coefficients of v in S*.

⊛ **Example 1.9**

1. $\mathrm{St}_n = \{e_i : 1 \leqslant i \leqslant n\}$ is a basis for $\mathbb{F}^n$.
2. In $\mathbb{F}^3$, the set
$$\{(1, 0, -1), (0, 1, -1), (0, 0, 1)\}$$
is a basis; it is linearly independent and spanning.
3. For $\mathbb{F}[t]_n$, the standard basis is
$$\{1, t, t^2, \ldots, t^n\}.$$
4. For $\mathbb{F}[t]$, the standard basis is
$$S := \{1, t, t^2, \ldots\} = \{t^n : n \in \mathbb{N}\}.$$
5. Let $\mathbb{F}[\![t]\!]$ denote the space of all formal power series $\sum_{n \in \mathbb{N}} a_n t^n$; polynomials are an example, but with only finite nonzero coefficients. Note that, then, the set S defined above is not a basis for this "extended" set. We *can* in fact find a basis for this set; we need more tools first.

$\hookrightarrow$ Theorem 1.2

Every vector space has a basis.

Remark 1.6. *This theorem relies on assuming the Axiom of Choice.*

$\hookrightarrow$ Lecture 05; Last Updated: Mon Mar 25 13:48:03 EDT 2024

Proof (Attempt). (Of theorem 1.2) We will try to "inductively" build a maximally independent set, as follows:

Begin with an empty set $S_0 := \emptyset$, and iteratively add more vectors to it. Let $v_0 \in V$ be a non-zero vector, and let $S_1 := \{v_0\}$.

If S_1 is maximal, then we are done. Otherwise, there exists a new vector $v_1 \in V \setminus S_1$ s.t. $S_2 := \{v_0, v_1\}$ is still independent.

If S_2 is maximal, then we are done. Otherwise, there exists a new vector $v_2 \in V \setminus S_2$ s.t. $S_3 := \{v_0, v_1, v_2\}$ is still independent.

Continue in this manner; this would take arbitrarily many finite, or even infinite, steps; we would need some "choice function" that would "allow" us to choose any particular ith vector v_i.

We can make this construction precise via the Axiom of Choice and transfinite induction (on ordinals); alternatively, we will prove a statement equivalent to the Axiom of Choice, Zorn's Lemma. ■

Remark 1.7. *Before stating Zorn's Lemma, we introduce the following terminology.*

$\hookrightarrow$ Axiom 1.1: Axiom of Choice

Let X be a set of nonempty sets. Then, there exists a choice function f defined on X that maps each set of X to an element of that set.

$\hookrightarrow$ Definition 1.11: Inclusion-Maximal Element

A *inclusion-maximal* element of I is a set $S \in I$ s.t. there is no strict super set $S' \supsetneq S$ s.t. $S' \in I$.

$\hookrightarrow$ Definition 1.12: Chain

Let X a set. Call a collection $C \subseteq \mathcal{P}(X)$ a *chain* if any two $A, B \in C$ are comparable, ie, $A \subseteq B$ or $B \subseteq A$.

$\hookrightarrow$ Definition 1.13: Upper Bound

An *upper bound* of a collection $\tau \subseteq \mathcal{P}(X)$ is a set $U \subseteq X$ s.t. $U \supseteq J \, \forall J \in \tau$; U contains the union of all sets in J.

⊛ Example 1.10: Of The Previous Definitions

Let $X := \mathbb{N}, I := \{\emptyset, \{0\}, \{1,2\}, \{1,2,3\}\} \subseteq \mathcal{P}(\mathbb{N})$.

The maximal elements of I would be $\{0\}$ and $\{1,2,3\}$.

Chains would include $C_0 := \{\varnothing, \{1,2\}, \{1,2,3\}\}, C_1 := \{\varnothing, \{0\}\}, C_2 := \{\varnothing\}$ (or any set containing a single element).

The sets $\{0,1,2,3\}$ and $\{0,1,2,3,4,5\}$ are upper bounds for I, while neither is an element of I. The set $\{1,2,3\}$ is an upper bound for C_0. A chain $\{\varnothing, \{0\}, \{0,1\}, \{0,1,2\}, \dots\}$ has an upper bound of $\mathbb{N}$.

↪ Lemma 1.4: Zorn's Lemma

Let X be an ambient set and $I \subseteq \mathcal{P}(X)$ be a nonempty collection of subsets of X. If every chain $C \subseteq I$ has an upper bound in I, then I has a maximal element.

"Proof". This is equivalent to the Axiom of Choice; proving it is beyond the scope of this course :(. ■

Proof of theorem 1.2, cnt'd. We obtain a maximal independent set using Zorn's Lemma.

Let I be the collection of all linearly independent subsets of V. I is nonempty; $\varnothing \in I$, as is $\{v\} \in I$ for any nonzero $v \in V$. To apply Zorn's, we need to show that every chain C if sets in I has an upper bound in I; that is, every linearly independent set has an upper bound that itself is linearly independent.

Let C be a chain in I. Let $S := \bigcup C$ be the union of all sets in C. To show S is linearly independent, it suffices to show that every finite subset $\{v_1, \dots, v_n\} \subseteq S$ is linearly independent. Let $S_i \in C$ be s.t. $v_i \in S_i$ for each i. Because C a chain, for each i, j we have either $S_i \subseteq S_j$ or $S_j \subseteq S_i$, and so we can order $S_1, \dots, S_n$ in increasing order w.r.t $\subseteq$. This implies, then, there is a maximal S_{i_0} s.t. $S_{i_0} \supseteq S_i \, \forall \, i \in \{1, \dots, n\}$. Moreover, we have that $\{v_1, \dots, v_n\} \in S_{i_0}$, and that S_{i_0} is linearly independent and thus $\{v_1, v_2, \dots, v_n\}$ is also linearly independent.

Thus, as we can apply Zorn's Lemma, we conclude that I has a maximal element, ie, there is a maximal independent set, and thus a V indeed has a basis. ■

↪ Lecture 06; Last Updated: Mon Mar 25 13:48:03 EDT 2024

↪ Theorem 1.3

For every vector space V over a field $\mathbb{F}$, any two bases $\mathcal{B}_1, \mathcal{B}_2$ are equinumerous/of equal size/cardinality, ie, there is a bijection between $\mathcal{B}_1$ and $\mathcal{B}_2$.

Remark 1.8. *We will only prove this for vector spaces that admit a finite basis.*

↪ Lemma 1.5: Steinitz Substitution

Let V be a vector space over a field $\mathbb{F}$. Let $Y \subseteq V$ be a (possibly infinite) linearly independent set and let $Z \subseteq V$ be a finite spanning set. Then:

1. $k := |Y| \leqslant |Z| =: n$
2. There is $Z' \subseteq Z$ of size $n - k$ s.t. $Y \cup Z'$ is still spanning.

Proof. Remark first that if Z finite and spanning for V, then we cannot have a infinite linearly independent Y subset of V. Thus, wlog assume that Y finite.

We prove by induction on k.

$k = 0$ gives that $Y = \emptyset$, and so $Z' = Z$ itself works ($Z' \cup Y = Z$) as a spanning set.

Suppose the statement holds for some $k \geqslant 0$. Let Y be an independent set such that $|Y| = k + 1$, ie

$$Y := \{y_1, y_2, \dots, y_k, y_{k+1}\}, \quad y \in V.$$

By our inductive assumption, we can consider $Y' := \{y_1, \dots, y_k\} \subseteq Y$ of size k, to obtain a set

$$Z' = \{z_1, z_2, \dots, z_{n-k}\} \subseteq Z, \text{ s.t. } Y' \cup Z' = \{y_1, \dots, y_k, z_1, \dots, z_{n-k}\}$$

is spanning. As this is spanning, we can write y_{k+1} as a linear combination of vectors in $Y' \cup Z'$, ie

$$y_{k+1} = a_1 y_1 + \cdots + a_k y_k + b_1 z_1 + \cdots + b_{n-k} z_{n-k}, \quad a_i, b_j \in \mathbb{F}.$$

It must be that at least one of b_j's must be nonzero; if they were all zero, then y_{k+1} would simply be a linear combination of vector y_i giving that y_{k+1} linearly dependent, contradicting our construction of Y linearly independent.

Assume, wlog, $b_{n-k} \neq 0$. Then, we can write

$$z_{n-k} = b_{n-k}^{-1} y_{k+1} - b_{n-k}^{-1} a_1 y_1 - \cdots - b_{n-k}^{-1} a_k y_k - b_{n-k}^{-1} b_1 z_1 - \cdots - b_{n-k}^{-1} b_{n-k-1} z_{n-k-1},$$

and hence

$$z_{n-k} \in \mathrm{Span}\{y_1, \dots, y_{k+1}, z_1, \dots, z_{n-k-1}\} = \mathrm{Span}\left(\underbrace{\{y_1, \dots, y_{k+1}\}}_{Y} \cup \underbrace{\{z_1, \dots, z_{n-k-1}\}}_{:=Z''} \right).$$

We had that $Y' \cup Z'$ was spanning, and $(Y' \cup Z') \setminus (Y \cup Z'') = \{z_{n-k}\} \subseteq \mathrm{Span}(Y \cup Z'')$, and we thus have that $Y \cup Z''$ is also spanning. ■

↪ Corollary 1.2: Finite Basis Case for theorem 1.3

Let V be a vector space that admits a finite basis. Then, any two bases of V are equinumerous.

Proof. Let Y, Z be two finite bases for V. Then, Y is independent and Z is spanning, so by Steinitz Substitution, $|Y| \leqslant |Z|$. OTOH, Z is independent, and Y is spanning, so by Steinitz Substitution, $|Z| \leqslant |Y|$, and we conclude that $|Y| = |Z|$. Let $n := |Y|$.

It remains to show that there exist no infinite bases for V; it suffices to show that there is no independent set of size $n + 1$. To this end, let $I \subseteq V$ such that $|I| = n + 1$ be an independent set. Y is still spanning, hence, by the substitution lemma, $n + 1 \leqslant n$, a contradiction. Hence, I as defined cannot exist and so any basis of V must be of size n. ■

↪ **Definition 1.14: Dimension**

Let V be a vector space over a field $\mathbb{F}$. The *dimension* of V, denote

$$\dim(V)$$

as the cardinality/size of any basis for V. We call V *finite dimensional* if $\dim(V)$ is a natural number, i.e. V admits a finite basis. Otherwise, we say V is infinite dimensional.

↪ **Corollary 1.3: of Steinitz Substitution**

Let V be a finite dimensional vector space over $\mathbb{F}$ and denote $n \coloneqq \dim(V)$. Then:

1. Every linearly independent subset $I \subseteq V$ has size $\leqslant n$;
2. Every spanning set $S \subseteq V$ for V has size $\geqslant n$;
3. Every independent set I can be completed to a basis to V, ie, there exists a basis B for V s.t. $I \subseteq B$.

Proof. Fix a basis B for V, $|B| =: n$.

1. If I is a independent set, then because B spanning, Steinitz Substitution gives $|I| \leqslant |B|$.
2. If S spanning for V, then because B is linearly independent, Steinitz Substitution gives $|B| \leqslant |S|$.
3. Let I be an independent set. Then, because B is spanning, Steinitz Substitution gives $B' \subseteq B$ of size $n - |I|$ s.t. $I \cup B'$ is spanning. Moreover, $|I \cup B'| \leqslant n$, and by 2. it must have size $\geqslant n$, and thus has size precisely n and is thus a minimally spanning set and thus a basis.

■

↪ **Corollary 1.4: Monotonicity of Dimension**

Let V be a vector space over a field $\mathbb{F}$. For any subspace $W \subseteq$, $\dim W \leqslant \dim V$, and

$$\dim W = \dim V \iff W = V.$$

Proof. Let $B \subseteq W$ be a basis for W. Because B is independent, $|B| \leqslant \dim(V)$ by 1. of corollary 1.3, so $\dim(W) = |B| \leqslant \dim(V)$.

If $|B| = \dim(V)$, then B is a basis for V again by 1. of corollary 1.3, so $W = \mathrm{Span}(B) = V$. ■

↪ *Lecture 07; Last Updated: Mon Mar 25 13:48:03 EDT 2024*

2 Linear Transformations, Matrices

2.1 Introduction: Definitions, Basic Properties

↪ **Definition 2.1: Linear Transformation**

Let V, W be vector spaces over a field $\mathbb{F}$. A function $T : V \to W$ is called a *linear transformation* if it preserves the vector space structures, that is,

1. $T(v_0 + v_1) = T(v_0) + T(v_1)$, $\forall\, v_0, v_1 \in V$;
2. $T(\alpha \cdot v) = \alpha \cdot T(v)$, $\forall\, \alpha \in \mathbb{F}, v \in V$;
3. $T(0_V) = 0_W$.

Remark 2.1. *Note that 3. is redundant, implied by 2., but included for emphasis:*

$$T(0_V) = T(0_{\mathbb{F}} \cdot 0_V) = 0_{\mathbb{F}} \cdot T(0_V) = 0_W.$$

⊛ **Example 2.1: Linear Transformations**

1. $T : \mathbb{F}^2 \to \mathbb{F}^2, T(a_1, a_2) := (a_1 + 2a_2, a_1)$.
2. Let $\theta \in \mathbb{R}$, and let $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the rotation by θ. The linearity of this is perhaps most obvious in polar coordinates, ie $v \in \mathbb{R}^2, v = r(\cos\alpha, \sin\alpha)$ for appropriate r, α, and $T_\theta(v) = r(\cos(\alpha + \theta), \sin(\alpha + \theta))$.
3. $T : \mathbb{R}^2 \to \mathbb{R}^2$, a reflection about the x-axis, ie, $T(x, y) = (x, -y)$.
4. Projections, $T : \mathbb{F}^n \to \mathbb{F}^n$.
5. The transpose on $M_n(\mathbb{F})$, ie, $T : M_n(\mathbb{F}) \to M_n(\mathbb{F})$, where $A \mapsto A^t$.
6. The derivative on space of polynomials of degree leq n, $D : \mathbb{F}[t]_{n+1} \to \mathbb{F}[t]_n, p(t) \mapsto p'(t)$.

↪ **Theorem 2.1**

Linear transformations are completely determined by their values on a basis.

That is, let $\mathcal{B} := \{v_1, \dots, v_n\}$ be a basis for a vector space V over $\mathbb{F}$. Let W also be a vector space over $\mathbb{F}$ and let $w_1, \dots, w_n \in W$ be arbitrary vectors. Then, there is a unique linear transformation $T : V \to W$ s.t. $T(v_i) = w_i \,\forall\, i = 1, \dots, n$.

Proof. We aim to define $T(v)$ for arbitrary $v \in V$. We can write

$$v = a_1 v_1 + \cdots + a_n v_n$$

as the unique representation of v in terms of the basis $\mathcal{B}$. Then, we simply define

$$T(v) \coloneqq a_1 w_1 + \cdots + a_n w_n,$$

for our given w_i's. Then, $T(v_i) = 1 \cdot w_i = w_i$, as desired, and T is linear;

1. Let $u, v \in V; u \coloneqq \sum_n a_i v_i, v \coloneqq \sum_n b_i v_i$. Then,

$$T(u+v) = T(\sum_n a_i v_i + \sum_n b_i v_i) = T(\sum_n (a_i + b_i) v_i) = \sum_n (a_i + b_i) w_i = \sum_n a_i w_i + \sum_n b_i w_i = T(u) + T(v).$$

2. Scalar multiplication follows similarly.

To show uniqueness, suppose T_0, T_1 are two linear transformations satisfying $T_0(v_i) = w_i = T_1(v_i)$. Let $v \in V$, and write $v = \sum_n a_i v_i$. By linearity,

$$T_k(v) = T_k(\sum_n a_i v_i) = \sum_n a_i T(v_i) = \sum_n a_i w_i,$$

for $k = 0, 1$, hence, $T_1(v) = T_0(v)$ for arbitrary v, hence the transformations are equivalent. ■

↪ **Definition 2.2: Some Important Transformations**

We denote $T_0 : V \to W$ by $T_0(v) \coloneqq 0_W \, \forall v \in V$ the *zero transformation*. We denote $I_V : V \to V$, $I_V(v) \coloneqq v \, \forall v \in V$, as the *identity transformation*.

↪ *Lecture 08; Last Updated: Sat Apr 6 12:28:02 EDT 2024*

2.2 Isomorphisms, Kernel, Image

↪ **Definition 2.3: Isomorphism**

Let V, W be vector spaces over $\mathbb{F}$. An *isomorphism* from V to W is a linear transformation $T : V \to W$ (a homomorphism for vector spaces) which admits an inverse T^{-1} that is also linear.

If such an isomorphism exists, we say V and W are *isomorphic*.

↪ **Proposition 2.1**

$T : V \to W$ is an isomorphism $\iff$ T is linear and bijective.

Proof. The direction $\implies$ is trivial.

Suppose $T : V \to W$ is linear and bijective, ie T^{-1} exists. We need to show that T^{-1} is linear. Let $w_1, w_2 \in W, a_1, a_2 \in \mathbb{F}$. Then:

$$\begin{aligned} T^{-1}(a_1 w_1 + a_2 w_2) &= T^{-1}(a_1 T(T^{-1}(w_1)) + a_2 T(T^{-1}(w_2))) \\ \textit{(by linearity of } T\textit{)} \quad &= T^{-1}(T(a_1 T^{-1}(w_1) + a_2 T^{-1}(w_2))) \\ &= a_1 T^{-1}(w_1) + a_2 T^{-1}(w_2). \end{aligned}$$

■

Remark 2.2. *This proposition holds for all structures that only have operations; it does not for those with relations, such as graphs, orders, etc..*

↪ **Theorem 2.2**

For $n \in \mathbb{N}$, every n-dimensional vector space V over $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$. In particular, all n-dim vector spaces over $\mathbb{F}$ are isomorphic.

Proof. Fix a basis $\mathcal{B} := \{v_1, \ldots, v_n\}$ for V, and let $T : V \to \mathbb{F}^n$ be the unique linear transformation determined by $\mathcal{B}$ with $T(v_i) = e_i$, where $\{e_1, \ldots, e_n\}$ is the standard basis for $\mathbb{F}^n$. We show that T is a bijection.

(Injective) Suppose $T(x) = T(y), x, y \in V$. Write $x = a_1 v_1 + \cdots + a_n v_n, y = b_1 v_1 + \cdots + b_n v_n$, the unique representation of x, y in the basis $\mathcal{B}$. We have:

$$a_1 e_1 + \cdots + a_n e_n = a_1 T(v_1) + \cdots + a_n T(v_n) = T(a_1 v_1 + \cdots + a_n v_n) = T(x) = T(y) = \cdots = b_1 e_1 + \cdots + b_n e_n,$$

but by the uniqueness of representation in a basis, it follows that each $a_i = b_i$, hence, $x = y$.

(Surjective) Let $w \in \mathbb{F}^n$. Then, $w = a_1 e_1 + \cdots + a_n e_n$ (uniquely). But then,

$$w = a_1 T(v_1) + \cdots + a_n T(v_n) = T(a_1 v_1 + \cdots + a_n v_n),$$

where $a_1 v_1 + \cdots + a_n v_n \in V$, hence T indeed surjective. ■

Remark 2.3. *Replacing $\mathbb{F}^n$ with an arbitrary n-dim vector space W over $\mathbb{F}$ yields the following.*

↪ **Theorem 2.3: Freeness of Vector Spaces**

Let W, V be vector spaces over $\mathbb{F}$ and let β, γ be bases for V, W respectively. Every bijection $T : \beta \to \gamma$ can be extended to an isomorphism $\hat{T} : V \to W$.

In particular, all vector spaces over $\mathbb{F}$ with equinumerous bases are isomorphic.

Remark 2.4. *The proof follows very similarly to the previous theorem, but extended to arbitrary, possible infinite, spaces.*

Proof. Homework exercise. ■

↪ **Definition 2.4: Image/Kernel**

For a linear transformation $T : V \to W$, where V, W are vector spaces over $\mathbb{F}$, we define the *image*

$$\operatorname{Im}(T) := T(V),$$

and its *kernel*

$$\operatorname{Ker}(T) := T^{-1}(\{0_W\}).$$

↪ **Proposition 2.2**

$\operatorname{Ker}(T)$ and $\operatorname{Im}(T)$ are subspaces of V, W resp.

Proof. ($\operatorname{Ker}(T)$) Let $v_0, v_1 \in \operatorname{Ker} T$ and $a_0, a_1 \in \mathbb{F}$, then

$$T(a_0v_0 + a_1v_1) = a_0T(v_0) + a_1T(v_1) = 0_W \implies a_0v_0 + a_1v_1 \in \operatorname{Ker} T.$$

($\operatorname{Im}(T)$) Let $w_0, w_1 \in \operatorname{Im} T$, $a_0, a_1 \in \mathbb{F}$. Then $w_i = T(v_i), v_i \in V$, and so

$$a_0w_0 + a_1w_1 = a_0T(v_0) + a_1T(v_1) = T(a_0v_0 + a_1v_1) \implies a_0w_0 + a_1w_1 \in \operatorname{Im} T.$$

■

↪ **Proposition 2.3**

Let $T : V \to W$ be a linear transformation, where V, W vector spaces over $\mathbb{F}$. Let β be a (possibly infinite) basis for V. Then, $T(\beta)$ spans $\operatorname{Im}(T)$.

In particular, T is surjective iff $T(\beta)$ spans W.

Proof. Let $w \in \operatorname{Im}(T)$, so $w = T(v)$ for some $v \in V$, where we have $v := a_1v_1 + \cdots + a_nv_n, v_i \in \beta$. Then,

$$w = T(v) = a_1T(v_1) + \cdots + a_nT(v_n) \in \operatorname{Span}(\{T(v_1), \ldots, T(v_n)\}) \subseteq \operatorname{Span}(T(\beta)).$$

■

↪ Lecture 09; Last Updated: Fri Apr 5 08:27:50 EDT 2024

$\hookrightarrow$ **Proposition 2.4**

Let $T : V \to W$ be a linear transformation, where V, W vector spaces over $\mathbb{F}$. TFAE:

1. T is injective.
2. $\mathrm{Ker}(T)$ is the trivial subspace $\{0_V\}$.
3. $T(\beta)$ is independent for each basis β for V.

3'. $T(\beta)$ is independent for some basis β for V.

Proof. (1. $\implies$ 2.) Trivial; only 0_V can be mapped to 0_W.

(2. $\implies$ 1.) Suppose $\mathrm{Ker}(T) = \{0_V\}$ and let $T(x) = T(y), x, y \in V$. By linearity,

$$T(x-y) = T(x) - T(y) = 0_W \implies x - y \in \mathrm{Ker}(T) \implies x - y = 0_V \implies x = y.$$

(2. $\implies$ 3.) Fix a basis β for V. To show that $T(\beta)$ linearly independent, take an arbitrary linear combination $a_1 w_1 + \cdots + a_n w_n \in T(\beta)$. Suppose $\sum_i a_i w_i = 0_W$. Since $w_i \in T(\beta)$, $w_i = T(v_i), v_i \in \beta$, hence

$$\begin{aligned} 0_W = a_1 w_1 + \cdots + a_n w_n = a_1 T(v_1) + \cdots + a_n T(v_n) &= T(a_1 v_1 + \cdots + a_n v_n) \\ &\implies a_1 v_1 + \cdots + a_n v_n \in \mathrm{Ker}(T) \\ &\implies a_1 v_1 + \cdots + a_n v_n = 0_V, \end{aligned}$$

but each v_i is linearly independent, hence this must be a trivial linear combination, and thus $a_i = 0 \,\forall\, i$.

(3) $\implies$ (3') Trivial; stronger statement implies weaker statement.

(3') $\implies$ (2) Suppose $T(\beta)$ linearly independent for some basis β for V. Suppose $T(v) = 0_W, v \in V$. We write

$$v = a_1 v_1 + \cdots a_n v_n, v_i \in \beta.$$

Then,

$$0_W = T(v) = T(a_1 v_1 + \cdots + a_n v_n) = a_1 T(v_1) + \cdots + a_n T(v_n),$$

but $\{T(v_i)\} \subseteq T(\beta)$ is linearly independent, hence, this combination must be trivial and each $a_i = 0$, and thus $v = 0_V$ and so $\mathrm{Ker}(T) = \{0_V\}$ is trivial. ■

↪ **Definition 2.5: Rank, nullity**

Let V, W be vector spaces over $\mathbb{F}$ and $T : V \to W$ be linear. Define *rank* of T as

$$\operatorname{rank}(T) := \dim(\operatorname{Im}(T)),$$

and *nullity* of T as

$$\operatorname{nullity}(T) := \dim(\operatorname{Ker}(T)).$$

↪ **Theorem 2.4: Rank-Nullity Theorem**

Let V, W be vector spaces over $\mathbb{F}$, $\dim(V) < \infty$. Let $T : V \to W$ be a linear transformation. Then,

$$\operatorname{nullity}(T) + \operatorname{rank}(T) = \dim(V).$$

Remark 2.5. *Intuitively: the nullity is the number of vectors we "collapse"; the rank is what is left. Together, we have the entire space.*

Remark 2.6. *This follows directly from the first isomorphism theorem for vector spaces, and the fact that* $\dim(V/\operatorname{Ker}(T)) = \dim(V) - \dim(\operatorname{Ker}(T))$*; however, we will prove it without this result below.*

Proof. Let $\{v_1, \ldots, v_k\}$ be a basis for $\operatorname{Ker}(T)$, and complete it to a basis $\beta := \{v_1, \ldots, v_k, u_1, \ldots, u_{n-k}\}$ for V, where $n := \dim(V)$. We need to show that $\dim(\operatorname{Im}(T)) = n - k$.

Recall that $\{T(v_1), \ldots, T(v_k), T(u_1), \ldots, T(u_{n-k})\}$ spans $\operatorname{Im}(T)$. But $v_1, \ldots, v_k \in \operatorname{Ker}(T)$, so $T(v_i) = 0_W \ \forall\, i = 1, \ldots, k$. Hence, letting $\gamma := \{T(u_1), \ldots, T(u_{n-k})\}$ spans $\operatorname{Im}(T)$. It remains to show that γ is independent.

Let $a_1 T(u_1) + \cdots a_{n-k} T(u_{n-k}) = 0_W$; by linearity,

$$\begin{aligned} T(a_1 u_1 + \cdots + a_{n-k} u_{n-k}) = 0_W \\ \implies a_1 u_1 + \cdots + a_{n-k} u_{n-k} \in \operatorname{Ker}(T) \\ \implies a_1 u_1 + \cdots + a_{n-k} u_{n-k} = b_1 v_1 + \cdots + b_k v_k, \end{aligned}$$

but each of these $u_i, v_j \in \beta$, hence, each coefficient must be identically zero as β linearly independent, and thus $\dim(\operatorname{Im}(T)) = n - k$. This completes the proof. ■

↪ **Corollary 2.1: Pigeonhole Principle for Dimension**

Let $T : V \to W$ be a linear transformation. If T injective, then $\dim(W) \geqslant \dim(V)$.

Proof. If $\dim(V) < \infty$, then $\dim(\operatorname{Im}(T)) = \dim(V)$, and we have that $\dim(\operatorname{Im}(T)) \leqslant \dim(W)$ and conclude $\dim(V) \leqslant \dim(W)$.

If $\dim(V) = \infty$, then $\dim(\operatorname{Im}(T)) = \infty$ and $\dim(W) \geqslant \dim(\operatorname{Im}(T)) = \infty$. ■

↪ **Corollary 2.2**

Let $n \in \mathbb{N}$ and V, W be n-dimensional vector spaces over $\mathbb{F}$. For a linear transformation $T : V \longrightarrow W$, TFAE:

1. T injective;
2. T surjective;
3. $\operatorname{rank}(T) = n$.

Proof. (2. $\iff$ 3.) Follows from $\operatorname{rank}(T) = \dim(\operatorname{Im}(T)) = n \iff \operatorname{Im}(T) = W$.

(1. $\implies$ 3.) We have $\operatorname{nullity}(T) = 0$ so $\operatorname{rank}(T) = \dim(V) = n$.

(3. $\implies$ 1.) If $\operatorname{rank}(T) = n$, then $\operatorname{nullity}(T) = 0$. ■

↪ Lecture 10; Last Updated: Mon Mar 25 13:48:03 EDT 2024

↪ **Theorem 2.5: First Isomorphism Theorem for Vector Spaces**

Let V, W be vector spaces over $\mathbb{F}$. Let $T : V \longrightarrow W$ be a linear transformation. Then,

$$V/\operatorname{Ker}(T) \cong \operatorname{Im}(T),$$

by the isomorphism given by $v + \operatorname{Ker}(T) \mapsto T(v)$.

Proof. From group theory, we know that $\hat{T} : V/\operatorname{Ker}(T) \longrightarrow \operatorname{Im}(T)$, where $\hat{T}(v + \operatorname{Ker}(T)) := T(v)$ is well-defined, and is an isomorphism of abelian groups. We need only to check that $\hat{T}$ is linear, namely, that is respects scalar multiplication. We have

$$\begin{aligned}
\hat{T}(a \cdot (v + \operatorname{Ker}(T))) &= \hat{T}((a \cdot v) + \operatorname{Ker}(T)) \\
&= T(av) = a \cdot T(v) \\
&= a\hat{T}(v + \operatorname{Ker}(T)),
\end{aligned}$$

as desired. ■

2.3 The Space $\operatorname{Hom}(V, W)$

↪ **Definition 2.6: Homomorphism Space**

For vector spaces V, W over $\mathbb{F}$, let $\operatorname{Hom}(V, W)$ (also denoted $\ell(V, W)$) denote the set of all linear transformations from V to W. We can turn this into a vector space over $\mathbb{F}$ as follows:

1. *Addition of linear transformations:* for $T_0, T_1 \in \operatorname{Hom}(V, W)$, define

$$(T_0 + T_1) : V \to W, \quad v \mapsto T_0(v) + T_1(v).$$

   $(T_0 + T_1)$ is clearly a linear transformation, as the linear combination of linear transformations T_0, T_1.

2. *Scalar multiplication of linear transformations:* for $T \in \operatorname{Hom}(V, W)$, $a \in \mathbb{F}$, define

$$(a \cdot T) : V \to W, \quad v \mapsto a \cdot T(v),$$

   which is again clearly linear in its own right.

↪ **Proposition 2.5**

Endowed with the operations described above, $\operatorname{Hom}(V, W)$ is a vector space over $\mathbb{F}$.

Proof. Follows easily from the definitions. ■

↪ **Theorem 2.6: Basis for** $\operatorname{Hom}(V, W)$

For vector spaces V, W over $\mathbb{F}$ and bases β, γ for V, W resp., the following set

$$\{T_{v,w} : v \in \beta, w \in \gamma\},$$

is a basis for $\operatorname{Hom}(V, W)$, where for each $v \in \beta$ and $w \in \gamma$, $T_{v,w} \in \operatorname{Hom}(V, W)$ defined as the unique linear transformation such that

$$T_{v,w}(v') = \begin{cases} w & v' = v \\ 0_W & v' \neq v \iff \in \beta \setminus \{v\} \end{cases}.$$

Proof. Left as a (homework) exercise. ■

↪ **Corollary 2.3**

If V, W finite dimensional, then $\dim(\operatorname{Hom}(V, W)) = \dim(V) \cdot \dim(W)$.

↪ **Proposition 2.6**

Let $\beta = \{v_1, \ldots, v_n\}$, $\gamma = \{w_1, \ldots, w_m\}$ be bases for V, W resp. Then, by theorem 2.6,

$$\{T_{v_i, w_j} : i \in \{1, \ldots, n\}, j \in \{1, \ldots, m\}\}$$

is a basis for $\mathrm{Hom}(V, W)$, and it has $n \cdot m$ vectors by construction.

2.4 Matrix Representation of Linear Transformations, Finite Fields

Consider a linear transformation $T : \mathbb{F}^n \to \mathbb{F}^m$ between finite fields. We know that T is uniquely determined by its value of basis vectors, so fix the standard bases

$$\beta = \{e_1^{(n)}, \ldots, e_n^{(n)}\} = \{v_1, \ldots, v_n\},$$

and note that T is determined by $\{T(v_1), \ldots, T(v_n)\} \subseteq \mathbb{F}^m$.

Remark 2.7. *We denote vectors in $\mathbb{F}^n$ as column vectors, ie* $\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{F}^n$.

Each $T(v_i)$ is a column vector in $\mathbb{F}^m$, and we an put these into a $m \times n$ matrix, namely:[7]

$$[T] := \begin{pmatrix} | & & | \\ T(v_1) & \cdots & T(v_n) \\ | & & | \end{pmatrix} = \underbrace{\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}}_{n}$$

We call this the *matrix representation* of T in the standard bases. The operation of multiplying an $m \times n$ matrix and a $n \times 1$ vector is precisely defined so that

↪ **Proposition 2.7**

$T(v) = [T] \cdot v$ for all $v \in \mathbb{F}^n$.

[7]Where $[T]$ denotes a matrix named "T".

Proof. Let $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, where $v = x_1 v_1 + \cdots + x_n v_n$. Then

$$T(v) = x_1 T(v_1) + \cdots + x_n T(v_n)$$
$$T(v_i) = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix}$$

so

$$T(v) = \begin{pmatrix} a_{11} \cdot x_1 + \cdots + a_{1n} \cdot x_n \\ \ddots \\ a_{m1} \cdot x_1 + \cdots + a_{mn} \cdot x_n \end{pmatrix} = [T] \cdot v$$

■

↪ **Definition 2.7**

For a given $m \times n$ matrix A over $\mathbb{F}$, define $L_A : \mathbb{F}^n \longrightarrow \mathbb{F}^m$ by $L_A(v) \coloneqq A \cdot v$, where v is viewed as an $n \times 1$ column. It follows from definition that the L_A is linear.

In other words, every $T \in \mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^m)$ is equal to L_A for some A.

↪ *Lecture 11; Last Updated: Sun Apr 7 23:03:11 EDT 2024*

↪ **Proposition 2.8**

The map

$$\mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^m) \longrightarrow M_{m\times n}(\mathbb{F})$$
$$T \longmapsto [T]$$

is an isomorphism of vector spaces, with inverse

$$M_{m\times n}(\mathbb{F}) \longrightarrow \mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^m)$$
$$A \longmapsto L_A.$$

Proof. Linearity: Let $\beta = \{v_1, \ldots, v_n\}$ be the standard basis for $\mathbb{F}^n$. Fix $T_1, T_2 \in \mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^m)$ and $\alpha \in \mathbb{F}$.

1.

$$
\begin{aligned}
[T_1 + T_2] &= \begin{pmatrix} & | & \\ \cdots & (T_1 + T_2)(v_i) & \cdots \\ & | & \end{pmatrix} = \begin{pmatrix} & | & \\ \cdots & T_1(v_i) + T_2(v_i) & \cdots \\ & | & \end{pmatrix} \\
&= \begin{pmatrix} & | & \\ \cdots & T_1(v_i) & \cdots \\ & | & \end{pmatrix} + \begin{pmatrix} & | & \\ \cdots & T_2(v_i) & \cdots \\ & | & \end{pmatrix} \\
&= [T_1] + [T_2]
\end{aligned}
$$

2. It remains to show that $\alpha \cdot [T] = [\alpha \cdot T]$; the proof follows similarly to 1.

Inverse: We need to show that 1. $A \mapsto L_A \mapsto [L_A]$ is the identity on $M_{m\times n}(\mathbb{F})$, and conversely, that 2. $T \mapsto [T] \mapsto L_{[T]}$ is the identity on $\mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^m)$.

1. We need to show that $[L_A] = A$. The jth column of $[L_A]$ is $L_A(v_j) = A \cdot v_j = j$th column of $A =: A^{(j)}$. Hence, the jth column of $[L_A]$ is equal to the jth column of A, and thus they are equal.
2. We showed this in proposition 2.7.

■

↪ Corollary 2.4

$\dim(\mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^m)) = \dim(M_{m\times n}(\mathbb{F})) = m \cdot n$.

Remark 2.8. *This was stated previously in proposition 2.6 by constructing an explicit basis. Indeed, this basis is precisely the image of the standard basis for $M_{m\times n}(\mathbb{F})$ under the map $A \mapsto L_A$.*

2.5 Matrix Representation of Linear Transformations, General Spaces

Remark 2.9. *The previous section was concerned with representing transformations between finite fields $\mathbb{F}^n, \mathbb{F}^m$; this section aims to make the same construction for any finite dimensional V, W.*

↪ Definition 2.8: Coordinate Vector

Let V be a finite dimensional space over $\mathbb{F}$ and let $\beta := \{v_1, \dots, v_n\}$ be a basis for V. Let $v \in V$, with (unique) representation $v = a_1 v_1 + \cdots + a_n v_n$. We denote

$$[v]_\beta := \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{F}^n$$

the *coordinate vector* of v in base β.

Remark 2.10. *Recall that $V \cong \mathbb{F}^n$ where $\dim(V) = n$, by the unique linear transformation $v_i \mapsto e_i$, where $\{e_1, \ldots, e_n\}$ the standard basis for $\mathbb{F}^n$. We denote this transformation*

$$I_\beta : V \longrightarrow \mathbb{F}.$$

For an arbitrary $v \in V$, $I_\beta(v)$ maps v to its coordinate vector:

$$\begin{aligned} I_\beta(v) = I_\beta(a_1 v_1 + \cdots + a_n v_n) &= a_1 I_\beta(v_1) + \cdots a_n I_\beta(v_n) \\ &= a_1 e_1 + \cdots + a_n e_n = [v]_\beta. \end{aligned}$$

↪ **Proposition 2.9**

The map

$$I_\beta : V \longrightarrow \mathbb{F}^n, \quad v \mapsto [v]_\beta$$

is an isomorphism.

Suppose we are given a linear transformation $T : V \longrightarrow W$, where V, W finite dimensional spaces over $\mathbb{F}$. Fix $\beta := \{v_1, \ldots, v_n\}$ and $\gamma := \{w_1, \ldots, w_m\}$ as bases for V, W resp. We can denote $[T(v_i)]_\gamma$ as $T(v_i)$ in base γ (in the field m), and construct a matrix for T:[8]

$$[T]_\beta^\gamma := \begin{pmatrix} | & & | \\ [T(v_1)]_\gamma & \cdots & [T(v_n)]_\gamma \\ | & & | \end{pmatrix}$$

We call this the *matrix representation* of T from β to γ.

↪ **Theorem 2.7**

Let $T : V \longrightarrow W, \beta, \gamma$ as above.

1. The following diagram commutes:

$$\begin{tikzcd} \bullet V \arrow[r, "T"] \arrow[d, "I_\beta"'] & \bullet W \arrow[d, "I_\gamma"] \\ \bullet \mathbb{F}^n \arrow[r, dashed, "L_{[T]_\beta^\gamma}"'] & \bullet \mathbb{F}^m \end{tikzcd}$$

   Namely, $I_\gamma \circ T = L_{[T]_\beta^\gamma} \circ I_\beta$, or equivalently, given $v \in V$, $[T(v)]_\gamma = [T]_\beta^\gamma \cdot [v]_\beta$.

2. The map $\mathrm{Hom}(V, W) \longrightarrow M_{m \times n}(\mathbb{F}), T \mapsto [T]_\beta^\gamma$ is a vector space isomorphism with inverse begin the map $M_{m \times n}(\mathbb{F}) \longrightarrow \mathrm{Hom}(V, W), A \mapsto I_\gamma^{-1} \circ L_A \circ I_\beta$

[8]Where we denote $[T]_\beta^\gamma$ as the matrix representation of the transform $T : V \longrightarrow W$, with basis β, γ for V, W respectively.

Proof. 2. is left as a (homework) exercise; it follows directly from 1.

Fix $v \in V$. We need to show that $I_\gamma \circ T(v) = L_{[T]_\beta^\gamma} \circ I_\beta(v)$. We have

$$I_\gamma \circ T(v) = [T(v)]_\gamma.$$

OTOH,

$$L_{[T]_\beta^\gamma} \circ I_\beta(v) = L_{[T]_\beta^\gamma}([v]_\beta) = [T]_\beta^\gamma \cdot [v]_\beta.$$

We need to show, then, that $[T(v)]_\gamma = [T]_\beta^\gamma \cdot [v]_\beta$. Let $v = a_1 v_1 + \cdots + a_n v_n$, so $[v]_\beta = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$. Recall that $[T]_\beta^\gamma = \begin{pmatrix} | & & | \\ [T(v_1)]_\gamma & \cdots & [T(v_n)]_\gamma \\ | & & | \end{pmatrix}$. Thus, we have

$$\begin{aligned} [T]_\beta^\gamma \cdot [v]_\beta = a_1[T(v_1)]_\gamma + \cdots + a_n[T(v_n)]_\gamma &= [a_1 T(v_1) + \cdots + a_n T(v_n)]_\gamma \quad \textit{(by linearly of } I_\gamma\textit{)} \\ &= [T(a_1 v_1 + \cdots + a_n v_n)]_\gamma \quad \textit{(by linearity of } T\textit{)} \\ &= [T(v)]_\gamma, \end{aligned}$$

which is precisely what we wanted to show. ■

Remark 2.11. *For $A \in M_{m\times n}(\mathbb{F})$ and $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n$, we have*

$$A \cdot x = x_1 \cdot A^{(1)} + x_2 \cdot A^{(2)} + \cdots + x_n \cdot A^{(n)},$$

where $A^{(j)}$ is the jth column of A; thus $A \cdot x$ is a linear combination of A, with coefficients given by the vector x; this interpretation can make it easier to make sense of computations.

↪ Lecture 12; Last Updated: Sat Apr 6 10:19:07 EDT 2024

2.6 Composition of Linear Transformations, Matrix Multiplication

↪ Proposition 2.10

Composition is associative; given $T : V \to W$, $S : W \to U$, and $R : U \to X$, then

$$(R \circ S) \circ T = R \circ (S \circ T).$$

Proof. Fix $v \in V$. Then

$$(R \circ S) \circ T(v) = (R \circ S)(T(v)) = R(S(T(v)))$$

OTOH:

$$R \circ (S \circ T)(v) = R((S \circ T)(v)) = R(S(T(v))).$$

■

Let $A \in M_{m \times n}(\mathbb{F})$ and $B \in M_{l \times m}(\mathbb{F})$. Then, $L_A : \mathbb{F}^n \to \mathbb{F}^m$ and $L_B : \mathbb{F}^m \to \mathbb{F}^l$, and have composition $L_B \circ L_A : \mathbb{F}^n \to \mathbb{F}^l$. We know that $L_B \circ L_A$ is a linear transformation, and thus must be equal to L_C for some matrix $C \in M_{l \times n}(\mathbb{F})$. Indeed, C is the matrix representation of the transformation $[L_B \circ L_A]$, as proven previously.

Let $\beta = \{e_1, \dots, e_n\}$ for $\mathbb{F}^n$, then

$$[L_B \circ L_A] = \begin{pmatrix} | & & | \\ L_B \circ L_A(e_1) & \cdots & L_B \circ L_A(e_n) \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ B \cdot (A \cdot e_1) & \cdots & B \cdot (A \cdot e_n) \\ | & & | \end{pmatrix}$$

↪ **Definition 2.9: Matrix Multiplication**

For matrices $A \in M_{m \times n}(\mathbb{F})$ and $B \in M_{l \times m}(\mathbb{F})$, define their product $B \cdot A$ to be the matrix

$$[L_B \circ L_A] = \begin{pmatrix} | & & | \\ B \cdot (A \cdot e_1) & \cdots & B \cdot (A \cdot e_n) \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ B \cdot A^{(1)} & \cdots & B \cdot A^{(2)} \\ | & & | \end{pmatrix} = (c_{ij})_{1 \leqslant i \leqslant l}^{1 \leqslant j \leqslant n}$$

where $A^{(j)}$ is the jth column of A, $c_{ij} := \begin{pmatrix} - & B_{(i)} & - \end{pmatrix} \cdot \begin{pmatrix} | \\ A^{(j)} \\ | \end{pmatrix}$.

↪ **Proposition 2.11**

$[L_B \circ L_A] = B \cdot A$, ie $L_B \circ L_A = L_{B \cdot A}$.

Proof. Follows from our definition. ■

↪ **Corollary 2.5**

Matrix multiplication is association; $C \cdot (B \cdot A) = (C \cdot B) \cdot A$ for $A \in M_{m \times n}(\mathbb{F}), B \in M_{l \times m}(\mathbb{F}), C \in M_{k \times l}(\mathbb{F})$.

Proof. $C \cdot (B \cdot A) = [L_C \circ (L_B \circ L_A)] = [(L_C \circ L_B) \circ L_A] = (C \cdot B) \cdot A$. ■

Remark 2.12. *This is proven by the linear transformation representation of matrices; try proving this* directly *from our definition.*

↪ **Corollary 2.6**

Let V, W, U be finite-dimensional vector spaces over $\mathbb{F}$, $T : V \to W$, $S : W \to U$ be linear transformations and α, β, γ be bases for V, W, U resp. Then,

$$[S \circ T]^{\gamma}_{\alpha} = [S]^{\gamma}_{\beta} \cdot [T]^{\beta}_{\alpha}.$$

Proof. Follows from the commutativity of the diagrams:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\wr"'] & W \arrow[r, "S"] \arrow[d, "\wr"'] & U \arrow[d, "\wr"'] \\ \mathbb{F}^n \arrow[r, "{[T]^{\beta}_{\alpha}}"'] & \mathbb{F}^m \arrow[r, "{[S]^{\gamma}_{\beta}}"'] & \mathbb{F}^l \end{tikzcd} \iff \begin{tikzcd} V \arrow[r, "T \circ S"] \arrow[d, "\wr"'] & U \arrow[d, "\wr"'] \\ \mathbb{F}^n \arrow[r, "{[S \circ T]^{\gamma}_{\alpha}}"'] & \mathbb{F}^l \end{tikzcd}$$

In "words", for $v \in V$,

$$[S \circ T]^{\gamma}_{\alpha} \cdot [v]_{\alpha} = [(S \circ T)(v)]^{\gamma}_{\alpha} = [S(T(v))]_{\alpha} = [S]^{\gamma}_{\beta} \cdot [T(v)]_{\beta} = [S]^{\gamma}_{\beta} \cdot [T]^{\beta}_{\alpha} \cdot [v]_{\alpha},$$

ie we have shown that $L_{[S \circ T]^{\gamma}_{\alpha}} = L_{[S]^{\gamma}_{\beta} \cdot [T]^{\beta}_{\alpha}}$. Because $A \mapsto L_A$ is an isomorphism, it follows that $[S \circ T]^{\gamma}_{\alpha} = [S]^{\gamma}_{\beta} \cdot [T]^{\beta}_{\alpha}$. ■

↪ Lecture 13; Last Updated: Mon Mar 25 13:48:03 EDT 2024

2.7 Inverses of Transformations and Matrices

Remark 2.13. *Recall that, given a function $f : X \to Y$, a function $g : Y \to X$ is called*

1. *a* left inverse *of f if $g \circ f = \mathrm{Id}_X$;*
2. *a* right inverse *of f if $f \circ g = \mathrm{Id}_X$;*
3. *a (two-sided)* inverse *of f if g both a left and right inverse of f.*

If an inverse exists, it is unique; let g_0, g_1 be inverse of f, then, $g_0 = g_0 \circ (f \circ g_1) = (g_0 \circ f) \circ g_1 = g_1$.

↪ **Proposition 2.12**

Let $f : X \to Y$. Then,

1. f has a left-inverse $\iff$ f injective;
2. f has a right-inverse $\iff$ f surjective;
3. f has an inverse $\iff$ f bijective.

Proof. ((a), $\implies$) Suppose $g : Y \to X$ is a left-inverse of f and $f(x_1) = f(x_2)$. Then, $g \circ f(x_1) = g \circ f(x_2) \implies x_1 = x_2$ and so f injective.

((b), $\implies$) Suppose $g : Y \to X$ is a right-inverse of f and let $y \in Y$. Then, $f(g(y)) = y \implies y \in f(X)$.

The remainder of the cases and directions are left as an exercise. ∎

Remark 2.14. *Proof of* (b), $\impliedby$ *uses Axiom of Choice.*

⊛ **Example 2.2**

1. The differentiation transform $\delta : \mathbb{F}[t]_{n+1} \to \mathbb{F}[t]_n, p(t) \mapsto p'(t)$ has a right inverse, the integration transform, $\iota : \mathbb{F}[t]_n \to \mathbb{F}[t]_{n+1}, p(t) \mapsto$ antiderivative of $p(t)$; conversely, ι has left inverse δ; they do not admit inverses.

2. Let $f : \mathbb{F}[\![t]\!] \to \mathbb{F}[\![t]\!]$ be the left-shift map, where $\sum_{n=0}^{\infty} a_n t^n \mapsto \sum_{n=1}^{\infty} a_n t^{n-1}$. Then, $g : \mathbb{F}[\![t]\!] \to \mathbb{F}[\![t]\!]$ with $\sum_{n=0}^{\infty} a_n t^n \mapsto \sum_{n=0} a_n t^{n+1}$, the right-shift map, is a right inverse of f, but f has no left inverse (it is not injective).

Remark 2.15. *The existence of only one-sided inverses existing happens only when in infinite-dimensional vectors spaces, or when the dimension of the domain is not the same as the dimension of the codomain.*

↪ **Corollary 2.7: Of Rank-Nullity Theorem**

Let $T : V \to W$ s.t. $\dim(V) = \dim(W) < \infty$. TFAE:

1. T has a left-inverse;
2. T has a right-inverse;
3. T is invertible (has an inverse).

Proof. We have already that T injective $\iff$ T surjective $\iff$ T bijective. ∎

↪ **Definition 2.10: Matrix Inverse**

We call a $n \times n$ matrix B over $\mathbb{F}$ the *inverse* of an $n \times n$ matrix A over $\mathbb{F}$ if $A \cdot B = B \cdot A = I_n$. We denote $B = A^{-1}$.

↪ **Proposition 2.13**

Let $A \in M_n(\mathbb{F})$. Then,

1. L_A is invertible $\iff$ A is invertible, in which case $L_A^{-1} = L_{A^{-1}}$;
2. A is invertible $\iff$ it has a left-inverse, ie $B \cdot A = I_n$ $\iff$ it has a right-inverse, ie $A \cdot B = I_n$.

Proof. 1. L_A invertible $\iff$ $\exists T : \mathbb{F}^n \to \mathbb{F}^n$-linear s.t. $L_A \circ T = T \circ L_A = I_{\mathbb{F}^n}$ $\iff$ $\exists$ a matrix $B \in M_n(\mathbb{F})$ such that $L_A \circ L_B = L_B \circ L_A = I_{\mathbb{F}^n}$ $\iff$ there is a matrix $B \in M_n(\mathbb{F})$ s.t. $L_{AB} = L_{BA} = I_{\mathbb{F}^n}$ $\iff$ there is a $B \in M_n(\mathbb{F})$ s.t. $A \cdot B = B \cdot A = I_n$.

2. Follows directly from corollary 2.7 and part 1.

■

2.8 Invariant Subspaces and Nilpotent Transformations

↪ Definition 2.11: T-Invariant

Let $T : V \to V$ be a linear transformation.[9] We call a subspace $W \subseteq V$ *T-invariant* if $T(W) \subseteq W$.

⊛ Example 2.3: Examples of Invariant Subspaces

1. For any $T : V \to V$, $\mathrm{Im}(T)$ is T-invariant.
2. For any $T : V \to V$, $\mathrm{Ker}(T)$ is T-invariant, since $T(v) = 0_V \in \mathrm{Ker}(T)\,\forall\, v \in \mathrm{Ker}(T)$. Moreover, for any $n \in \mathbb{N}$, the space $\mathrm{Ker}(T^n)$ is T-invariant.[10]

↪ Lecture 14; Last Updated: Mon Mar 25 13:48:03 EDT 2024

↪ Proposition 2.14

For a linear operator $T : V \to V$, the following hold:

1. $V \supseteq \mathrm{Im}(T) \supseteq \mathrm{Im}(T^2) \supseteq \cdots \supseteq \mathrm{Im}(T^n) \supseteq \cdots$. Moreover, $\mathrm{Im}(T^n)$ is T-invariant for any $n \in \mathbb{N}$.
2. $\{0_V\} \subseteq \mathrm{Ker}(T) \subseteq \mathrm{Ker}(T^2) \subseteq \cdots \subseteq \mathrm{Ker}(T^n) \subseteq \cdots$. Moreover, $\mathrm{Ker}(T^n)$ is T-invariant for any $n \in \mathbb{N}$.

Proof. 1. If $x \in \mathrm{Im}(T^{n+1})$, then $x = T^{n+1}(y) = T^n(T(y)) \in \mathrm{Im}(T^n)$ for some $y \in V$, hence $\mathrm{Im}(T^{n+1}) \subseteq \mathrm{Im}(T^n)$. If $x \in \mathrm{Im}(T^n)$, then $x = T^n(y)$ so $T(x) = T(T^n(y)) = T^n(T(y)) \in \mathrm{Im}(T^n)$, so $T(\mathrm{Im}(T^n)) \subseteq \mathrm{Im}(T^n)$.

2. If $x \in \mathrm{Ker}(T^n)$, then $T^{n+1}(x) = T(T^n(x)) = T(0_V) = 0_V$ hence $x \in \mathrm{Ker}(T^{n+1})$ so $\mathrm{Ker}(T^n) \subseteq \mathrm{Ker}(T^{n+1})$. Moreover, $T(x) \in \mathrm{Ker}(T^n)$ since $T(x) \in \mathrm{Ker}(T^{n-1}) \subseteq \mathrm{Ker}(T^n)$, since $T^{n-1}(T(x)) = T^n(x) = 0_V$ so $T(\mathrm{Ker}(T^n)) \subseteq \mathrm{Ker}(T^n)$.

■

⊛ Example 2.4: More Examples of Invariant Subspaces

Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(x, y, z) := (2x + y, 3x - y, 7z)$. Then, the $x - y$ plane, $\{(x, y, z) \in \mathbb{R}^3 : z = 0\}$

[9] Because the domain and codomain are the same, we often call T a “linear operator”.
[10] $T^n := T \circ T \circ \cdots \circ T$, n times; $T^0 := I_V$.

is T-invariant, as is the z axis, $\{(x, y, z) \in \mathbb{R}^3 : x = y = 0\}$. Hence, we can decompose $\mathbb{R}^3$ into two T-invariant subspaces, namely $x - y$ plane $\oplus$ z-axis.

↪ Definition 2.12: Nilpotent

In a ring R, an element $r \in R$ is called *nilpotent* if $r^n = 0$ for some $n \in \mathbb{N}^+$.

A linear transformation $T : V \to V$ is called nilpotent if $T^n = 0$ for some $n \in \mathbb{N}^+$.[11]

For a matrix $A \in M_n(\mathbb{F})$, A is called nilpotent if $A^n = 0_n$ for some $n \in \mathbb{N}^+$.

⊛ Example 2.5: Examples of Nilpotent Transformations

1. Let V, n-dimensional vector space over $\mathbb{F}$ with basis $\beta := \{v_1, \ldots, v_n\}$. Let $T : V \to V$ be the unique linear transformation that "shifts" β: ie, $T(v_1) := 0_V, T(v_2) := v_1, \ldots, T(v_n) = v_{n-1}$.
2. The differentiation operation, $\delta : \mathbb{F}[t]_n \to \mathbb{F}[t]_n$ is nilpotent, since $\delta^{n+1} = 0$ for any polynomial.
3. For any matrix $A \in M_n(\mathbb{F})$, A is nilpotent iff $L_A : \mathbb{F}^n \to \mathbb{F}^n$ is nilpotent.

   Proof. $L_{A^k} = L_A^k \implies A^k = 0 \iff L_{A^k} = 0 \iff L_A^k = 0$ ■

4. $n \times n$ matrices that are strictly upper triangular[12] are nilpotent. For instance, for 3×3, we need to show[13]

$$\begin{pmatrix} 0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0 \end{pmatrix}^3 = 0 \iff \begin{pmatrix} 0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0 \end{pmatrix}^3 \cdot \begin{pmatrix} \star \\ \star \\ \star \end{pmatrix} = 0$$

[11] One can verify that all linear transformations $T : V \to V$ from a vector space to itself form a ring with $(\circ, +)$, ie composition and ("standard") addition of transformations. The same holds for linear operators defined over an abelian group (where the same + operation is endowed by the ring).

We have:

$$\begin{pmatrix}0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0\end{pmatrix}^2 \begin{pmatrix}0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\star \\ \star \\ \star\end{pmatrix} = \begin{pmatrix}0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0\end{pmatrix}^2 \begin{pmatrix}\star \\ \star \\ 0\end{pmatrix}$$
$$= \begin{pmatrix}0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\star \\ \star \\ 0\end{pmatrix}$$
$$= \begin{pmatrix}0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\star \\ 0 \\ 0\end{pmatrix}$$
$$= \begin{pmatrix}0 \\ 0 \\ 0\end{pmatrix}.$$

↪ **Proposition 2.15**

If V is n-dimensional and $T : V \to V$ is a linear nilpotent transformation, then $T^n = 0$.

Proof. Left as a (homework) exercise. ∎

↪ **Definition 2.13: Domain Restriction**

For a function $f : X \to Y$ and $A \subseteq X$, we define the *restriction* of f to A as the function $f|_A : A \to Y$ given by $a \mapsto f(a)$.

↪ **Definition 2.14: Direct Sum**

Let V be a vector space over $\mathbb{F}$, and let $W_0, W_1 \subseteq V$ be subspaces of V. If

1. $W_0 \cap W_1 = \{0_V\}$ (the subspaces are *linearly independent*), and
2. $W_0 + W_1 = \{w_0 + w_1 : w_0 \in W_0, w_1 \in W_1\} = V$,

we write $V = W_0 \oplus W_1$, and say V is the *direct sum* if W_0, W_1.

[13]ie zeros everywhere except cells strictly above diagonal.
[13]Where we denote arbitrary elements $\star$; different $\star$s are not necessarily equal.

↪ **Theorem 2.8: Fitting's Lemma**

For finite dimensional vector space V over $\mathbb{F}$ and a linear transformation $T : V \to V$, there is a decomposition

$$V = U \oplus W$$

as a direct sum of T-invariant subspaces U, W such that $T|_U : U \to U$ is nilpotent and $T|_W : W \to W$ is an isomorphism.

Proof. Recall that $\mathrm{Im}(T) \supseteq \cdots \supseteq \mathrm{Im}(T^n)$ and $\mathrm{Ker}(T) \subseteq \cdots \subseteq \mathrm{Ker}(T^n)$. Both of these become constant eventually, ie the inequalities become strict equalities, hence $\exists N \in \mathbb{N}^+$ such that $\forall k \in \mathbb{N}$, $\mathrm{Im}(T^{N+k}) = \mathrm{Im}(T^N)$ and $\mathrm{Ker}(T^{N+k}) = \mathrm{Ker}(T^N)$.

Let $U := \mathrm{Ker}(T^N)$ and $W := \mathrm{Im}(T^N)$. These are clearly T-invariant.

$T^N(\mathrm{Ker}(T^N)) = \{0_V\}$, and $T(\mathrm{Im}(T^N)) = \mathrm{Im}(T^{N+1}) = \mathrm{Im}(T^N) = W$ and thus $T|_W : W \to W$ is surjective and hence $T|_W$ must be injective and thus an isomorphism.

It remains to show that $V = U \oplus W$. If $v \in U \cap W$, $T^N(v) = 0_V$ but $T|_W$ an isomorphism so $T^N(v) = 0 \iff v = 0_V$, hence $U \cap W = \{0_V\}$.

Thus, we have $\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W) = \dim(U) + \dim(W) = \dim(V)$; moreover, it must be that $U + W = V$.[14] ■

↪ *Lecture 15; Last Updated: Mon Mar 25 13:48:03 EDT 2024*

2.9 Dual Spaces

↪ **Definition 2.15: Dual Space**

For a vector space V over a field $\mathbb{F}$, linear transformations from $V \to \mathbb{F}$ (where we view $\mathbb{F}$ as a one-dimensional vector space over $\mathbb{F}$) are called *linear functionals*. The space of linear functionals (namely, $\mathrm{Hom}(V, \mathbb{F})$) is denoted V^*, and called the *dual space* of V.

↪ **Proposition 2.16**

If V is finite dimensional, $\dim(V^*) = \dim(V)$.[15]

Proof. For finite dimensional V, we know that $\dim(\mathrm{Hom}(V, \mathbb{F})) = \dim(V) \cdot \dim(\mathbb{F}) = \dim(V)$, hence $\dim(V^*) = \dim(V)$. In the same notation with which we proved this originally in proposition 2.6; fix a basis $\beta := \{v_1, \ldots, v_n\}$ for V and the standard basis $\gamma := \{1\}$ for $\mathbb{F}$, and defined $\beta^* := \{f_1, \ldots, f_n\}$, where $f_i := T_{v_i,1} : V \to \mathbb{F}$ maps $v_i \mapsto 1$ and every other basis vector to $0_{\mathbb{F}}$. ■

Remark 2.16. *The basis β^* for V^* is called the* dual basis. *Explicitly, we have:*

[14] It is precisely here that we use finiteness of V.
[15] This does *not* hold for infinite dimensional spaces.

↪ **Corollary 2.8**

Let V be a finite dimensional vector space over $\mathbb{F}$ and let $\beta := \{v_1, \dots, v_n\}$ be a basis for V. Then,

$$\beta^* := \{f_1, \dots, f_n\}$$

is a basis for V^*. Moreover, for each linear functional $f \in V^*$,

$$f = \sum_{i=1}^{n} f(v_i) \cdot f_i.$$

Proof. Linear indepedence: let $a_1 f_1 + \cdots + a_n f_n = 0_{V^*} =: 0$. Then,

$$(a_1 f_1 + \cdots + a_n f_n)(v_i) = a_i f_i(v_i) = a_i \cdot 1 = a_i \implies a_i = 0,$$

hence β^* indeed linearly independent.

Spanning: let $f \in V^*$. We claim that $f = \sum_{i=1}^{n} f(v_i) f_i$. It suffices to show these two sides are equal on the basis vectors, as linear transformations are determined by their effect on basis vectors. We have:

$$\left(\sum_{i=1}^{n} f(v_i) f_i\right)(v_j) = \sum_{i=1}^{n} f(v_i) f_i(v_j) = \sum_{i=1}^{n} f(v_i) \cdot \delta_{ij} = f(v_j),$$

as desired.[16] ■

⊛ **Example 2.6**

1. Let $V := \mathbb{F}^n$ and $\beta := \{v_1, \dots, v_n\}$ be a basis for $\mathbb{F}^n$, viewed as column vectors, and let $\beta^* := \{f_1, \dots, f_n\}$ be the dual basis for V^*. Recall that $f_i : \mathbb{F}^n \to \mathbb{F}$, hence $f_i := L_{A_i}$ for some matrix $A_i \in M_{1 \times n}(\mathbb{F}) :=$ space of $1 \times n$ row vectors. Hence, $A_i = e_i^t$.
2. Consider V^{**}, the dual of the dual. If V is finite-dimensional, then as $\dim(V) = \dim(V^*)$, we have $\dim(V) = \dim(V^*) = \dim(V^{**})$, ie, they are (abstractly) isomorphic.

   We have that $T : V \to V^*, v_i \mapsto f_i$ is an isomorphism; we define an explicit isomorphism to V^{**} below.

↪ **Definition 2.16**

Let V be an arbitrary vector space over $\mathbb{F}$. For each $x \in V$, define $\hat{x} \in V^{**}$ by $\hat{x} : V^* \to \mathbb{F}$, where $\hat{x}(f) := f(x)$.

Remark 2.17. *Note that $\hat{x}$ is linear.*

[16] Where $\delta_{ij} := \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$ is the Kronecker delta.

↪ **Theorem 2.9**

The map $x \mapsto \hat{x} : V \to V^{**}$ is a linear injection. In particular, if V is finite dimensional, it is an isomorphism.

Proof. Let $x \in V$ and suppose $\hat{x} = 0_{V^{**}}$. Let β be a basis for V and β^* its dual basis. Let $x = a_1 v_1 + \cdots + a_n v_n$ for $v_i \in \beta, a_i \in \mathbb{F}$. Let f_i such that $f_i(v_j) = \delta_{ij} v_j$. Then,

$$\hat{x} f_i = f_i(x) = f_i(a_1 v_1 + \cdots a_n v_n) = a_i = 0,$$

hence, $a_i = 0 \, \forall \, i$. Hence, $x = 0$, and thus $\hat{x}$ has a trivial kernel and is thus injective. ■

↪ *Lecture 16; Last Updated: Sat Apr 6 10:19:07 EDT 2024*

Remark 2.18. *Notice that to get an isomorphism $V \cong V^*$, we fixed a basis for V to define it. However, for $V \cong V^{**}$, we had a canonical isomorphism independent of choice of basis. Writing $S \subseteq V$, $\hat{S} := \{\hat{x} : x \in S\} \subseteq V^{**}$, our theorem says that $\hat{V} = V^{**}$ for finite-dimensional V.*

↪ **Definition 2.17: Annihilator**

Let V be a vector space over $\mathbb{F}$ and $S \subseteq V$. We call

$$S^\perp := \{f \in V^* : f|_S = 0\} = \{f \in V^* : f(u) = 0 \, \forall \, u \in S\}$$

the *annihilator* of S.

↪ **Proposition 2.17**

Let V be a vector space over $\mathbb{F}$ and $S \subseteq V$.

1. $S^\perp$ is a subspace of V^*[17]
2. $S_1 \subseteq S_2 \subseteq V \implies S_1^\perp \supseteq S_2^\perp$
3. $S^\perp = (\mathrm{Span}(S))^\perp$

Proof.

1. If $f_1, f_2 \in S^\perp, a \in \mathbb{F}$, then $\forall \, u \in S$,

$$(a f_1 + f_2)(u) = a f_1(u) + f_2(u) = a \cdot 0 + 0,$$

so $a f_1 + f_2 \in S^\perp$.

2. Clear.
3. If $f \in V^*$ takes all vectors in S to 0, then it does the same for linear combinations.

■

---

[17]Even if S is not a subspace itself.

↪ **Proposition 2.18**

If V is finite dimensional and $U \subseteq V$ a subspace, then $(U^{\perp})^{\perp} = \hat{U}$.

Proof. We know that $V^{**} = \hat{V}$, so we fix $\hat{x} \in \hat{V}$ and show that

$$\hat{x} \in (U^{\perp})^{\perp} \iff \hat{x} \in \hat{U} \iff x \in U.$$

We have

$$\hat{x} \in (U^{\perp})^{\perp} :\iff \forall f \in U^{\perp}, \hat{x}(f) = f(x) = 0$$

hence if $x \in U$, then $\hat{x} \in (U^{\perp})^{\perp}$, so $\hat{U} \subseteq (U^{\perp})^{\perp}$.

Conversely, let $\hat{x} \in (U^{\perp})^{\perp}$. Then, $\forall f \in U^{\perp}$, $f(x) = 0$.

Suppose towards a contradiction that $x \notin U$. We aim to define $f \in U^{\perp}$ s.t. $f(x) = 1$, obtaining a contradiction. Let $\{u_1, \dots, u_k\}$ be a basis for U, noting that $\{u_1, \dots, u_k, x\}$ still linearly independent by assumption of $x \notin U = \text{Span}(\{u_1, \dots, u_k\})$. Thus, we can extend this to a basis $\beta = \{u_1, \dots, u_k, x, v_1, \dots, v_m\}$ for V. Define $f : V \to \mathbb{F} \in V^*$ as the unique linear transformation such that $f(u_i) = f(v_j) = 0$ and $f(x) = 1$. Then, $f \in U^{\perp}$ by definition, and $f(x) = 1$ by definition. This is a contradiction that $x \notin U$. ■

↪ **Corollary 2.9**

For a finite dimensional V and subspace $U \subseteq V$,

$$U = \{x \in V : \forall f \in U^{\perp}, f(x) = 0\}.$$

↪ **Definition 2.18: Dual/Transpose of T**

Let V, W be vector spaces over a field $\mathbb{F}$ and $T : V \to W$. We define the *dual/transpose* of T as the map $T^t : W^* \to V^*$, given by $g \mapsto g \circ T$. Ie, $T^t(g)(v) := g \circ T(v) = g(T(v))$.

↪ **Proposition 2.19**

Let V, W be vector spaces over a field $\mathbb{F}$ and $T : V \to W$.

1. T^t is linear.
2. $\text{Ker}(T^t) = (\text{Im}(T))^{\perp}$.
3. $\text{Im}(T^t) \subseteq (\text{Ker}(T))^{\perp}$ and is equal if V, W are finite dimensional.
4. If V, W are finite dimensional and β, γ are bases resp., then

$$[T^t]_{\gamma^*}^{\beta^*} = ([T]_{\beta}^{\gamma})^t.$$

Proof. 1. $T^t(ag_1+g_2) = (ag_1+g_2)\circ T = a\cdot g_1\circ T + g_2\circ T = a\cdot T^t(g_1)+T^*(g_2)$, $\forall\, g_1, g_2 \in W^*, a\in\mathbb{F}$.

2. For $g\in W^*$,

$$\begin{aligned} g\in \mathrm{Ker}(T^t) :\iff T^t(g) = 0_{V^*} &\iff T^t(g)(v) = 0\,\forall\, v\in V \\ &\iff g(T(v)) = 0\,\forall\, v\in V \\ &\iff g(w) = 0\,\forall\, w\in \mathrm{Im}(T) \\ &\iff g\in (\mathrm{Im}(T))^\perp \end{aligned}$$

3. Fix $f = T^t(g)\in \mathrm{Im}(T^t)$, $g\in W^*$, and $u\in\mathrm{Ker}(T)$, noting that $f(u) = T^t(g)(u) = g(T(u)) = g(0_W) = 0$ so $f\in(\mathrm{Ker}(T))^\perp$.

   Suppose now V, W are finite dimensional; we've shown an inclusion, so it suffices now to show that $\dim(\mathrm{Im}(T^t)) = \dim(\mathrm{Ker}(T))^\perp$. We have:

$$\begin{aligned} \dim(\mathrm{Im}(T^t)) &= \dim(W^*) - \dim(\mathrm{Ker}(T^t)) \\ &= \dim(W) - \dim(\mathrm{Im}(T)^\perp) \\ &= \dim(W) - \dim(W) + \dim(\mathrm{Im}(T)) \\ &= \dim(\mathrm{Im}(T)) \end{aligned}$$

   OTOH:

$$\dim(\mathrm{Ker}(T)^\perp) = \dim(V) - \dim(\mathrm{Ker}(T)) = \dim(\mathrm{Im}(T)),$$

   and thus $\dim(\mathrm{Im}(T^t)) = \dim(\mathrm{Ker}(T))^\perp$ (remarking that the first equality follows from 1. of the following theorem, and 2. from the dimension theorem).

4. Let $\beta := \{v_1,\dots,v_n\}, \gamma := \{w_1,\dots,w_m\}$ be finite bases for V, W resp. Recall that

$$A := [T]_\beta^\gamma := \begin{pmatrix} | & & | \\ [T(v_1)]_\gamma & \cdots & [T(v_n)]_\gamma \\ | & & | \end{pmatrix},$$

   ie $A^{(j)} = [T(v_j)]_\gamma$ hence $T(v_j) = \sum_{k=1}^m A_{kj}w_k$.

   Similarly, write $\gamma^* := \{g_1,\dots,g_m\}$ and $\beta^* := \{f_1,\dots,f_n\}$, then

$$B := [T^t]_{\gamma^*}^{\beta^*} := \begin{pmatrix} | & & | \\ [T^t(g_1)]_{\beta^*} & \cdots & [T^t(g_m)]_{\beta^*} \\ | & & | \end{pmatrix},$$

   so $T^t(g_i) = \sum_{\ell=1}^n B_{\ell i}f_\ell = \sum_{\ell=1}^n T^t(g_i)(v_\ell)f_\ell$, so $B_{\ell i} = T^t(g_i)(v_\ell)$. To complete the proof, we must show that

$A_{ij} = B_{ji}$ for all i, j:

$$B_{ji} = T^t(g_i)(v_j) = g_i(T(v_j)) = g_i(\sum_{k=1}^m A_{kj} w_k) = \sum_{k=1}^m A_{kj} g_i(w_k) = A_{ij},$$

where the last equality $g_i(w_k) = \delta_{ik}$, by construction.

■

↪ Lecture 17; Last Updated: Mon Mar 25 13:48:03 EDT 2024

↪ Theorem 2.10

Let V be a finite-dimensional vector space over $\mathbb{F}$ and $U \subseteq V$ be a subspace.

1. $\dim(U^\perp) = \dim(V) - \dim(U)$. In fact, if $\{v_1, \ldots, v_k\}$ is a basis for U and $\beta := \{v_1, \ldots, v_k, v_{k+1} \ldots, v_n\}$ is a basis for V with the dual basis $\beta^* = \{f_1, \ldots, f_n\}$, then $\{f_{k+1}, \ldots, f_n\}$ is a basis for $U^\perp$.

2. $(V/U)^* \cong U^\perp$ by the map $f \mapsto f_U$, where $f_U : V \to \mathbb{F}$ given by $f_U(v) := f(v + U)$.

Proof. Left as a (homework) exercise.

■

↪ Corollary 2.10: of proposition 2.19

Let V, W be vector spaces over $\mathbb{F}$ and $T : V \to W$ be a linear transformation.

1. T^t injective $\iff$ T surjective.

2. If V, W finite dimensional, then T^t surjective $\iff$ T injective.

Proof.

1. T^t injective $\iff$ $\mathrm{Ker}(T^t) = \{0_{W^*}\}$ $\iff$ $\mathrm{Im}(T)^\perp = \{0_{W^*}\}$ $\implies$ $^{\circledast}\,\mathrm{Im}(T) = W$ $\iff$ T surjective. Conversely, if $\mathrm{Im}(T) = W \implies (\mathrm{Im}(T))^t = \{0_{W^*}\}$ (and the rest follows identically).

2. $\mathrm{Im}(T^t) = \mathrm{Ker}(T)^\perp \implies \mathrm{Im}(T^\perp) = V^* \iff \mathrm{Ker}(T) = \{0_V\}$, following similar logic to above.

■

Remark 2.19. *Part 4. of proposition 2.19 establishes a dependency between the columns and rows of a matrix; precisely:*

↪ Lecture 18; Last Updated: Mon Mar 25 13:48:03 EDT 2024

2.9.1 Application to Matrix Rank

$\hookrightarrow$ **Definition 2.19: Matrix Rank/C-Rank,R-Rank**

For a matrix $A \in M_{m\times n}(\mathbb{F})$, we define

$$\operatorname{rank}(A) := \operatorname{rank}(L_A)$$

and the *column rank* of

$$\text{c-rank}(A) := \text{size of maximal indep. subset of columns } \{A^{(1)}, \dots, A^{(n)}\}$$

and *row rank* of

$$\text{r-rank}(A) := \text{size of maximal indep. subset of rows } \{A_{(1)}, \dots, A_{(m)}\}.$$

Remark 2.20. *Notice that* $\operatorname{rank}(A) = \text{c-rank}(A)$.

$\hookrightarrow$ **Corollary 2.11**

$$\operatorname{rank}(A) = \operatorname{rank}(A^t) = \text{r-rank}(A)$$

Proof. We know already that $\operatorname{rank}(A^t) = \text{c-rank}(A^t) = \text{r-rank}(A)$, as remarked previously, hence we need only to show that $\operatorname{rank}(A^t) = \operatorname{rank}(A)$. But $A = [L_A]$ and $A^t = [L_{A^t}] = [L_A]^t = [L_A^t]$. Thus, $\operatorname{rank}(A) = \operatorname{rank}(L_A) = \operatorname{rank}(L_A^t) = \operatorname{rank}(A^t)$. ■

$\hookrightarrow$ **Corollary 2.12**

$$\operatorname{rank}(A) = \text{c-rank}(A) = \text{r-rank}(A), \quad \forall A \in M_{m\times n}(\mathbb{F})$$

3 Elementary Matrices, Matrix Operations

3.1 Systems of Linear Equations

We can write a system of m equations of n unknowns x_i

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \quad \ddots \qquad \ddots \qquad \ddots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{cases}$$

succinctly as a matrix equation

$$A \cdot \vec{x} = \vec{b},$$

where $A := (a_{ij}) \in M_{m\times n}(\mathbb{F})$, $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, and $\vec{b} := \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{F}^m$. Hence, $\vec{x}$ solves $A\vec{x} = \vec{b} \iff L_A(\vec{x}) = \vec{b} \iff \vec{x} \in L_A^{-1}(\vec{b})$. In other words, a solution exists iff $\vec{b} \in \mathrm{Im}(L_A) = \mathrm{Span}(A^{(1)}, \dots, A^{(n)})$. In particular, when $\vec{b} = \vec{0}$, a solution always exists, $\vec{x} = \vec{0}$. We call $A \cdot \vec{x} = \vec{0}$ the *homogeneous system of equations* of A.

It follows that $A \cdot \vec{x} = \vec{0}$ has nonzero solutions $\iff$ $\mathrm{Ker}(L_A)$ non-trivial. Moreover, if $A \cdot \vec{x} = \vec{b}$ and $A \cdot \vec{y} = \vec{0}$, then $A \cdot (\vec{x} + \vec{y}) = \vec{b}$ as well by linearity.

$\hookrightarrow$ **Proposition 3.1**

For $A \in M_{m\times n}(\mathbb{F})$ and $b \in \mathrm{Im}(L_A)$ the set of solutions to $A\vec{x} = \vec{b}$ is precisely the coset $\vec{v} + \mathrm{Ker}(L_A)$ where $\vec{v} \in \mathbb{F}^n$ is a particular solution to $A\vec{x} = \vec{b}$; $A\vec{v} = \vec{b}$.

Proof. $\vec{v}+$ an element of $\mathrm{Ker}(L_A)$ is a solution to $A\vec{x} = \vec{b}$. Conversely, if $\vec{v}, \vec{w}$ are solutions to $A\vec{x} = \vec{b}$, then $A \cdot (\vec{v} - \vec{w}) = \vec{b} - \vec{b} = \vec{0}$ so $\vec{v} - \vec{w} \in \mathrm{Ker}(L_A)$, thus $\vec{w} = \vec{v} + (\vec{v} - \vec{w}) \in \vec{v} + \mathrm{Ker}(L_A)$. ■

$\hookrightarrow$ **Corollary 3.1**

If $m < n$ and $A \in M_{m\times n}(\mathbb{F})$, then there is always a nonzero solution to the homogeneous equation $A\vec{x} = \vec{0}$

Proof. $\mathrm{nullity}\,(L_A) = n - \mathrm{rank}(L_A) = n - \dim(\mathrm{Im}(L_A)) \geqslant n - m > 0$ hence $\mathrm{Ker}(L_A)$ nontrivial. ■

$\hookrightarrow$ Lecture 19; Last Updated: Mon Mar 25 13:48:03 EDT 2024

$\hookrightarrow$ **Corollary 3.2**

For $A \in M_{m\times n}(\mathbb{F})$,

1. $\mathrm{Ker}(L_A) = \{0_{\mathbb{F}^n}\} \iff A\vec{x} = \vec{b}$ has at most one solution, for each $\vec{b} \in \mathbb{F}^m$.
2. If $n = m$, A is invertible $\iff A\vec{x} = \vec{b}$ has exactly one solution for each $\vec{b} \in \mathbb{F}^m$.

Proof. 1. follows from proposition 3.1. 2. follows from 1. ■

We would like to determine whether $A\vec{x} = \vec{b}$ has a solution (equivalently, if $\vec{b} \in \mathrm{Im}(L_A)$), and to solve it, determining a particular solution, and $\mathrm{Ker}\, L_A$.

3.2 Elementary Row/Column Operations, Matrices

↪ **Definition 3.1: Elementary Row (Column) Operations**

Let $A \in M_{m\times n}(\mathbb{F})$. An *elementary row (column) operation* is one of the following operations applied to A:

1. Interchanging any two rows (columns) of A;
2. Multiplying a row (column) by a nonzero scalar from $\mathbb{F}$;
3. Adding a scalar multiple of one row (column) to another.

Remark 3.1. *All of these operations are (clearly) invertible. Moreover, each of these operations can be seen as linear transformations $M_{m\times n}(\mathbb{F}) \to M_{m\times n}(\mathbb{F})$, and can thus be represented as $(m \cdot n) \times (m \cdot n)$ matrices.*

↪ **Definition 3.2: Elementary Matrix**

A matrix $E \in M_n(\mathbb{F})$ is called *elementary* if it is obtained from I_n by an elementary row/column operation.

⊛ **Example 3.1**

1. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ is obtained from I_3 by operation 1.; indeed, either swapping the last two rows or columns yields the same result.
2. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ is obtained from I_3 by operation 2.; again, either the row or column view yields the same.
3. $\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ is obtained from I_3 by operation 3.; again, either viewed as adding 2 times the second column to the first or 2 times the first row to the second.

↪ **Theorem 3.1: Elementary Matrices and Operations**

Each elementary matrix can be obtained either by a row or column operation of the same kind.

Proof. Clear by example. ■

↪ **Theorem 3.2**

For matrices $A, B \in M_{m\times n}(\mathbb{F})$, if B is obtained from A by an elementary row (column) operation of type (i), then $B = E \cdot A$ ($B = A \cdot E$) for the elementary matrix $E \in M_m(\mathbb{F})$ ($M_n(\mathbb{F})$) obtained from the identity matrix by the same operation as in obtaining B from A.

Conversely, if E is an elementary matrix then $E \cdot A$ ($A \cdot E$) is obtained from A by applying the same elementary operations as in obtaining E from the identity matrix.

↪ **Proposition 3.2**

Elementary matrices are invertible, and the inverse is also an elementary matrix of the same type.

Proof. This follows from the fact that each elementary operation is invertible, and as each elementary operation can be representing as an elementary matrix, the result is clear. ■

↪ Lecture 20; Last Updated: Thu Feb 22 21:48:02 EST 2024

↪ **Proposition 3.3**

1. If $A \in M_{m\times n}(\mathbb{F})$, $P \in \mathrm{GL}_m(\mathbb{F})$[18], and $Q \in \mathrm{GL}_n(\mathbb{F})$, then $\mathrm{rank}(P \cdot A) = \mathrm{rank}(A) = \mathrm{rank}(A \cdot Q)$
2. More generally, if $T : V \to W$ is a linear transformation, where V, W finite dimensional, and $S : W \to W$ and $R : V \to V$ are linear and invertible, then $\mathrm{rank}(S \circ T) = \mathrm{rank}(T) = \mathrm{rank}(T \circ R)$.

Proof. 1. follows directly from part 2., being a special case where $T = L_A, S = L_P, R = L_Q$.

We have that $\mathrm{rank}(T) = \dim(\mathrm{Im}(T))$, and as S an isomorphism, $S|_{\mathrm{Im}(T)}$ is injective and thus $S(\mathrm{Im}(T)) \cong \mathrm{Im}(T)$, by S, so in particular, $\mathrm{rank}(S \circ T) = \dim(S(\mathrm{Im}(T))) = \mathrm{rank}(\mathrm{Im}(T)) = \mathrm{rank}(T)$.

For the other equality, we have that $\mathrm{Im}(T \circ R) = T(R(V)) = T(V) = \mathrm{Im}(T)$ so $\mathrm{rank}(T) = \dim(\mathrm{Im}(T)) = \dim(\mathrm{Im}(T \circ R)) = \mathrm{rank}(T \circ R)$. ■

↪ **Corollary 3.3**

Elementary row/column operations (equivalently, multiplication by elementary matrices) are rank-preserving; if B obtained from A by a row/column operation, then $\mathrm{rank}(B) = \mathrm{rank}(A)$.

Proof. Elementary operations correspond to multiplication by elementary matrices as we have shown previously, which are further invertible by proposition 3.2, which hence do not change the rank by proposition 3.3. ■

---

[18] Denoting the space of invertible $m \times m$ matrices.

↪ **Theorem 3.3: Diagonal Matrix Form**

Every matrix $A \in M_n(\mathbb{F})$ can be transformed into a matrix B of the form

$$\left(\begin{bmatrix} I_r \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right),$$

where the top right and bottom left $[0]$'s are $n - r \times r$, the bottom $[0]$ is $n - r \times n - r$, using row, column operations. In particular, $r = \operatorname{rank}(A)$.

Proof. We prove by induction on n.

Base: If $n = 0$, $A = ()$ and we are done.

Inductive Step: Suppose $n \geqslant 1$ and the statement holds for $n - 1$. If A is all zeros, we are done. Else, A has some nonzero entry, and by swapping two rows and columns such that the entry is in the top left ($a_1 1$) of the matrix, and then multiplying by $a_1 1^{-1}$ such that it is equal to 1,

$$\begin{pmatrix} 1 & \star & \cdots & \star \\ \star & \ddots & & \\ \vdots & & \ddots & \\ \star & & & \ddots \end{pmatrix}.$$

We can then use row (resp. column) operations such that each cell below (resp. to the right of) the top left 1 is equal to 0 by subtracting $\star\cdot$ row (resp. column) one from each,

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

Applying induction the the $n - 1 \times n - 1$ matrix we have left over in the bottom right block, we can transform this block into the desired form by row/column operations, not affecting A itself. This gives us the desired form of A. ∎

↪ **Corollary 3.4**

For each $A \in M_n(\mathbb{F})$, there are invertible matrices $P, Q \in \mathrm{GL}_n(\mathbb{F})$ such that

$$B := P \cdot A \cdot Q$$

is of the form in theorem 3.3. Moreover, P and Q are products of elementary matrices.

Proof. Follows from row/column operations corresponding to left/right multiplication by elementary matrices.

■

↪ **Corollary 3.5**

Every invertible matrix $A \in \mathrm{GL}_n(\mathbb{F})$ is a product of elementary matrices.

Proof. Let $A \in \mathrm{GL}_n(\mathbb{F})$, so $\mathrm{rank}(A) = n$. Then, by corollary 3.4, there exists matrices $P, Q \in \mathrm{GL}_n(\mathbb{F})$ such that $PAQ = I_n$ hence $A = P^{-1}Q^{-1}$. P, Q are themselves products of elementary matrices and thus their inverses are, hence A itself is a product of elementary matrices. ■

↪ **Corollary 3.6**

$\mathrm{rank}(A) = \mathrm{rank}(A^t) \, \forall A \in M_n(\mathbb{F})$.

Remark 3.2. *We've already proven this, but we present an alternative approach.*

Proof. There are $P, Q \in \mathrm{GL}_n(\mathbb{F})$ such that $B = PAQ$ of the desired diagonal form where $r = \mathrm{rank}(A)$. Then, $B^t = Q^t A^t P^t$, and thus $\mathrm{rank}(B^t) = \mathrm{rank}(A^t)$. But $B^t = B$ so $\mathrm{rank}(B^t) = \mathrm{rank}(B) = \mathrm{rank}(A)$ and thus $\mathrm{rank}(A) = \mathrm{rank}(A^t)$ as desired. ■

↪ **Corollary 3.7**

The transpose of an invertible matrix is invertible, with $(A^t)^{-1} = (A^{-1})^t$.

Proof. $A \cdot A^{-1} = I_n = A^{-1} \cdot A \implies (A^{-1})^t \cdot A^t = I_n^t = I_n = A^t \cdot (A^{-1})^t$. ■

↪ Lecture 21; Last Updated: Sat Apr 6 10:19:07 EDT 2024

3.2.1 Application to Finding Inverse Matrix

If $A \in M_n(\mathbb{F})$ is invertible, then $A = E_1 \cdot \cdots \cdot E_k$ for some elementary matrices E_i, so $A^{-1} = E_k^{-1} \cdot \cdots \cdot E_1^{-1} \cdot I_n$.

Consider the augmented matrix $(A|I_n)$. Remark that $B \cdot (A|I_n) = (BA|BI_n)$, and in particular, $E_k^{-1} \cdots E_1^{-1} \cdot (A|I_n) = (I_n|A^{-1})$, ie, there are row operations that turn $(A|I_n)$ to $(I_n|A^{-1})$.

↪ **Theorem 3.4**

Let $A \in M_n(\mathbb{F})$ be invertible.

1. There are row operations that turn $(A|I_n)$ into $(I_n|A^{-1})$.
2. If row operations turn $(A|I_n)$ into $(I_n|B)$ then $B = A^{-1}$.

3.2.2 Solving Systems of Linear Equations

↪ **Definition 3.3**

For matrices $A_1, A_2 \in M_{m\times n}(\mathbb{F})$ and $\vec{b}_1, \vec{b}_2 \in \mathbb{F}^m$, the systems of linear equations $A_1 \cdot \vec{x} = \vec{b}_1$ and $A_2 \cdot \vec{x} = \vec{b}_2$ are called *equivalent* if their sets of solutions are equal.

In particular, any two systems with no solutions are equivalent.

↪ **Proposition 3.4**

If $G \in \mathrm{GL}_m(\mathbb{F})$ and $A \in M_{m\times n}(\mathbb{F}), \vec{b} \in \mathbb{F}^m$, then $G \cdot A\vec{x} = G \cdot \vec{b}$ is equivalent to $A\vec{x} = \vec{b}$

Proof. Multiply both sides from the left by G^{-1}. ■

↪ **Corollary 3.8**

Row operations applied to $(A|b)$ do not change the solution set of $A\vec{x} = \vec{b}$.

↪ **Definition 3.4: ref/rref**

Let $B \in M_{m\times n}(\mathbb{F})$. We say B is in *row echelon form* if

1. All zero rows are at the bottom, ie each nonzero row is above each zero row;
2. The first nonzero entry (called a pivot) of each row is the only nonzero entry in its column;
3. The pivot of each row appears to the right of the pivot of the previous row.

If all pivots are 1, then we say that B is in *reduced row echelon form*.

↪ **Theorem 3.5: Gaussian Elimination Theorem**

There is a sequence of row operations of types 1. and 3. that bring any matrix $A \in M_{m\times n}(\mathbb{F})$ to a row echelon form. Moreover, applying row operations of type 2. to a matrix in row echelon form results in a reduced row echelon form.

↪ *Lecture 22; Last Updated: Sat Mar 9 09:25:26 EST 2024*

⊛ **Example 3.2**

$$\begin{aligned} 3x_1 + \ 2x_2 + \ 3x_3 - \ 2x_4 &= 1 \\ x_1 + \ x_2 + \ x_3 \quad &= 3 \\ x_1 + \ 2x_2 + \ x_3 - \ x_4 &= 2 \end{aligned} \quad \rightsquigarrow \quad A := \begin{pmatrix} 3 & 2 & 3 & -2 \\ 1 & 1 & 1 & 0 \\ 1 & 2 & 1 & -1 \end{pmatrix}, \vec{b} := \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix},$$

so we have agumented matrix

$$(A|b) = \begin{pmatrix} 3 & 2 & 3 & -2 & | & 1 \\ 1 & 1 & 1 & 0 & | & 3 \\ 1 & 2 & 1 & -1 & | & 2 \end{pmatrix} \overset{\text{Gaussian Elimination}}{\rightsquigarrow} \begin{pmatrix} 1 & 0 & 1 & 0 & | & 1 \\ 0 & 1 & 0 & 0 & | & 2 \\ 0 & 0 & 0 & 1 & | & 3 \end{pmatrix},$$

so $r := \operatorname{rank}(A) = 3$ and $\operatorname{nullity}(L_A) = 4 - 3 = 1$, so we expect a solution as a particular solution plus an ideal (the kernel). Rewriting, we see that

$$\begin{aligned} x_1 \qquad + x_3 &= 1 \\ x_2 \qquad &= 2 \\ x_4 &= 3 \end{aligned} \quad \Longrightarrow \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 - t_1 \\ 2 \\ t_1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 3 \end{pmatrix} + t_1 \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix},$$

where $t_1 \in \mathbb{F}$ arbitrary. Moreover, since setting $t_1 = 0$ gives that $\vec{v} := (1, 2, 0, 3)^t$ a solution, then $t_1(-1, 0, 1, 0)^t$ is a solution to the homogeneous system $A\vec{x} = \vec{0}$, ie, $\vec{u} := (-1, 0, 1, 0)^t$ is a basis for the kernel of $\operatorname{Ker}(L_A)$.

↪ Theorem 3.6

For any system $A\vec{x} = \vec{b}$, using Gaussian elimination we obtain another system $A_1\vec{x} = \vec{b_1}$ where $(A_1|\vec{b_1})$ is the reduced echelon form of $(A|\vec{b})$. Then:

1. $A\vec{x} = \vec{b}$ has a solution $\iff$ $\operatorname{rank}(A_1|\vec{b_1}) = \operatorname{rank}(A_1) = \sharp$ of non-zero rows of A_1.
2. If a solution exists, then, denoting $r := \operatorname{rank}(A)$ and $n := \sharp$ columns of A, we have the general solution to $A\vec{x} = \vec{b}$ of the form
$$\vec{v} + t_1\vec{u}_1 + \cdots + t_{n-r}\vec{u}_{n-r}$$
where $\vec{v} \in \mathbb{F}^n$ and $\{\vec{u}_1, \ldots, \vec{u}_{n-r}\}$ a basis for $\operatorname{Ker}(L_A)$ = space of solutions to $A\vec{x} = \vec{0}$.

Proof. We will only prove 1.

Recall that $A\vec{x} = \vec{b}$ has a solution $\iff$ $\vec{b} \in \operatorname{Im}(L_A) = \operatorname{Span}(\text{columns of } A)$ $\iff$ $\operatorname{Span}(\text{columns of } A) = \operatorname{Span}(\text{columns of } (A|b))$ $\iff$ $\operatorname{rank}(A) = \operatorname{rank}((A|b))$. ■

↪ **Corollary 3.9**

The system $A\vec{x} = \vec{b}$ has a solution $\iff$ in the reduced echelon form $(A_1|\vec{b}_1)$ of the augmented matrix, we do not have a pivot in the last column.

↪ **Lemma 3.1**

Let $B \in M_{m\times n}(\mathbb{F})$ be obtained from $A \in M_{m\times n}(\mathbb{F})$ via a row operation. Then, for all $a_1, \ldots, a_n \in \mathbb{F}$,

$$a_1 A^{(1)} + \cdots + a_n A^{(n)} = \vec{0} \iff a_1 B^{(1)} + \cdots + a_n B^{(n)} = \vec{0}.$$

In particular, columns in A are linearly (in)dependent iff the corresponding columns in B are linearly (in)dependent.

Proof. Left as a (homework) exercise. ∎

↪ **Lemma 3.2**

Let B be the reduced row echelon form of $A \in M_{m\times n}(\mathbb{F})$. Then:

1. $\sharp$ non-zero rows of $B = \operatorname{rank}(B) = \operatorname{rank}(A) =: r$.
2. For each $i = 1, \ldots, r$, denote by j_i the pivot of the ith row. Then, $B^{(j_i)} = e_i \in \mathbb{F}^m$. In particular, $\{B^{(j_1)}, \ldots, B^{(j_r)}\}$ is linearly independent.
3. Each column of B without a pivot is in the span of the previous columns.

Proof. Follows from the definition of rref. ∎

↪ **Corollary 3.10**

The rref of a matrix is unique.

Proof. Left as a (homework) exercise. ∎

↪ Lecture 23; Last Updated: Mon Mar 25 13:48:03 EDT 2024

3.3 Determinant

The determinant, denoted $\det(A)$, of a square matrix $A \in M_n(\mathbb{F})$ is a scalar from $\mathbb{F}$, meant to equal 0 iff A is not invertible.

↪ Proposition 3.5

$A \in M_n(\mathbb{F})$ is invertible $\iff$ the columns of A are linearly independent $\iff$ the rows of A are linearly independent $\iff$ $\operatorname{rank}(A) = n$

Proof. A invertible $\iff$ L_A invertible $\iff$ L_A bijection $\iff$ L_A surjection $\iff$ $\operatorname{rank}(L_A) = \operatorname{rank}(A) = n$ ■

⊛ Example 3.3

Let $A \in M_3(\mathbb{R})$, $A = \begin{pmatrix} - & v_1 & - \\ - & v_2 & - \\ - & v_3 & - \end{pmatrix}$. If $\{v_1, v_2, v_3\}$ linear dependent, then $\dim(\operatorname{Span}(v_1, v_2, v_3)) \leqslant 2$, which happens iff the parallelepiped formed with sides v_1, v_2, v_3 is contained in a plane (is "flat"), iff the parallelepiped is a parallelogram, ie, has 0 volume. As such, we can make the notion of volume dependent on the orientation of v_1, v_2, v_3 such that permuting v_1, v_2, v_3 changes the sign of the volume. This gives us the idea of an "oriented volume", which we can define as our determinant. This has a clear meaning in $\mathbb{R}$, but it remains to show how we can generalize this to arbitrary fields, where such a "volume" does not have a concrete meaning.

We now aim to derive a general formula for the determinant of a matrix over an arbitrary field by observing several key characteristics of our parallelepiped constructed above, and using these to define a unique determinant formula with geometric motivations.

Observation 1

Scaling a vector in a parallelepiped scales the volume of the parallelepiped by the same scalar.

↪ Definition 3.5: multiinear form

A function $\delta : M_n(\mathbb{F}) \to \mathbb{F}$ is called (row) multilinear, or n-linear, if it is linear in every row, i.e. for each $i = 1, \ldots, n$,

$$\delta\begin{pmatrix} - & v_1 & - \\ & \vdots & \\ - & v_{i-1} & - \\ - & c \cdot \vec{x} + \vec{y} & - \\ - & v_{i+1} & - \\ & \vdots & \\ - & v_n & - \end{pmatrix} = c \cdot \delta\begin{pmatrix} - & v_1 & - \\ & \vdots & \\ - & v_{i-1} & - \\ - & \vec{x} & - \\ - & v_{i+1} & - \\ & \vdots & \\ - & v_n & - \end{pmatrix} + \delta\begin{pmatrix} - & v_1 & - \\ & \vdots & \\ - & v_{i-1} & - \\ - & \vec{y} & - \\ - & v_{i+1} & - \\ & \vdots & \\ - & v_n & - \end{pmatrix}.$$

⊛ Example 3.4

1. $\delta(A) := a_{11} \cdot a_{22} \cdot \cdots \cdot a_{nn}$ is n-linear.

2. Fix $j \in \{1, \ldots, n\}$. The function $\delta_j(A) := a_{1j} \cdot a_{2j} \cdot \cdots \cdot a_{nj}$ is n-linear.

*3. However, $\mathrm{tr}(A) := \sum_{i=1}^{n} a_{ii}$ is *not* n-linear; scalar multiplication fails.

↪ **Proposition 3.6**

For an n-linear form $\delta : M_n(\mathbb{F}) \to \mathbb{F}$, if $A \in M_n(\mathbb{F})$ has zero row, then $\delta(A) = 0$.

Proof. $\delta(A) = \delta\left(\begin{pmatrix} \vec{0} \\ \vdots \end{pmatrix}\right) = \delta\left(\begin{pmatrix} \vec{0} \\ \vdots \end{pmatrix} + \begin{pmatrix} \vec{0} \\ \vdots \end{pmatrix}\right) = \delta\left(\begin{pmatrix} \vec{0} \\ \vdots \end{pmatrix}\right) + \delta\left(\begin{pmatrix} \vec{0} \\ \vdots \end{pmatrix}\right) = \delta(A) + \delta(A) \implies \delta(A) = 0.$ ■

Observation 2

If two sides of the parallelepiped are equal, then the volume is 0 (the shape is "flat").

↪ **Definition 3.6: Alternating**

A n-linear form $\delta : M_n(\mathbb{F}) \to \mathbb{F}$ is called *alternating* if $\delta(A) = 0$ for any matrix A whose two equal rows.

↪ **Proposition 3.7**

Let $\delta : M_n(\mathbb{F}) \to \mathbb{F}$ be an alternating n-linear form. Then, if B is obtained from A by swapping two rows, then $\delta(B) = -\delta(A)$.

Proof. It suffices to show that swapping two consecutive rows changes the sign of the result. Suppose B is obtained from A by swapping rows 1 and 2, namely

$$B = \begin{pmatrix} - & A_{(2)} & - \\ - & A_{(1)} & - \\ & \vdots & \end{pmatrix}.$$

Then,

$$\delta \begin{pmatrix} - & A_{(1)} + A_{(2)} & - \\ - & A_{(1)} + A_{(2)} & - \\ & \vdots & \end{pmatrix} = 0,$$

since its first two rows are equal; OTOH,

$$\delta \begin{pmatrix} - & A_{(1)} + A_{(2)} & - \\ - & A_{(1)} + A_{(2)} & - \\ & \vdots & \end{pmatrix} = \delta(A) + \delta(B),$$

so $\delta(B) = -\delta(A)$. ■

↪ **Proposition 3.8**

A multilinear form $\delta : M_n(\mathbb{F}) \to \mathbb{F}$ is alternating $\iff$ $\delta(A) = 0$ for every matrix A with two equal consecutive rows.

Proof. Left as a (homework) exercise. ■

Observation 3

If $v_i = e_i$ for $i = 1, \ldots, n$, ie, our parallelepiped is the unit cube, then the volume, aptly, equals 1; it is "normalized".

↪ *Lecture 24; Last Updated: Mon Mar 25 13:48:03 EDT 2024*

↪ **Proposition 3.9**

Let $\delta : M_n(\mathbb{F}) \to \mathbb{F}$ be an alternating multilinear form. Then, for each matrix $A \coloneqq (a_{ij}) \in M_n(\mathbb{F})$, we have

$$\delta(A) = \sum_{\pi \in S_n} a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)} \delta(\pi I),$$

where

$$\pi I_n \coloneqq \begin{pmatrix} - & e_{\pi(1)} & - \\ & \vdots & \\ - & e_{\pi(n)} & - \end{pmatrix}.$$

Proof. Left as a (homework) exercise. ■

Remark 3.3. *Since δ alternating, we can use row swaps to bring any πI_n to I_n, thus $\delta(\pi I_n) = \pm\delta(I_n)$; $\pm$ depends on the number of row swaps needed, ie, the parity of the given permutation π.*

↪ **Definition 3.7: Parity**

For a permutation $\pi \in S_n$, we let $\sharp\pi \coloneqq$ number of inversions = number of pairs $i, j \in \{1, \ldots, n\}$ such that $i < j$ but $\pi(i) > \pi(j)$. We say π *even* (resp. *odd*) if $\sharp\pi$ *even* (resp. *odd*), and define $\operatorname{sgn}(\pi) \coloneqq (-1)^{\sharp\pi}$ the *sign* of π.

↪ **Proposition 3.10**

$\operatorname{sgn} : S_n \to (\{1, -1\}, \cdot)$ is a group homomorphism, that is -1 of transpositions. In particular,

1. $\operatorname{sgn}(\pi^{-1}) = \operatorname{sgn}(\pi)$
2. If π a product of k transpositions, $\tau_1 \cdot \tau_2 \cdots \tau_k$, then $k = \sharp\pi \mod 2$.

Proof. See Goren, Lemma 4.2.1.

For (a), we have that $\text{sgn}(\pi^{-1}) = \text{sgn}(\pi)^{-1} = \text{sgn}(\pi)$.

For (b), $\text{sgn}(\pi) = \text{sgn}(\tau_1 \cdots \tau_k) = \text{sgn}(\tau_1) \cdots \text{sgn}(\tau_k) = (-1)^k$ so $(-1)^{\sharp\pi} = (-1)^k$ and thus $k = \sharp\pi \mod 2$. ■

↪ Corollary 3.11: Of proposition 3.9

For any alternating multilinear form $\delta : M_n(\mathbb{F}) \to \mathbb{F}$ and $A := (a_{ij}) \in M_n(\mathbb{F})$,

$$\delta(A) = \sum_{\pi \in S_n} a_{1\pi(1)} \cdots a_{n\pi(n)} \cdot \text{sgn}(\pi) \cdot \delta(I_n).$$

In particular, δ is uniquely determined by its value on I_n.

Proof. By proposition 3.9, $\delta(A) = \sum_{\pi \in S_n} a_{1\pi(1)} \cdots a_{n\pi(n)} \delta(\pi I_n)$, so we need only to show that $\delta(\pi I_n) = \text{sgn}(\pi) \cdot \delta(I_n)$. Writing $\pi^= \tau_1 \cdots \tau_k$ as transpositions, we know that $(-1)^k = \text{sgn}(\pi)$ and each row swap corresponding to a τ_i changes the sign of δ. Applying each τ_i row swaps to I_n, we obtain πI_n and thus $\delta(\pi I_n) = (-1)^k \cdot \delta(I_n) = \text{sgn}(\pi) \cdot \delta(I_n)$. ■

↪ Theorem 3.7: Characterization of the Determinant

There is a *unique* normalized (ie is 1 on I_n) alternating multilinear form; we call such a form the *determinant* and denote det; namely,

$$\det(A) := \sum_{\pi \in S_n} \text{sgn}(\pi) \cdot a_{1\pi(1)} \cdots a_{n\pi(n)}.$$

Proof. Uniqueness follows from corollary 3.11. It remains to show that the given definition for det is a normalized, alternating, multilinear form.

Normalized: $\det(I_n) = \sum_{\pi \in S_n} \text{sgn}(\pi) \cdot a_{1\pi(1)} \cdots a_{n\pi(n)} = (-1)^0 \cdot 1 \cdots 1 = 1$, since each summand will be zero for any permutation other than the identity.

Multilinear: A linear combination of n-linear forms is itself an n-linear form, so it suffices to prove that for a fixed $\pi \in S_n$, $\delta_\pi : M_n(\mathbb{F}) \to \mathbb{F}$ given by $\delta_\pi(A) := a_{1\pi(1)} \cdots a_{n\pi(n)}$ is n-linear, which should be clear as a product of matrix entries.

Alternating: Suppose A has two equal rows, wlog $A_{(1)}, A_{(2)}$. We partition S_n into the disjoint union of even and odd permutations, denoting A_n the even permutations. Note that $S_n \setminus A_n = A_n \cdot (12)$, ie the coset of the transposition (12) of the subgroup A_n. Thus, $A_n \to A_n \cdot (12)$ via $\pi \mapsto \pi' := \pi \cdot (12)$ is a bijection, and our partition has two equal parts. Thus, we can rewrite det as

$$\begin{aligned}
\det(A) &= \sum_{\pi \in S_n} \text{sgn}(\pi) \cdot a_{1\pi(1)} \cdots a_{n\pi(n)} \\
&= \sum_{\pi \in A_n} \text{sgn}(\pi) a_{1\pi(1)} \cdots a_{n\pi(n)} + \sum_{\pi \in A_n} \underbrace{\text{sgn}(\pi')}_{=-\text{sgn}(\pi)} \underbrace{a_{1\pi'(1)}}_{a_{1\pi(2)}} \cdots \underbrace{a_{n\pi'(n)}}_{=a_{n\pi(n)}} \\
&= \sum_{\pi \in A_n} \text{sgn}(\pi) a_{1\pi(1)} \cdots a_{n\pi(n)} - \sum_{\pi \in A_n} \text{sgn}(\pi) a_{1\pi(1)} \cdots a_{n\pi(n)} = 0,
\end{aligned}$$

where the last line follows from $a_{1\pi(2)} = a_{2\pi(2)}$ and conversely $a_{2\pi(1)} = a_{1\pi(1)}$ by assumption, and thus the two partitioned summands are equal, of opposite sign. ■

3.3.1 Properties of the Determinant

↪ **Lemma 3.3**

Let $\delta : M_n(\mathbb{F}) \to \mathbb{F}$ be an alternating multilinear form. Then, for $A \in M_n(\mathbb{F})$ and an elementary matrix E, if E is of type

1. 1, then $\delta(E \cdot A) = -\delta(A)$;
2. 2, representing multiplying by a scalar $c \in \mathbb{F}$, then $\delta(E \cdot A) = c\delta(A)$;
3. 3, then $\delta(E \cdot A) = \delta(A)$.

Proof. 1. is a restatement of the alternating property, proposition 3.7, 2. is the definition of multilinearity.

For 3., suppose E adds $c\cdot$ row i to row j, and suppose wlog $i = 1, j = 2$. Then,

$$\delta(E \cdot A) = \delta(A_{(1)}, A_{(2)} + c \cdot A_{(1)}, A_{(3)}, \ldots, A_{(n)}) = \delta(A) + c \cdot \delta(A_{(1)}, A_{(1)}, A_{(3)}, \ldots, A_{(n)}) = \delta(A),$$

by definition of δ being alternating. ■

↪ **Theorem 3.8**

For $A \in M_n(\mathbb{F})$, $\det(A) = 0$ iff A noninvertible.

Proof. Let $E_1, \ldots, E_k$ be elementary matrices such that $A' \coloneqq E_1 \cdots E_k \cdot A$ is in rref, remaring that then $\det(A') = c \cdot \det(A)$ for some $c \in \mathbb{F}, c \neq 0$, by lemma 3.3. We also have that $\mathrm{rank}(A) = \mathrm{rank}(A')$, and $\mathrm{rank}(A') < n \iff A'$ has a zero row.

($\Longleftarrow$) if A' has a zero row, then by multilinearity, $\det(A') = 0$ and thus $\det(A) = 0$ as well.

($\Longrightarrow$) if A' has no zero row, then $A' = I_n$ and thus $\det(A') = 1$, and $\det(A) = c^{-1} \cdot 1 \neq 0$. ■

↪ **Theorem 3.9**

The determinant respects products, $\det(A \cdot B) = \det(A) \cdot \det(B)$, for all $A, B \in M_n(\mathbb{F})$.

Proof. Suppose first A noninvertible, so $\mathrm{rank}(A) < n$ and $\det(A) = 0$. Then

$$\mathrm{rank}(A \cdot B) = \mathrm{rank}(L_{AB}) = \mathrm{rank}(L_A \circ L_B) \leqslant \mathrm{rank}(L_A) = \mathrm{rank}(A) < n,$$

so $A \cdot B$ also noninvertible and $\det(A \cdot B) = 0$. Hence, $\det(A) \cdot \det(B) = 0 \cdot \det(B) = 0 = \det(A \cdot B)$.

Suppose now A invertible. Then, writing $A = E_1 \cdots E_k$ as a product of elementary matrices; it suffices to show, by induction, for a single E. By lemma 3.3, $\det(A) = \det(E \cdot I) = c$ for some non-zero constant $c \in \mathbb{F}$, so $\det(A) \cdot \det(B) = c \cdot \det(B)$. On the other hand, $\det(A \cdot B) = \det(E \cdot B) = c \cdot \det(B)$, also by lemma 3.3. ■

↪ **Corollary 3.12**

$\det(A^{-1}) = \det(A)^{-1}$, $\forall A \in \mathrm{GL}_n(\mathbb{F})$.

Proof. $1 = \det(I_n) = \det(A \cdot A^{-1}) = \det(A) \cdot (A^{-1}) \implies \det(A^{-1}) = \det(A)^{-1}$. ■

↪ **Corollary 3.13**

$\det(A^t) = \det(A) \, \forall A \in M_n(\mathbb{F})$.

Proof. If A noninvertible, then $\mathrm{rank}(A^t) = \mathrm{rank}(A) < n$ so both are noninvertible, and thus $\det(A^t) = \det(A) = 0$.

If A invertible, writing $A = E_1 \cdots E_k$, we have $A^t = E_k^t \cdots E_1^t$. For each $i = 1, \ldots, k$, E_i^t is an elementary matrix of the same type, with the same constant if of type 2, and thus $\det(E_i) = \det(E_i^t)$, and so

$$\det(A^t) = \det(E_k^t) \cdots \det(E_1^t) = \det(E_1) \cdots \det(E_k) = \det(A).$$

■

↪ Lecture 25; Last Updated: Sat Apr 6 10:19:07 EDT 2024

4 Diagonalization of Linear Operators

4.1 Introduction: Definitions of Diagonalization

This section will be concerned with decomposing a linear operator $T : V \to V$ for a finite dimensional V into a direct sum of simpler linear operators.

The simplest linear operator we could consider is multiplication by a fixed scalar; ideally, then, we would like to be able, for any operator $T : V \to V$, to decompose $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$ of T-invariant subspaces such that $T|_{V_i}$ is just multiplication by some scalar λ_i.

↪ **Definition 4.1: Linearly Independent Subspaces**

For subspaces $V_1, V_2, \ldots, V_k \subseteq V$, we say that $\{V_1, \ldots, V_k\}$ is *linearly independent* if

$$V_i \cap \sum_{j \neq i} V_j = \{0_V\},$$

then, we call $V_1 + V_2 + \cdots + V_k$ a *direct sum* and denote $V_1 \oplus V_2 \oplus \cdots \oplus V_k$.

↪ **Definition 4.2: Diagonalization**

Call a linear operator $T : V \to V$ *diagonalizable* if it admits a *diagonalization*, ie

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k,$$

where each V_i is a subspace of V, such that $T|_{V_i}$ is just multiplication by a fixed scalar $\lambda_i \in \mathbb{F}$.

⊛ **Example 4.1**

1. If A a diagonal matrix, $A = \begin{pmatrix} \lambda_1 & 0 & \cdots \\ 0 & \ddots & 0 \\ \cdots & 0 & \lambda_n \end{pmatrix}$, then L_A is diagonalizable; take $V_i := \mathrm{Span}(\{e_i\})$, then $\mathbb{F}^n = V_1 \oplus \cdots \oplus V_n$.

2. If A not diagonal, but is similar to a diagonal matrix D as above ie $\exists Q \in \mathrm{GL}_n(\mathbb{F})$ s.t. $A = QDQ^{-1}$. Then, as any invertible matrix $Q = [I_n]_\alpha^\beta$ is a change of basis matrix, denoting $\beta := \{v_1, \ldots, v_n\}$, then letting $V_i := \mathrm{Span}(\{v_i\})$ gives the appropriate decomposition such that $L_A|_{V_i} =$ mult. by λ_i. We generalize this below.

↪ **Proposition 4.1**

Let V, $\dim(V) < \infty$. A linear operator $T : V \to V$ is diagonalizable iff there is a basis β for V such that $[T]_\beta^\beta$ is diagonal.

Proof. ($\Longrightarrow$) Suppose $V = V_1 \oplus \cdots \oplus V_k$ such that $T|_{V_i} =$ mult. by λ_i. Let β_i be a basis for V_i, then, $\beta := \cup_{i=1}^k \beta_i$ is a basis for V. Then, for each $v \in \beta$, $v \in \beta_i$ for some i and so $T(v) = \lambda_i \cdot v$ and thus $[T(v)]_\beta = \begin{pmatrix} 0 \\ \vdots \\ \lambda_i \\ \vdots \\ 0 \end{pmatrix}$, and so

$$[T]_\beta = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

($\Longleftarrow$) Suppose $\beta := \{v_1, \ldots, v_n\}$ a basis such that $[T]_\beta$ is diagonal. Then, taking $V_i := \mathrm{Span}(\{v_i\})$, $[T(v_i)] = \lambda_i \cdot e_i = \lambda_i \cdot [v_i]_\beta = [\lambda_i v_i]_\beta$. $v \mapsto [v]_\beta$ injective, and thus $Tv_i = \lambda_i v_i$. ∎

4.2 Eigenvalues/vectors/spaces

↪ **Definition 4.3: Eigenvalue/eigenvector**

For a linear operator $T : V \to V$ and $\lambda \in \mathbb{F}$, λ is called an *eigenvalue* of T if there is a non-zero vector $v \in V$ such that $T(v) = \lambda \cdot v$. Then, v is called an *eigenvector*.

↪ Lecture 26; Last Updated: Sat Apr 6 12:29:01 EDT 2024

↪ **Proposition 4.2**

For a finite dimensional vector space V and a linear transformation $T : V \to V$, TFAE:

1. T is diagonalizable, ie $V = \bigoplus_{i=1}^{k} V_i$ s.t. $T|_{V_i}$ scalar multiplication for each i.
2. There is a basis β for V such that $[T]_\beta^\beta$ is diagonal.
3. There is a basis β consisting of eigenvectors of T.

Proof. (1. $\iff$ 2.) proposition 4.1.

(2. $\implies$ 3.) Suppose $\beta := \{v_1, \ldots, v_n\}$ a basis such that $[T]_\beta$ a diagonal matrix with entries λ_i. Then, $[T(v_j)]_\beta = \lambda_j e_j$ so $T(v_j) = \lambda_j v_j$ and thus v_j an eigenvector.

(3. $\implies$ 2.) Let $\beta := \{v_1, \ldots, v_n\}$ a basis of eigenvectors such that $T(v_j) = \lambda_j v_j$ for some $\lambda_j \in \mathbb{F}$. Then

$$[T]_\beta = \begin{pmatrix} | & | & & | \\ [T(v_1)]_\beta & [T(v_2)]_\beta & \cdots & [T(v_n)]_\beta \\ | & | & & | \end{pmatrix}$$

But $[T(v_j)]_\beta = [\lambda_j v_j]_\beta = \lambda_j e_j$, so this matrix is diagonal with entries λ_j. ■

↪ **Proposition 4.3**

For $A \in M_n(\mathbb{F})$, A is diagonalizable, ie L_A diagonalizable, $\iff \exists Q \in GL_n(\mathbb{F})$ s.t. $Q^{-1}AQ$ is diagonal; the columns of Q are eigenvectors, forming a basis for $\mathbb{F}^n$.

Proof. A diagonalizable $\iff$ there is a basis β for $\mathbb{F}^n$ such that $[L_A]_\beta$ diagonal. Then, letting α be the standard basis, we have that $A = [L_A]_\alpha = [I]_\beta^\alpha \cdot [L_A]_\beta \cdot [I]_\alpha^\beta = [I]_\beta^\alpha \cdot [L_A]_\beta \cdot ([I]_\beta^\alpha)^{-1}$ so $[L_A]_\beta = ([I]_\beta^\alpha)^{-1} \cdot A \cdot [I]_\beta^\alpha$. Letting $Q := [I]_\beta^\alpha$, we get $Q^{-1}AQ$ diagonal. The columns of Q are exactly the vectors in β, and thus eigenvectors. ■

↪ **Definition 4.4: Eigenspace**

For an eigenvalue λ of $T : V \to V$, let $\mathrm{Eig}_V(\lambda) := \{v \in V : Tv = \lambda v\}$, called the *eigenspace* of T corresponding to λ.

↪ **Proposition 4.4**

$\text{Eig}_V(\lambda)$ a subspace of V.

Remark 4.1. *Diagonalizability is a conjugate-invariant property; if $A \sim B$ and A diagonalizable, then so is B.*

↪ **Proposition 4.5**

The trace, tr, and determinant, det, functions $M_n(\mathbb{F}) \to \mathbb{F}$ are conjugation-invariant.

↪ **Definition 4.5**

Let V, $\dim(V) = n.$ and $T : V \to V$ a linear operator. Define tr (resp. det) of T as $\text{tr}(T) \coloneqq \text{tr}\big([T]_\beta\big)$ $(\det(T) \coloneqq \det\big([T]_\beta\big))$ for some/any basis β for V.

Remark 4.2. *This is well-defined (doesn't depend on the choice of basis), $[T]_\alpha$, $[T]_\beta$ are conjugate for any two bases, and* tr, det *are conjugate-invariant.*

↪ **Proposition 4.6**

$\dim(V) = n, T : V \to V$ invertible $\iff \det(T) \neq 0$.

Proof. T invertible $\iff [T]_\beta$ invertible $\iff \det\big([T]_\beta\big) \neq 0$ for some basis β. ■

↪ **Proposition 4.7**

Let $T : V \to V, \dim(V) < \infty$.

1. $v \in V$ an eigenvector of T with eigenvalue $\lambda \iff v \in \text{Ker}(\lambda I - T)$.
2. $\lambda \in \mathbb{F}$ an eigenvalue $\iff \lambda I - T$ non-invertible $\iff \det(\lambda I - T) = 0$.

Proof. 1. $T(v) = \lambda v \iff \lambda v - T(v) = 0 \iff (\lambda I_V - T)(v) = 0 \iff v \in \text{Ker}(\lambda I_V - T)$.

2. follows from 1. by the dimension theorem. ■

↪ Lecture 27; Last Updated: Mon Apr 8 11:43:09 EDT 2024

↪ **Corollary 4.1**

For $A \in M_n(\mathbb{F})$, $\lambda \in \mathbb{F}$ an eigenvalue of A (that is, if L_A) $\iff \det(\lambda I - A) = 0$.

Proof. Follows from the previous proposition by noting that $[\lambda I_{\mathbb{F}^n} - L_A]$ in the standard basis of $\mathbb{F}^n$ is just $\lambda I_n - A$. ■

↪ **Proposition 4.8**

1. For $A \in M_n(\mathbb{F})$, the function $t \mapsto \det(tI_n - A)$ is a polynomial in t of the form

$$p_A(t) := t^n - \operatorname{tr}(A)t^{n-1} + \cdots + (-1)^n \det(A)$$

and is called the *characteristic polynomial* of A.

2. For a n-dim V and $T : V \longrightarrow V$, the function $t \mapsto \det(tI_V - T)$ is a polynomial of the form

$$p_T(t) := t^n - \operatorname{tr}(T)t^{n-1} + \cdots + (-1)^n \det(T).$$

Proof. 1. a homework exercise; 2. follows immediately. ■

Hence, this proposition gives that the eigenvalues of A are precisely the roots of $p_A(t)$.

↪ **Corollary 4.2**

$T : V \longrightarrow V$ has at most n distinct eigenvalues.

⊛ **Example 4.2**

Let $A := \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 4 \\ 0 & 0 & 4 \end{pmatrix}$. Then

$$-p_A(t) = \det(A - tI_n) = \det \begin{pmatrix} 3-t & 1 & 0 \\ 0 & 3-t & 4 \\ 0 & 0 & 4-t \end{pmatrix} = (3-t)^2(4-t),$$

with roots $t = 3, 4$ and thus A has two eigenvalues $\lambda_1 := 3$ mult. 2 and $\lambda_2 := 4$. Then:

$$\operatorname{Eig}_A(\lambda_1) = \operatorname{Ker}(3I - L_A) = \{\vec{x} \in \mathbb{F}^3 : (A - 3I)\vec{x} = 0\},$$

hence, $\vec{x} \in \operatorname{Eig}_A(\lambda_1)$ are the solutions to the homogeneous system $(A - 3I)\vec{x} = 0$:

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 4 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \iff \begin{cases} x_2 = 0 \\ x_3 = 0 \end{cases} \iff \vec{x} = ae_1, a \in \mathbb{F},$$

so $\operatorname{Eig}_A(3) = \operatorname{Span}(\{e_1\})$. A similar computation gives $\operatorname{Eig}_A(\lambda)(2) = \operatorname{Span}(\{(1, 1, \frac{1}{4})\})$.

We have hence found two 1-dimensional eigenspaces; A is thus not diagonalizable.

↪ Proposition 4.9

Let $\lambda_1, \dots, \lambda_k$ be distinct eigenvalues of $T : V \to V$ on V n-dim. Then if v_i an eigenvector of T corresponding to λ_i, then $\{v_1, \dots, v_k\}$ is linearly independent. In particular, $k \leqslant n$.

Proof. By induction on k. If $k = 1$ then $\{v_1\}$ is linear independent because $v_1 \neq 0_V$. Suppose the proposition holds for k. Let $\lambda_1, \dots, \lambda_{k+1}$ be distinct eigenvalues with corresponding $\{v_1, \dots, v_{k+1}\}$ eigenvectors. Let

$$① \qquad a_1 v_1 + \cdots + a_{k+1} v_{k+1} = 0_V.$$

Taking $T(①)$, we have

$$② \qquad \lambda_1 a_1 v_1 + \cdots + \lambda_{k+1} a_{k+1} v_{k+1} = 0_V.$$

Then, $② - \lambda_{k+1} \cdot ①$ yields

$$(\lambda_1 - \lambda_{k+1}) a_1 v_1 + \cdots + (\lambda_k - \lambda_{k+1}) a_k v_k = 0_V,$$

but $v_1, \dots, v_k$ linearly independent by assumption, so $(\lambda_i - \lambda_{k+1}) a_i = 0$ for $i = 1, \dots, k$. The λ_i's distinct, hence it must be that $a_i = 0$ for $i = 1, \dots, k$, and so ① gives that $a_{k+1} v_{k+1} = 0_V$. But v_{k+1} an eigenvalue, so this is only possible if $a_{k+1} = 0$ and the proof is complete. ■

↪ Corollary 4.3

For distinct eigenvalues $\lambda_1, \dots, \lambda_k$ of $T : V \to V$, $\dim(V) < \infty$, the corresponding eigenspaces $\mathrm{Eig}_T(\lambda_i)$ are linearly independent.

Proof. This follows directly proposition 4.9. ■

↪ Definition 4.6: Geometric Multiplicity

For eigenvalue λ of $T : V \to V$, denote by $m_g(\lambda) \coloneqq \dim(\mathrm{Eig}_T(\lambda))$ and call it the *geometric multiplicity* of λ.

↪ Corollary 4.4

For $T : V \to V$ with distinct eigenvalues $\lambda_1, \dots, \lambda_k$,

$$\sum_{i=1}^{k} m_g(\lambda_i) \leqslant n.$$

Proof. $\sum_{i=1}^{k} m_g(\lambda_i) = \dim(\bigoplus_{i=1}^{k} \mathrm{Eig}_T(\lambda_i)) \leqslant n$. ■

↪ Lecture 28; Last Updated: Fri Apr 5 13:26:26 EDT 2024

↪ **Theorem 4.1**

Let $V, n := \dim(V)$. A linear operator $T : V \to V$ is diagonalizable iff the sum of the geometric multiplicities of all of the eigenvalues $\lambda_1, \ldots, \lambda_k$ equals n, ie iff

$$\sum_{i=1}^{k} m_g(\lambda_i) = n.$$

Proof. Recall that T diagonalizable iff $\exists$ a basis consisting of eigenvectors.

($\Longrightarrow$) If $\beta := \{v_1, \ldots, v_n\}$ a basis for V of eigenvectors, then each $v_i \in \mathrm{Eig}_T(\lambda_j)$ for some j, so $\beta \subseteq \cup_{i=1}^{k} \mathrm{Eig}_T(\lambda_i)$ and $\beta \cap \mathrm{Eig}_T(\lambda_i)$ is linearly independent, hence $|\beta \cap \mathrm{Eig}_T(\lambda_i)| \leqslant m_g(\lambda_i)$. Thus, $n = |\beta| = \sum_{i=1}^{k} \left|\beta \cap \mathrm{Eig}_T(\lambda_i)\right| \leqslant \sum_{i=1}^{k} m_g(\lambda_i)$. By the previous corollary, it follows that $\sum_{i=1}^{k} m_g(\lambda_i) = n$.

($\Longleftarrow$) Suppose $\sum_{i=1}^{k} m_g(\lambda_i) = n$ and let β_i a basis for $\mathrm{Eig}_T(\lambda_i)$. By the linear independence of the eigenspaces, $\beta := \cup_{i=1}^{k} \beta_i$ still linearly independent and, having n elements, is a basis for V consisting of eigenvectors by construction. ■

⊛ **Example 4.3**

Let $D : \mathbb{F}[t]_2 \to \mathbb{F}[t]_2$ by $p(t) \mapsto p'(t)$. To find eigenvalues of D, we fix the basis $\alpha := \{1, t, t^2\}$ for D and find the corresponding matrix representation

$$[D]_\alpha = \begin{pmatrix} | & | & | \\ [D(1)]_\alpha & [D(t)]_\alpha & [D(t^2)]_\alpha \\ | & | & | \end{pmatrix} = \begin{pmatrix} | & | & | \\ [0]_\alpha & [1]_\alpha & [2t]_\alpha \\ | & | & | \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Thus,

$$p_D(t) = -\det([D]_\alpha - tI_3) = -\begin{pmatrix} -t & 1 & 0 \\ 0 & -t & 2 \\ 0 & 0 & -t \end{pmatrix} = t^3,$$

hence, the only eigenvalue is $\lambda = 0$, with corresponding $\mathrm{Eig}_D(0) = \mathrm{Ker}(D - 0 \cdot I) = \mathrm{Ker}(D)$, so $m_g(0) = \dim(\mathrm{Ker}(D)) = 3 - \mathrm{rank}(D) = 3 - \mathrm{rank}([D]_\alpha) = 1$. Moreover, D is not diagonalizable.

↪ **Definition 4.7: Algebraic Multiplicity**

For $V, \dim(V) < \infty$, and a linear operator $T : V \to V$ and an eigenvalue λ of T, we define the *algebraic multiplicity* of λ to be the multiplicity of λ as the root of $p_T(t)$, ie the largest $k \geqslant 1$ such that $(t - \lambda)^k \mid p_T(t)$. We denote this by

$$m_a(\lambda).$$

↪ Lemma 4.1

Let V, $\dim(V) < \infty$ and $T : V \to V$ be linear. For each T-invariant subspace $W \subseteq V$, let $T_W := T|_W : W \to W$. Then,

$$p_{T_W}(t) \mid p_T(t).$$

Proof. Let $\alpha := \{v_1, \ldots, v_k\}$ be a basis for W and extend it to a basis $\beta := \alpha \cup \{v_{k+1}, \ldots, v_n\}$ for V. Leting $A := [T_W]_\alpha$, we see that

$$\begin{aligned} [T]_\beta &= \begin{pmatrix} | & & | & | & & | \\ [T(v_1)]_\beta & \cdots & [T(v_k)]_\beta & [T(v_{k+1})]_\beta & \cdots & [T(v_n)]_\beta \\ | & & | & | & & | \end{pmatrix} \\ &= \begin{pmatrix} & \star & \\ A & & \star \\ & \star & \\ \mathbf{0} & & \star \\ & \star & \end{pmatrix}, \end{aligned}$$

where $\mathbf{0}$ is a $n - k \times k$ matrix of zeros. Hence,

$$p_T(t) = -\det\big([T]_\beta - tI_n\big) = -\det(\cdots) = -\det(A - tI_k) \cdot \det(B - tI_{n-k}) = -p_{T_W}(t)\det(B - tI_{n-k}),$$

and the proof is complete. ∎

↪ Proposition 4.10

Let V, $\dim(V) < \infty$, and $T : V \to V$. For each eigenvalue λ of T, $m_g(\lambda) \leqslant m_a(\lambda)$.

Proof. Let $W := \mathrm{Eig}_T(\lambda)$, which is T-invariant, so by lemma 4.1, $p_T(t) = p_{T_W}(t) \cdot q(t)$ for some $q(t) \in \mathbb{F}[t]$. But, fixing any basis $\alpha := \{v_1, \ldots, v_k\}$ for W, we have that $T_W(v_i) = T(v_i) = \lambda v_i$ so $[T(v_i)]_\alpha = \lambda e_i \in \mathbb{F}^k$ hence $[T_W]_\alpha$ is just a $k \times k$ diagonal matrix with λ entries. Thus, $p_{T_W}(t) = \det(tI_k - [T_W]_\alpha) = (t-\lambda)^k$, and so $p_T(t) = (t-\lambda)^k \cdot q(t)$ and thus $m_a(\lambda) \geqslant k = \dim(W) = m_g(\lambda)$. ∎

↪ Definition 4.8: Splits

A polynomial $p(t) \in \mathbb{F}[t]$ *splits* over $\mathbb{F}$ if $p(t) = a \cdot (t - r_1) \cdots (t - r_n)$ for some $a \in \mathbb{F}$, $r_1, \ldots, r_n \in \mathbb{F}$.

Remark 4.3. *If $\mathbb{F}$ is algebraically closed, then every polymomial over $\mathbb{F}$ splits over $\mathbb{F}$.*

Remark 4.4. *For an eigenvalue λ of $T : V \to V$, where V is n-dimensional, $p_T(t)$ splits iff $\sum_{i=1}^k m_a(\lambda_i) = n$.*

↪ Theorem 4.2: Main Criterion of Diagonalizability

Let V, $\dim(V) < \infty$, $T : V \to V$ linear. Then T diagonalizable iff $p_T(t)$ splits and $m_g(\lambda) = m_a(\lambda)$ for each eigenvalue λ of T.

Proof. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T. Then,

$$T \text{ diagonalizable} \iff \sum_{i=1}^{k} m_g(\lambda_i) = n := \dim(V)$$

since $m_g(\lambda_i) \leqslant m_a(\lambda_i)$ and $\sum_{i=1}^{k} m_a(\lambda_i) \leqslant n$, we have that

$$n = \sum_{i=1}^{k} m_g(\lambda_i) \iff m_g(\lambda_i) = m_a(\lambda_i), \quad i = 1, \ldots, k, \text{ and } \sum_{i=1}^{k} m_a(\lambda_i) = n,$$

but this last statement is equivalent to saying that $p_T(t)$ splits. ■

↪ *Lecture 29; Last Updated: Mon Mar 25 13:48:03 EDT 2024*

⊛ **Example 4.4**

1. $A := \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}$, so $L_A : \mathbb{F}^3 \longrightarrow \mathbb{F}^3$. Then,

$$p_A(t) = -\det \begin{pmatrix} 4-t & 0 & 1 \\ 2 & 3-t & 2 \\ 1 & 0 & 4-t \end{pmatrix} = -(4-t)(3-t)(4-t) + 1 \cdot (3-t) \cdot 2 = -(t-5)(t-3)^2.$$

Supposing $\operatorname{char}(\mathbb{F}) \neq 2$ ie $3 \neq 5$, then we have two distinct eigenvalues $\lambda_1 = 5, \lambda_2 = 3$ with $m_a(5) = 1, m_a(3) = 2$, so the polynomial splits (regardless of $\mathbb{F}$). We have that $1 \leqslant m_g(5) \leqslant m_a(5) = 1$, so $m_g(5) = m_a(5) = 1$. We need only to check that $m_g(3) = 2$; but we have that

$$\begin{aligned} m_g(3) &= \operatorname{nullity}(L_A - 3 \cdot I) = 3 - \operatorname{rank}(L_A - 3 \cdot I) = 3 - \operatorname{rank}(A - 3I) \\ &= 3 - \operatorname{rank} \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{pmatrix} = 3 - 1 = 2 = m_a(3), \end{aligned}$$

so A indeed diagonalizable. A conjugate of A that is diagonal is $D := \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$, and if v_1 an eigenvector for $\lambda_1 = 5$ and v_2, v_3 are linearly independent eigenvectors for $\lambda_2 = 3$, then

$$Q := \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} = [I_3]_{\beta}^{\alpha},$$

where $\alpha := \{e_1, e_2, e_3\}$ and $\beta := \{v_1, v_2, v_3\}$, is such that

$$D = Q^{-1}AQ.$$

In the case that $\mathrm{char}(\mathbb{F}) = 2$, $3 = 5$ so we hae a single eigenvalue $\lambda = 1 = 3 = 5$ with $m_a(1) = 3$. But we still have that $\mathrm{rank}(A - I) = \mathrm{rank}\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} = 1$ so $m_g(1) = 2 < 3$, hence A is not diagonalizable.

2. Let $T : \mathbb{F}^2 \to \mathbb{F}^2$ be a rotation by ninety degrees, so $T(e_1) = e_2$ and $T(e_2) = -e_1$. Then, $T = L_A$ with

$$A = [T]_\alpha = \begin{pmatrix} | & | \\ e_2 & -e_1 \\ | & | \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

with α the standard basis. Then

$$p_T(t) = p_A(t) = -\det\begin{pmatrix} -t & -1 \\ 1 & -t \end{pmatrix} = t^2 + 1,$$

which doesn't split over $\mathbb{F} := \mathbb{R}$, but does over $\mathbb{F} := \mathbb{C}$ or any $\mathbb{F}$ with characteristic 2 where $t^2 + 1 = (t+1)^2$.

When $\mathbb{F} := \mathbb{C}$, $p_T(t) = (t - i)(t + i)$ so we have 2 distinct eigenvalues with each having algebraic multiplicity 1, hence both have geometric multiplicity of 1 and thus T is diagonalizable.

When $\mathrm{char}(\mathbb{F}) = 2$, we have a single eigenvalue $\lambda = 1$, with

$$m_g(1) = \mathrm{nullity}(T - I) = 2 - \mathrm{rank}(T - I) = 2 - \mathrm{rank}\begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix} = 2 - \mathrm{rank}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = 1 < 2 = m_a(1),$$

so T is not diagonalizable.

Remark 4.5. *From the previous two examples, regard that the issue of diagonalizability is a field-related issue; not only because of the "splittability" of polynomials, but because of characteristic.*

4.3 T-cyclic Vectors and the Cayley-Hamilton Theorem

↪ **Definition 4.9: T-cyclic subspace**

Let V be any vector space, $T : V \to V$ a linear operator, and $v \in V$. The *T-cyclic subspace* of/generated by v is the space

$$\mathrm{Span}(\{v, T(v), T^2(v), \ldots, \}) = \mathrm{Span}(\{T^n(v) : n \in \mathbb{N}\}).$$

Remark 4.6. *Note that T-cyclic subspaces are T-invariant. In a sense, T-cyclic subspaces are "minimal T-invariant subspaces". Recall too that the characteristic polynomial of T restricted to T-invariant subspaces divides the characteristic polynomial of T by lemma 4.1.*

↪ **Lemma 4.2**

Let V be finite dimensional, $T : V \longrightarrow V$ linear, and $v \in V$. Let $W :=$ the T-cyclic subspace generated by v.

1. $\{v, T(v), \ldots, T^{k-1}(v)\}$ is a basis for W, where $k := \dim(W)$.
2. Since $T^k(v) \in \operatorname{Span}(\{v, T(v), \ldots, T^{k-1}(v)\})$, we have a unique representation $T^k(v) = a_0 v + a_1 T(v) + \cdots + a_{k-1}T^{k-1}(v)$. Then,
$$p_{T_W}(t) = t^k - a_{k-1}t^{k-1} - \cdots - a_1 t - a_0$$

Proof. Left as homework.

Hint for 2.: use $\beta := \{v, \ldots, T^{k-1}(v)\}$ representation of $[T_W]_\beta$. ■

Remark 4.7. *Note that if V itself T-cyclic for some v, then T "satisfies" its own characteristic polynomial. Indeed, $p_T(t) = t^n - a_{n-1}t^{n-1} - \cdots - a_0$ and so*

$$p_T(T) := T^n - a_{n-1}T^{n-1} - \cdots - a_0 I_V$$

is equal to 0 on v, and hence on all vectors $u \in V$ since $V = \operatorname{Span}(\{v, T(v), \ldots, T^{n-1}(v)\})$ because

$$p_T(T)(T^i)(v) = T^{n+i}(v) - a_{n-1}T^{n-1+i}(v) - \cdots - a_0 T^i(v) = (T^i \circ p_T(T))(v) = T^i(p_T(v)) = T^i(0) = 0.$$

Even more generally, we have that this is true in general, precisely:

↪ **Theorem 4.3: Cayley-Hamilton Theorem**

Let V be finite dimensional and $T : V \longrightarrow V$ be linear. Then T satisfies its own characteristic polynomial $p_T(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_0$, ie

$$p_T(T) = T^n + a_{n-1}T^{n-1} + \cdots + a_0 I_V \equiv 0_V.$$

Proof. Fix $v \in V$. Let $W := T$-cyclic subspace generated by v, so $p_{T_W}(t) | p_T(t)$, ie $p_T(t) = q(t) \cdot p_{T_W}(t)$. Hence $p_T(T) = q(T) \circ p_{T_W}(T)$, and thus

$$p_T(T)(v) = q(T)(p_{T_W}(T)(v)) \overset{\text{lemma 4.2}}{=} q(T)(0) = 0.$$

■

↪ **Corollary 4.5: Cayley-Hamilton for Matrices**

For every $A \in M_n(\mathbb{F})$, $p_A(A) = 0$.

Proof. Follows immediately from $[L_A]_\alpha = A$. ■

↪ *Lecture 30; Last Updated: Mon Mar 25 13:48:03 EDT 2024*

5 Inner Product Spaces

5.1 Introduction: Inner Products, Norms, Basic Properties

For this section, $\mathbb{F}$ will always be either $\mathbb{R}$ or $\mathbb{C}$.

↪ Definition 5.1: Inner Product

Let V be a vector space over $\mathbb{F}$. An *inner product* on V is a function

$$V \times V \longrightarrow \mathbb{F}, \qquad (u, v) \mapsto \langle u, v \rangle,$$

satisfying, for all $u, v, w \in V$ and $\alpha \in \mathbb{F}$,

1. Linear in the first coordinate:
   (a) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$
   (b) $\langle \alpha u, v \rangle = \alpha \cdot \langle u, v \rangle$
2. Skew-symmetric:
   (a) $\langle u, v \rangle = \overline{\langle v, u \rangle}$
3. $\langle u, u \rangle \geqslant 0$, and equal to 0 iff $u = 0_V$.

V together with $\langle ., . \rangle$ is called an *inner product space*.

Unless otherwise stated, all vector spaces V should be considered as an inner product space from here on.

Remark 5.1. *Note that the third requirement is well-defined; that is, it follows from 2. that $\langle u, u \rangle \in \mathbb{R}$, since $\langle u, u \rangle = \overline{\langle u, u \rangle}$, ie $\langle u, u \rangle$ is equal to its own complex conjugate, which is only possible if its imaginary part is precisely 0. So, it makes sense to require it to be geq 0 (if it was complex, this would be meaningless).*

↪ Definition 5.2

Let $\langle ., . \rangle$ be an inner product on V. The *norm* associated to this inner product is defined

$$||v|| := \sqrt{\langle v, v \rangle}, \quad v \in V.$$

We call $v \in V$ a *unit vector* if $||v|| = 1$. For $v \in V, v \neq 0$, we call $||v||^{-1} \cdot v$ the *normalization of* v.

Remark 5.2. Never *work with a norm directly; working with the square of the norm is far easier.*

↪ **Proposition 5.1**

Let V be an inner product space. For each $u, v, w \in V$ and $\alpha \in \mathbb{F}$,

1. Conjugate linearity in the second coordinate holds:

   (a) $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$

   (b) $\langle u, \alpha v \rangle = \overline{\alpha} \langle u, v \rangle$

2. $||\alpha \cdot v|| = |\alpha| \cdot ||v||$

3. $||v, 0_V|| = 0 = ||0_V, v||$

Proof. 1.(a), (b) follow from skew-symmetry.

For 2., we have $||\alpha v||^2 = \langle \alpha v, \alpha v \rangle = \alpha \cdot \overline{\alpha} \langle v, v \rangle = |\alpha|^2 \cdot ||v||^2$.

For 3., follows from $\langle 0_V, v \rangle + \langle 0_V, v \rangle = \langle 0_V, v \rangle$. ■

⊛ **Example 5.1**

1. For $V := \mathbb{F}^n$, the standard inner product is the "dot product"; for $\vec{x} := (x_1, \dots, x_n), \vec{y} := (y_1, \dots, y_n)$,

$$\langle \vec{x}, \vec{y} \rangle := \vec{x} \cdot \vec{y} := \sum_{i=1}^{n} x_i \overline{y_i},$$

which gives

$$||\vec{x}|| = \sqrt{\sum_{i=1}^{n} |x_i|^2},$$

that is, the standard Euclidean norm.

↪ **Proposition 5.2**

For $\mathbb{F} := \mathbb{R}$ and $\vec{x}, \vec{y} \in \mathbb{R}^n$, $\vec{x} \cdot \vec{y} = ||\vec{x}|| ||\vec{y}|| \cos \alpha$, where α the angle from $\vec{x}$ to $\vec{y}$.

2. If $\langle ., . \rangle$ an inner product on V and r a positive real, then $\langle ., . \rangle_r := r \cdot \langle ., . \rangle$ is also an inner product.

3. Let $V := C[0, 1]$. Define for $f, g \in V$,

$$\langle f, g \rangle := \int_0^1 f(t) \cdot \overline{g(t)} \, \mathrm{d}t \, .$$

4. Let $V := \mathbb{F}[t]_n$. For $f(t) := a_0 + a_1 t + \cdots + a_n t^n, g(t) := b_0 + b_1 t + \cdots + b_n t^n$, define

$$\langle f, g \rangle_1 := \sum_{i=0}^{n} a_i \overline{b_i},$$

and

$$\langle f, g \rangle_2 := \int_0^1 f(t) \overline{g(t)} \, dt \, .$$

These are both inner products.

5. For $A \in M_{n \times m}(\mathbb{F})$, let $A^* := \overline{A}^t$ the *conjugate transpose of* A.[19] For $V := M_n(\mathbb{F})$ and $A, B \in V$, define

$$\langle A, B \rangle := \operatorname{tr}(B^* \cdot A).$$

It is left as a (homework) exercise to verify that this is a well-defined inner product.

↪ *Lecture 31; Last Updated: Sat Apr 6 12:27:25 EDT 2024*

5.2 Projections and Cauchy-Schwartz

↪ Definition 5.3: Orthogonal

Let V be an inner product space. Call $u, v \in V$ *orthogonal*, and write $u \perp v$, if $\langle u, v \rangle = 0$.

⊛ Example 5.2

In $\mathbb{R}^3$ equipped with the dot product, $(1, 0, -1) \perp (1, 0, 1)$.

↪ Theorem 5.1: Pythagorean Theorem

For an inner product space V and $u, v \in V$, if $u \perp v$ then

$$||u||^2 + ||v||^2 = ||u + v||^2.$$

In particular, $||u||, ||v|| \leqslant ||u + v||$.

Proof.

$$||u + v||^2 = \langle u + v, u + v \rangle = \langle u, u \rangle + \overset{=0}{\langle u, v \rangle} + \overset{=0}{\langle v, u \rangle} + \langle v, v \rangle = ||u||^2 + ||v||^2.$$

■

---

[19] Where $\overline{A} := (\overline{a_{ij}})$.

↪ **Definition 5.4**

For vectors u, v in an inner product space V, if u is a unit vector, then put

$$\operatorname{proj}_u(v) := \langle v, u\rangle \cdot u.$$

↪ **Proposition 5.3**

Let V be an inner product space and $u \in V$ a unit vector. For each $v \in V$, $v - \operatorname{proj}_u(v) \perp u$. In particular, $v = \operatorname{proj}_u(v) + w$ where $w := v - \operatorname{proj}_u(v) \perp \operatorname{proj}_u(v)$.

Proof.

$$\langle v - \operatorname{proj}_u(v), u\rangle = \langle v, u\rangle - \langle \operatorname{proj}_u(v), u\rangle = \langle v, u\rangle - \langle v, u\rangle \cdot \langle u, u\rangle = \langle v, u\rangle - \langle v, u\rangle = 0.$$

■

↪ **Corollary 5.1**

Let V be an inner product space and $u \in V$ a unit vector. For each $v \in V$, $||\operatorname{proj}_u(v)|| \leqslant ||v||$.

Proof. $\operatorname{proj}_u(v) \perp w := v - \operatorname{proj}_u(v)$, hence $||\operatorname{proj}_u(v)|| \leqslant ||\operatorname{proj}_u(v) + w|| = ||v||$ by the Pythagorean theorem. ■

↪ **Theorem 5.2**

Let V be an inner product space and $x, y \in V$.

(a) (Cauchy-Banyakovski-Schwartz inequality) $|\langle x, y\rangle| \leqslant ||x|| \cdot ||y||$.

(b) (Triangle inequality) $||x + y|| \leqslant ||x|| + ||y||$.

Proof. (a) If $||y|| = 0$ then $y = 0_V$ and $0 \leqslant 0$ and we are done. Suppose $||y|| \neq 0$ and divide both sides by $||y||$:

$$\langle x, ||y||^{-1} \cdot y\rangle \leqslant ||x||,$$

ie, we need to prove $|\langle x, y\rangle| \leqslant ||x||$, where u a unit. But

$$|\langle x, u\rangle| = ||\langle x, u\rangle \cdot u|| = ||\operatorname{proj}_u(x)|| \leqslant ||x||$$

by the previous corollary.

(b) We equivalently prove $||x + y||^2 \leqslant (||x|| + ||y||)^2$. We have:

$$\begin{aligned} ||x + y||^2 = \langle x + y, x + y\rangle &= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\ &\leqslant ||x||^2 + ||y||^2 + 2\,|\langle x, y\rangle| \\ &\overset{\text{(by CBS)}}{\leqslant} ||x||^2 + ||y||^2 + 2||x||||y|| = (||x|| + ||y||)^2. \end{aligned}$$

■

⊛ **Example 5.3**

1. For $\mathbb{F}^n$, CS claims that $\left|\sum_{i=1}^n x_i y_i\right| \leqslant \sqrt{\sum_{i=1}^n |x_i|^2}\sqrt{\sum_{i=1}^n |y_i|^2}$, but $\langle x, y\rangle = ||x||||y|| \cos\alpha$, so this simply follow from $|\cos\alpha| \leqslant 1$.
2. For $f, g \in C[0,1]$, $\int_0^1 f(t)g(t)\,\mathrm{d}t \leqslant \sqrt{\int_0^1 |f(t)|^2\,\mathrm{d}t}\sqrt{\int_0^1 |g(t)|^2\,\mathrm{d}t}$.

From the triangle inequality, it is natural to define $d : V \times V \longrightarrow [0,\infty)$ $d(u,v) \coloneqq ||u - v||$ as the "distance" between vectors u, v; indeed, one can show that such a d defines a metric on V.

↪ **Proposition 5.4: The Parallelogram Law**

For an inner product space V and $u, v \in V$,

(a) $2||u||^2 + 2||v||^2 = ||u+v||^2 + ||v-u||^2$.

(b) $\mathrm{Re}\langle u, v\rangle = \frac{1}{2}\left(||u||^2 + ||v||^2 - ||v-u||^2\right)$

Proof. Let as a (homework) exercise. ■

5.3 Orthogonality and Orthonormal Bases

↪ **Definition 5.5: Orthogonal/Orthonormal**

Call a set $S \subseteq V$ *orthogonal* (resp. *orthonormal*) if the vectors in S are pair-wise orthogonal to each (resp. in addition, they are unit).

↪ **Proposition 5.5**

Orthonormal sets of nonzero vectors are linearly independent.

Proof. Suppose $a_1 v_1 + \cdots + a_n v_n = 0_V$, $v_1, \ldots, v_n$ orthogonal. Then

$$\begin{aligned} \langle a_1 v_1 + \cdots + a_n v_n, v_i\rangle &= \langle 0_V, v_i\rangle = 0 \\ \implies \sum_{j=1}^n a_j \langle v_j, v_i\rangle &= a_i \underbrace{\langle v_i, v_i\rangle}_{\neq 0}, \end{aligned}$$

hence a_i's identically zero. ■

↪ *Lecture 32; Last Updated: Mon Apr 8 13:29:53 EDT 2024*

↪ **Definition 5.6: Orthonormal Basis**

Let V be an inner product space over $\mathbb{F}$. An *orthonormal basis* β for V is a basis that is orthonormal.

⊛ **Example 5.4: Of Orthognormal Bases**

(a) For $\mathbb{F}^n$, the standard basis is orthonormal with respect to the dot product; $\langle e_i, e_j\rangle = \delta_{ij}$.

(b) For $\mathbb{F}^4$ with the dot product, $\alpha := \{(1,0,1,0)^t, (1,0,-1,0)^t, (0,1,0,1)^t, (0,1,0,-1)^t\}$ is an orthogonal basis; remark that to show this we need only to show that each vector is orthogonal by proposition 5.5. We can turn this into an ortho*normal* basis by normalizing each vector:

$$||(1,0,1,0)||^2 = 1+0+1+0 = 2 \implies ||(1,0,1,0)|| = \sqrt{2},$$

and indeed each vector has norm $\sqrt{2}$, so

$$\beta := \{\frac{1}{\sqrt{2}} \cdot v : v \in \alpha\}$$

now an orthonormal basis.

↪ **Proposition 5.6: Benefits of Orthonormal Bases**

Let $\beta := \{u_1, u_2, \ldots, u_n\}$ be an orthonormal basis for V. Then:

(a) For every $v \in V$, the coordinates of v in β are just $\langle v, u_i\rangle$ ie

$$\begin{aligned} v &= \langle v, u_1\rangle \cdot u_1 + \langle v, u_2\rangle \cdot u_2 + \cdots + \langle v, u_n\rangle \cdot u_n \\ &= \mathrm{proj}_{u_1}(v) + \mathrm{proj}_{u_2}(v) + \cdots + \mathrm{proj}_{u_n}(v). \end{aligned}$$

In this case, the coefficients $\langle v, u_i\rangle$ are called the *Fourier coefficients* of v in β.

(b) For any linear operator $T : V \to V$, $[T]_\beta = (\langle Tu_j, u_i\rangle)_{i,j}$, ie

$$[T]_\beta = \begin{pmatrix} \langle Tu_1, u_1\rangle & \langle Tu_2, u_1\rangle & \cdots & \langle Tu_n, u_1\rangle \\ \langle Tu_1, u_2\rangle & \langle Tu_2, u_2\rangle & \cdots & \langle Tu_n, u_2\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle Tu_1, u_n\rangle & \langle Tu_2, u_n\rangle & \cdots & \langle Tu_n, u_n\rangle \end{pmatrix}.$$

In particular, remark that $\langle Tu_j, u_i\rangle$ is the (ij)th element.

Proof. (a) Let $v = a_1u_1 + \cdots + a_nv_n$ be the unique representation of v in β. Taking the inner product with u_i

on both sides, then, we get

$$\langle v, u_i \rangle = \sum_{j=1}^{n} a_j \langle u_j, u_i \rangle = \sum_{j=1}^{n} a_j \delta_{ji} = a_i.$$

(b) The jth column of $[T]_\beta$ is $[Tu_j]_\beta = (\langle Tu_j, u_1 \rangle, \langle Tu_j, u_2 \rangle, \ldots, \langle Tu_j, u_n \rangle)^t$, by part (a).

■

Clearly, orthonormal bases are quite convenient; but does one always exist? More precisely, does every inner product space admit an orthonormal basis? We will show that the finite dimensional ones always do.

↪ **Definition 5.7: Orthogonality to a Set**

For a set $S \subseteq V$ and $v \in V$, we say that v is *orthogonal to* S and write $v \perp S$ if v is orthogonal to all vectors in S.

↪ **Proposition 5.7**

$v \perp V \iff v = 0_V$

Proof. Let as a homework exercise. ■

↪ **Lemma 5.1**

Suppose $\alpha := \{u_1, \ldots, u_k\}$ is an orthonormal set. For each $v \in V$, the vector

$$\operatorname{proj}_\alpha(v) := \sum_{i=1}^{k} \operatorname{proj}_{u_i}(v) = \sum_{i=1}^{k} \langle v, u_i \rangle u_i$$

has the property that $v \operatorname{proj}_\alpha(v) \perp \alpha$, in particular, $v = \operatorname{proj}_\alpha(v) \perp \operatorname{proj}_\alpha(v)$.

Thus, $v = \operatorname{proj}_\alpha(v) + \operatorname{orth}_\alpha(v)$ where $\operatorname{orth}_\alpha(v) := v - \operatorname{proj}_\alpha$, where $\operatorname{proj}_\alpha(v) \perp \operatorname{orth}_\alpha(v)$.

Proof. We need to show that $v - \operatorname{proj}_\alpha(v) \perp u_j$ for each $j = 1, \ldots, k$. Fix j, then

$$\begin{aligned} \langle v - \operatorname{proj}_\alpha(v), u_j \rangle &= \langle v - u_j \rangle - \langle \operatorname{proj}_\alpha, u_i \rangle \\ &= \langle v, u_j \rangle - \sum_{i=1}^{k} \langle v, u_i \rangle \underbrace{\langle u_i, u_j \rangle}_{=\delta_{ij}} \\ &= \langle v, u_j \rangle - \langle v, u_j \rangle = 0. \end{aligned}$$

■

5.4 Gram-Schmidt Algorithm

We describe now a process to

$$\underset{\text{independent set}}{\{v_1, v_2, \ldots, v_k\}} \rightsquigarrow \underset{\text{orthonormal set}}{\{u_1, u_2, \ldots, u_k\}}$$

with the property that for all $\ell = 1, \ldots, k$, $\text{Span}(\{v_1, \ldots, v_\ell\}) = \text{Span}(\{u_1, \ldots, u_\ell\})$.

The ℓth step of the process takes

$$\{\underbrace{u_1, \ldots, u_{\ell-1}}_{\text{orthonormal}}, v_\ell\} \rightsquigarrow \underbrace{\{u_1, \ldots, u_{\ell-1}, u_\ell\}}_{\substack{\text{orthonormal} \\ \text{Span}(\{u_1,\ldots,u_{\ell-1},v_\ell\})=\text{Span}(\{u_1,\ldots,u_{\ell-1},u_\ell\})}}.$$

Concretely, we replace v_ℓ with

$$v'_\ell := \text{orth}_{\{u_1,\ldots,u_{\ell-1}\}}(v_\ell) \equiv v_\ell - \text{proj}_{\{u_1,\ldots,u_{\ell-1}\}}(v_\ell) \equiv v_\ell - \sum_{i=1}^{\ell-1} \langle v_\ell, u_i \rangle u_i.$$

By lemma 5.1, this is indeed orthogonal to the preceding vectors; we need simply now to normalize it, namely $u_\ell := ||v_\ell||^{-1} \cdot v'_\ell$.

⊛ Example 5.5

$v_1 := (1, 0, 1, 0), v_2 := (1, 1, 1, 1), v_3 := (0, 1, 2, 1)$.

First we take $u_1 := ||v_1||^{-1} v_1 = \frac{1}{\sqrt{2}}(1, 0, 1, 0)$.

Then $v'_2 = v_2 - \langle v_2, u_1 \rangle u_1 = v_2 - \frac{1}{\sqrt{2}}(1+1)\frac{1}{\sqrt{2}}(1, 0, 1, 0) = (1, 1, 1, 1) - (1, 0, 1, 0) = (0, 1, 0, 1)$. Normalizing, $u_2 := \frac{1}{\sqrt{2}}(0, 1, 0, 1)$.

Finally, $v'_3 = v_3 - \langle v_3, u_1 \rangle u_1 - \langle v_3, u_2 \rangle u_2 = (-1, 0, 1, 0)$, and so $u_3 := \frac{1}{\sqrt{2}}(-1, 0, 1, 0)$, giving us a final orthonormal set

$$\{\frac{1}{\sqrt{2}}(1, 0, 1, 0), \frac{1}{\sqrt{2}}(0, 1, 0, 1), \frac{1}{\sqrt{2}}(-1, 0, 1, 0).\}$$

↪ Corollary 5.2

Every finite dimensional inner product space admits an orthonormal basis.

Proof. Feed any basis to the process above. ■

↪ Lecture 33; Last Updated: Fri Apr 5 13:54:28 EDT 2024

5.5 Orthogonal Complements and Orthogonal Projections

↪ **Definition 5.8: Orthogonal Complement**

Let V be an inner product set. For a set $S \subseteq V$, its *orthogonal complement* is the subspace

$$S^{\perp} := \{v \in V : v \perp S\}.$$

↪ **Proposition 5.8**

$S^{\perp}$ indeed a subspace as in the definition (even if S is not).

Proof. Let $v, w \in S^{\perp}$, $a \in \mathbb{F}$. Then for each $s \in S$, $\langle v + aw, s \rangle = \langle v, s \rangle + a \cdot \langle w, s \rangle = 0 + a \cdot 0$, hence $v + aw \in S^{\perp}$. ■

Remark 5.3. *We previously used $S^{\perp}$ to denote the annihilator of S, with $S^{\perp} \subseteq V^*$, ie the linear functionals that are 0 on S, while now we are talking about $S^{\perp} \subseteq V$ as the set of vectors orthogonal to S; this is slightly abusive notation. We shall see why to follow (indeed, we have a natural bijection between the two, which we shall show).*

↪ **Theorem 5.3**

Let V be an inner product space and let $W \subseteq V$ be a finite dimensional subspace.

(a) For each $v \in V$, there is a unique decomposition $v = w + w_{\perp}$ such that $w \in W$ and $w_{\perp} \in W^{\perp}$. We call such a w the *orthogonal projection* of v onto W, and denote it $\mathrm{proj}_W(v)$.

(b) $V = W \oplus W^{\perp}$. In particular, if $\dim(V) < \infty$, then

$$\dim(W^{\perp}) = \dim(V) - \dim(W).$$

Proof. (a) Existence: Let $\alpha := \{w_1, w_2, \ldots, w_k\}$ be an orthonormal basis for W, which exists since $\dim(W) < \infty$ (corollary 5.2). Let $w := \mathrm{proj}_{\alpha}(v)$, then, $w_{\perp} := v - w$ is orthogonal to α by lemma 5.1, hence orthogonal to the span $\mathrm{Span}(\alpha) = W$.

Uniqueness: Suppose there exist two such decompositions, $w + w_{\perp} = v = w' + w'_{\perp}$. Note that since $v - w$ and $v - w'$ are both orthogonal to W, so is their difference, ie $v - w, v - w' \in W^{\perp} \implies (v - w) - (v - w') = w' - w \in W^{\perp}$, being a subspace. But $w - w' \in W$ as well, and is also orthogonal to 0, so it must be that $w - w' = 0_V$ and thus $w = w'$.

(b) By (a), $V = W + W^{\perp}$. It remains to show that $W \cap W^{\perp}\{0_V\}$; but for $w \in W$, $w \in W$ and $w \in W^{\perp}$ simultaneously only if $w = 0_V$.

■

Remark 5.4. *If α, β two different orthonormal bases for a finite dimensional subspace W, then $\mathrm{proj}_{\alpha}(v) = \mathrm{proj}_{\beta}(v)$ for all $v \in V$, because $\mathrm{proj}_W(v)$ is unique.*

↪ **Theorem 5.4**

For any finite dimensional subspace $W \subseteq V$ and for each $v \in V$, the orthogonal projection $\mathrm{proj}_W(v)$ is the unique closest vector to V in W.

Proof. Left as a (homework) exercise. ■

↪ **Proposition 5.9**

Let $W \subseteq V$ be a finite dimensional subspace. Then

(a) $\text{proj}_W : V \to V$ a linear operator.

(b) A linear operator $T : V \to V$ is a projection (onto $\text{Im}(T)$) operator iff $\text{Ker}(T) = \text{Im}(T)^\perp$.

Proof. Let as a (homework) exercise. ■

↪ **Corollary 5.3**

Let $W \subseteq V$ be a finite dimensional subspace. Then $(W^\perp)^\perp = W$.

Proof. By definition $W \subseteq (W^\perp)^\perp$; we show the converse. Let $v \in (W^\perp)^\perp$. Then, $v = w + w_\perp$ for some vectors $w \in W$ and $w_\perp \in W^\perp$. We know $\langle v, w_\perp \rangle = 0$, so

$$\begin{aligned} ||v||^2 = \langle v, v \rangle = \langle v, w + w_\perp \rangle &= \langle v, w \rangle + \langle v, w_\perp \rangle \\ &= \langle v, w \rangle = \langle v, w_\perp \rangle = \langle w + w_\perp, w_\perp \rangle = \langle w, w \rangle = ||w||^2. \end{aligned}$$

On the other hand, $||v||^2 = ||w||^2 + ||w_\perp||^2$, so it must be that $||w_\perp||^2 = 0$ hence $w_\perp = 0_V$ and thus $v = w \in W$ and the proof is complete. ■

5.6 Riesz Representation and Adjoint

Let V be an inner product space. For each $w \in V$, we can define a linear functional $f_w \in V^*$ as follows: $f_w(v) := \langle v, w \rangle$. It turns out that for a finite dimensional V, every linear functional is of this form.

↪ **Theorem 5.5: Riesz Representation Theorem**

Let V be a finite dimensional inner product space. Then, for each $f \in V^*$, there is a unique $w \in V$ such that $f = f_w$, ie $f(v) = \langle v, w \rangle$ for all $v \in V$.

On other words, the map $V \to V^*, w \mapsto f_w$ is a linear isomorphism.

Proof. Existence: fix $f \in V^*$ and let $\beta := \{v_1, \ldots, v_n\}$ be an orthonormal basis for V. Then, for each $v \in V$, $v = \langle v, v_1 \rangle v_1 + \cdots + \langle v, v_n \rangle v_n$ hence

$$\begin{aligned} f(v) &= \langle v, v_1 \rangle f(v_1) + \cdots + \langle v, v_n \rangle f(v_n) \\ &= \langle v, \overline{f(v_1)} v_1 \rangle + \cdots + \langle v, \overline{f(v_n)} v_n \rangle \\ &= \langle v, \overline{f(v_1)} v_1 + \cdots + \overline{f(v_n)} v_n \rangle, \end{aligned}$$

hence, taking $w := \overline{f(v_1)}v_1 + \cdots + \overline{f(v_n)}v_n$ gives us existence.

Uniqueness: Suppose $f_{w_1} = f = f_{w_2}$ so $f_{w_1 - w_2} = f_{w_1} - f_{w_2} = 0_{V^*}$ ie $\forall\, v \in V$, $\langle v, w_1 - w_2\rangle = f_{w_1 - w_2} = 0$. Hence, $w_1 - w_2 = 0 \implies w_1 = w_2$ and uniqueness holds.

As such, existence gives us injectivity and uniqueness gives us surjectivity of $w \mapsto f_w$. ∎

↪ Lecture 34; Last Updated: Mon Apr 8 13:46:12 EDT 2024

↪ Theorem 5.6: Adjoint

Let V be finite dimensional, $T : V \to V$. There exists a unique linear operator $T^* : V \to V$ called the *adjoint* of T such that for all two vectors $v, w \in V$,

$$\langle T(v), w\rangle = \langle v, T^*(w)\rangle.$$

Remark 5.5. *Because this is an implicit definition, we must always work with this definition; there's no real way to work with T^* directly*

Proof. For a fixed $w \in V$, define $\tilde{f}_w \in V^*$ by $\tilde{f}_w(v) := \langle Tv, w\rangle$, which is indeed a linear functional on V (to check). By theorem 5.5, there is a unique element $\tilde{w} \in V$ such that $\tilde{f}_w = f_{\tilde{w}}$, ie $\tilde{f}_w(v) = \langle Tv, w\rangle = \langle v, \tilde{w}\rangle = f_{\tilde{w}}$ for any $v \in V$. Setting $T^*(w) := \tilde{w}$, we find that T^* fulfills the required definition; we need only to check T^* linear.

Let $w_1, w_2 \in V$, $a \in \mathbb{F}$, then $T^*(aw_1 + w_2)$ the unique vector $u \in V$ such that $\langle Tv, aw_1 + w_2\rangle = \langle v, T^*(a_1 w_1 + w_2)\rangle$, so it suffices to check that $aT^*w_1 + T^*w_2$ also satisfies this (by uniqueness). Indeed,

$$\langle Tv, aw_1 + w_2\rangle = \overline{a}\langle Tv, w_1\rangle + \langle Tvw_2\rangle = \overline{a}\langle v, T^*w_1\rangle + \langle v, T^*w_2\rangle = \langle v, aT^*w_1 + T^*w_2\rangle,$$

and so this must equal $\langle v, T^*(aw_1 + w_2)\rangle$ by uniqueness. ∎

↪ Proposition 5.10: Matrix Representation of Adjoint

(a) Let $T : V \to V$ be a linear operator on a finite dimensional V and let β be an *orthonormal* basis for V. Then

$$[T^*]_\beta = [T]_\beta^*,$$

where, for $A \in M_n(\mathbb{F})$, A^* denotes its conjugate transpose/adjoint of A, for clear reasons.

(b) For any $A \in M_n(\mathbb{F})$, the adjoint of $L_A : \mathbb{F}^n to \mathbb{F}^n$ is L_{A^*} ie $L_A^* = L_{A^*}$.

Proof. (a) Recall that the (ij)th entry of $[T^*]_\beta$ with $\beta := \{v_1, \ldots, v_n\}$ is $\langle T^*v_j, v_i\rangle$, which equals $\overline{\langle v_i, T^*(v_j)\rangle} = \overline{\langle Tv_i, v_j\rangle} = \overline{(ji)\text{th entry of } [T]_\beta}$, hence $[T^*]_\beta = \overline{[T]_\beta^t} = [T]_\beta^*$.

(b) This is a special case of (a) with β being the standard basis, ie $v_i = e_i$. We have $[L_A^*]_\beta$ is the matrix B such that $L_A^* = L_B$, and by (a) $B = [L_A]_\beta^* = A^*$.

∎

↪ **Proposition 5.11: Adjoint versus Other Operations**

Let $T : V \longrightarrow V$ on V with V finite dimensional. Then:

(a) $T \mapsto T^* : \mathrm{Hom}(V, V) \longrightarrow \mathrm{Hom}(V, V)$ is conjugate linear.

(b) $(T_1 \circ T_2)^* = T_2^* \circ T_1^*$.

(c) $I_V^* = I_V$.

(d) $(T^*)^* = T$.

(e) If T invertible, so is T^* and $(T^*)^{-1} = (T^{-1})^*$.

Proof. We prove (a), the rest are left as (homework) exercises. For any $v, w \in V$,

$$\langle (T_1 + T_2)(v), w\rangle = \langle T_1 v, w\rangle + \langle T_2 v, w\rangle = \langle v, T_1^* w\rangle + \langle v, T_2^* w\rangle = \langle v, T_1^* w + T_2^* w\rangle = \langle v, (T_1^* + T_2^*)w\rangle.$$

Similarly, for $a \in \mathbb{F}$, we have for all $v, w \in V$,

$$\langle aT(v), w\rangle = a\langle Tv, w\rangle = \langle v, \overline{a}T^* w\rangle = \langle v, (\overline{a}T^*)w\rangle.$$

■

↪ **Proposition 5.12: Kernel and Image of Adjoint**

Let $T : V \longrightarrow V$, V finite dimensional. Then

(a) $\mathrm{Im}(T^*)^\perp = \mathrm{Ker}(T)$;

(b) $\mathrm{Ker}(T^*) = \mathrm{Im}(T)^\perp$.

Proof. Remark that because $\dim(V) < \infty$, $\mathrm{Im}(T^*) = \mathrm{Ker}(T)^\perp \iff \mathrm{Im}(T^*)^\perp = \mathrm{Ker}(T)$.

For each $v \in V$,

$$\begin{aligned} v \in \mathrm{Im}(T^*)^\perp &\iff \forall\, u \in \mathrm{Im}(T^*), \langle v, u\rangle = 0 \iff \forall\, w \in V, \langle v, T^* w\rangle = 0 \\ &\iff \forall\, w \in V, \langle Tv, w\rangle = 0 \iff Tv = 0_V \iff v \in \mathrm{Ker}(T) \end{aligned}$$

(b) Apply (a) to T^*, ie $\mathrm{Im}(T^{**})^\perp = \mathrm{Ker}(T^*)$, but $T^{**} = T$ and the proof is complete.

■

↪ **Corollary 5.4**

Let $T : V \longrightarrow V$ on V n-dimensional inner product space. Then $\mathrm{rank}(T) = \mathrm{rank}(T^*)$ and $\mathrm{nullity}(T) = \mathrm{nullity}(T^*)$.

Proof. $\operatorname{rank}(T^*) = \dim(\operatorname{Im}(T^*)) = \dim(\operatorname{Ker}(T)^\perp) = n - \operatorname{nullity}(T) = \operatorname{rank}(T)$ and it follows by the dimension theorem that $\operatorname{nullity}(T^*) = n - \operatorname{rank}(T^*) = n - \operatorname{rank}(T) = \operatorname{nullity}(T)$. ■

↪ Lecture 35; Last Updated: Wed Apr 10 13:40:28 EDT 2024

↪ Corollary 5.5

Let $T : V \to V$, V finite dimensional. For $\lambda \in \mathbb{F}$, λ an eigenvalue iff $\overline{\lambda}$ an eigenvalue of T^*.

Remark 5.6. *But the corresponding eigenvectors may be different in general.*

Proof. λ an eigenvalue of $T \iff \operatorname{nullity}(T - \lambda I_V) > 0 \iff \operatorname{nullity}((T - \lambda I_V)^*) = \operatorname{nullity}(T^* - \overline{\lambda} I_V) > 0 \iff \overline{\lambda}$ an eigenvalue of T^*. ■

↪ Lemma 5.2: Schur's Lemma (Orthonormal Version)

Let $T : V \to V$ on V finite dimensional and suppose that $p_T(t)$ splits. Then there is an orthonormal basis β for V such that $[T]_\beta$ upper triangular.

Proof. Because $p_T(t)$ splits, T, hence by corollary 5.5 also T^*, has eigenvalues. We prove by induction on $n := \dim(V)$. For $n = 1$, matrix is upper triangular so we are done.

Suppose $n \geqslant 2$ and the statement holds for $n - 1$. Let λ be an eigenvalue and v_n a corresponding normal (wlog by normalizing it) eigenvector for T^*, ie $T^*(v_n) = \lambda v_n$. Let $W := \operatorname{Span}(\{v_n\})$. Then, $W^\perp$ is T-invariant: indeed, if $v \perp W$, then $v \perp v_n$ ie $\langle v, v_n \rangle = 0$, then $\langle Tv, v_n \rangle = \langle v, T^*(v_n) \rangle = \langle v, \lambda v_n \rangle = \overline{\lambda} \langle v, v_n \rangle = 0$ so $Tv \perp W$.

Now, $\dim(W^\perp) = n - \dim(W) = n - 1$ and $T_{W^\perp} : W^\perp \to W^\perp$, so by induction applied to $T_{W^\perp}$, there is an orthonormal basis $\alpha := \{v_1, \ldots, v_{n-1}\}$ of $W^\perp$ such that $[T_{W^\perp}]_\alpha$ is upper triangular. Then, $\beta := \alpha \cup \{v_n\} = \{v_1, \ldots, v_{n-1}, v_n\}$ is an orthonormal basis for V, and

$$[T]_\beta = \begin{pmatrix} | & & | & | \\ [T(v_1)]_\beta & \cdots & [T(v_{n-1})]_\beta & [T(v_n)]_\beta \\ | & & | & | \end{pmatrix} = \begin{pmatrix} | & & | & | \\ [T_{W^\perp}(v_1)]_\alpha & \cdots & [T_{W^\perp}(v_{n-1})]_\alpha & [T(v_n)]_\beta \\ | & & | & | \\ 0 & & 0 & | \end{pmatrix}$$

$$\text{(by induction assumption)} \qquad = \begin{pmatrix} \star & \star & \star & \cdots & \star \\ 0 & \star & \ddots & \cdots & \star \\ 0 & 0 & \ddots & \ddots & \star \\ 0 & 0 & \ddots & \star & \star \\ 0 & 0 & \cdots & 0 & \star \end{pmatrix},$$

which is upper triangular. ■

Remark 5.7. *If T, T^* had precisely the same eigenvectors, then using precisely the same proof, we could get that $[T]_\beta$ diagonal, since then $Tv_n = \overline{\lambda} v_n$. This would happen, for instance, if $T = T^*$, but this condition can be relaxed:*

↪ **Definition 5.9: Normality**

$T : V \to V$ is called

- *normal* if T and T^* commute, ie $T \circ T^* = T^* \circ T$;
- *self-adjoint* if $T = T^*$.

⊛ **Example 5.6**

(a) Orthogonal projections are self-adjoint.

Let $W \subseteq V$ a subspace and P the orthogonal projection onto W. Fix $u, v \in V$. Then $u = P(u) + u', v = P(v) + v', u', v' \in W^\perp$. Then

$$\langle Pu, v\rangle = \langle Pu, Pu + v'\rangle = \langle Pu, Pv\rangle + \underbrace{\langle Pu, v'\rangle}_{=0} = \langle Pu, Pv\rangle,$$

and similarly,

$$\langle u, Pv\rangle = \langle Pu + u', Pv\rangle = \langle Pu, Pv\rangle + \langle u', Pv\rangle = \langle Pu, Pv\rangle,$$

hence $\langle Pu, v\rangle = \langle u, Pv\rangle$.

(b) If $P : V \to V$ an orthogonal projection and $\lambda \in \mathbb{C} \backslash \mathbb{R}$ then $(\lambda P)^* = \overline{\lambda} P \neq \lambda P$ so λP not self-adjoint, but it is still normal;

$$(\lambda P)(\lambda P)^* = (\lambda P)(\overline{\lambda} P) = (\lambda^2)(P^2) = (\overline{\lambda} P)(\lambda P) = (\lambda P)^*(\lambda P).$$

(c) Let $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$, where $W_i | W_j, i \neq j$. Then for any $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$, the operator $T := \lambda_1 \operatorname{proj}_{W_1} + \cdots + \lambda_k \operatorname{proj}_{W_k}$ is normal.

↪ **Proposition 5.13: Properties of Normal Operators**

Let $T : V \to V$ be a normal linear operator on V finite dimensional.

(a) $||Tv|| = ||T^* v||$ for all $v \in V$.

(b) $T - aI_V$ (or more generally $p(T)$ for any polynomial $p(t)$, ie the powers of T are normal) is normal.

(c) For all $v \in V$, v an eigenvector of T corresponding to eigenvalue $\lambda \iff v$ an eigenvector of T^* corresponding to $\overline{\lambda}$.

(d) For distinct eigenvectors $\lambda_1 \neq \lambda_2$, $\operatorname{Eig}_T(\lambda_1) \perp \operatorname{Eig}_T(\lambda_2)$.

Proof. $\overset{!}{=}$ indicates use of the normality assumption.

(a) $||Tv||^2 = \langle Tv, Tv \rangle = \langle v, T^*Tv \rangle \overset{!}{=} \langle v, TT^*v \rangle = \langle v, T^{**}T^*v \rangle = \langle T^*v, T^*v \rangle = ||T^*v||^2$.

(b) $(T - aI_V)(T^* - \overline{a}I_V) = TT^* - aT^* - \overline{a}T - a\overline{a}I_V \overset{!}{=} T^*T - aT^* - \overline{a}T - a\overline{a}I_V = (T^* - \overline{a}I_V)(T - \overline{aI_V})$. Similar proof follows for general polynomials.

(c) v an eigenvector of T corresponding to $\lambda \iff (T - \lambda I_V)(v) = 0 \iff ||(T - \lambda I_V)(v)|| = 0 \overset{!\text{ by (a)}}{\iff} ||(T^* - \overline{\lambda}I_V)(v)|| = 0 \iff v$ an eigenvector of T^* corresponding to $\overline{\lambda}$.

(d) Let $v_1 \in \mathrm{Eig}_T(\lambda_1)$, $v_2 \in \mathrm{Eig}_T(\lambda_2)$. Then $\lambda_1\langle v_1, v_2 \rangle = \langle \lambda_1 v_1, v_2 \rangle = \langle Tv_1, v_2 \rangle \overset{!}{=} \langle v_1, T^*v_2 \rangle = \langle v_1, \overline{\lambda_2}v_2 \rangle = \lambda_2\langle v_1, v_2 \rangle$ so $(\lambda_1 - \lambda_2)(\langle v_1, v_2 \rangle) = 0$, but λ_1, λ_2 assumed distinct hence $\langle v_1, v_2 \rangle = 0$ and $v_1 \perp v_2$.

■

↪ *Lecture 36; Last Updated: Wed Apr 10 14:27:27 EDT 2024*