MATH 251 CLASS NOTES

McGILL UNIVERSITY

NICHOLAS HAYEK

Based on lectures by Prof. Anush Tserunyan

CONTENTS

I Vector Spaces

AXIOMS

In previous calculus courses, we've considered the set of tuples $\langle a_1, ..., a_n \rangle$, where, in particular, $a_i \in \mathbb{R}$. These are examples of *vector spaces*, the construction which we will primarily study in this course.

Define a *vector space* V over the field $\mathbb{F}$ to be an abelian group under the operation $+$ and an identity element $\mathbb{0}_V$, which one calls the *zero vector*. Members of V are called *vectors*. Finally, V is equipped with scalar multiplication by members of $\mathbb{F}$, and satisfy the following axioms:

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

Recall also the properties of abelian groups from MATH 235, which apply to V:

$$u(vw) = (uv)w \qquad v + \mathbb{0}_V = v \qquad \exists(-v) \text{ s.t. } v + (-v) = \mathbb{0}_V \qquad uv = vu$$

for all $u, v, w \in V$.

PROPOSITION 1.1

Some formal consequences of the vector space axioms:

$$\mathbb{0}_{\mathbb{F}} v = \mathbb{0}_V \text{ for all } v \in V \qquad -\mathbb{1}_{\mathbb{F}} v = -v \qquad \alpha \mathbb{0}_V = \mathbb{0}_V$$

PROOFS.

(1): $\mathbb{0}_{\mathbb{F}} v = (\mathbb{0}_{\mathbb{F}} + \mathbb{0}_{\mathbb{F}})v = \mathbb{0}_{\mathbb{F}} v + \mathbb{0}_{\mathbb{F}} v \implies \mathbb{0}_V = \mathbb{0}_{\mathbb{F}} v$

(2): $-\mathbb{1}_{\mathbb{F}} v + v = (-\mathbb{1}_{\mathbb{F}} + \mathbb{1}_{\mathbb{F}})v = \mathbb{0}_{\mathbb{F}} v = \mathbb{0}_V$

(3): $\alpha \mathbb{0}_V = \alpha(\mathbb{0}_V + \mathbb{0}_V) \implies \alpha \mathbb{0}_V = \mathbb{0}_V$ □

Examples: Most of the pedagogical examples of vector spaces we'll see do not bear much resemblance to the $\mathbb{R}^n$, $\langle x, y, z \rangle$-like form we are familiar with:

1. The set of real, continuous functions, denoted $C[\mathbb{R}] := \{f : \mathbb{R} \to \mathbb{R}\}$, is a vector space over $\mathbb{R}$.

2. $\mathbb{F}[t]$, the set of polynomials with coefficients in $\mathbb{F}$, where addition and scalar multiplication are defined as usual, is a vector space over $\mathbb{F}$.

FURTHER CONSTRUCTIONS

Define a *product*, sometimes called the *direct sum*, of two vector spaces U, V over the same field $\mathbb{F}$ to be the Cartesian product $U \times V$ equipped with the following:

$$(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2) \quad \text{and} \quad \lambda(u, v) = (\lambda u, \lambda v)$$

$\forall u_1, u_2 \in U, v_1, v_2 \in V, \lambda \in \mathbb{F}$. One notates this as $U \oplus V$. This is itself a vector space. Note that the coordinate-vise addition and scalar multiplication are defined as in the original vector spaces.

A good exercise to prove.

For example, consider $\mathbb{F}^2$ over the field $\mathbb{F}$. One can conceptualize $\mathbb{F}$ as a vector space over $\mathbb{F}$, and thus the direct product of $\mathbb{F}$ with itself is a vector space.

Subspaces

We have constructed from a vector space one larger than it. Here is one smaller: define a *subspace* to be a set $W \subseteq V$ satisfying the following conditions

$$\mathbb{0}_V \in W \qquad u + v \in W \ \forall u, v \in W \qquad \alpha u \in W \ \forall u \in W, \alpha \in \mathbb{F}$$

If W were non-empty, then choose $u \in W$. Then $\mathbb{0}_{\mathbb{F}} u \in W$, so $\mathbb{0}_V \in W$ as required.

There a few equivalent characterizations of subspaces: $W \subseteq V$ is a vector space; or, $W \subseteq V$ is non-empty and satisfies the latter two conditions from above.

Examples: Consider $\mathbb{F}^n$ over the field $\mathbb{F}$. This is a vector space. The following are subspaces of $\mathbb{F}^n$:

1. $\{(0, x_2, ..., x_n) \in \mathbb{F}^n \ : \ x_i \in \mathbb{F}\}$.
2. $W = \{(x_1, ..., x_n) \in \mathbb{F}^n \ : \ x_1 + 2x_2 = 0)\}$. One can choose $x_3, ..., x_n$ all 0, and since $x_1 = x_2 = 0$ satisfy $x_1 + 2x_2 = 0$, one sees that $\mathbb{0}_V \in W$. If $x_1 + 2x_2 = 0$, then $\lambda x_1 + 2\lambda x_2 = 0$ as well, so W is closed under scalar multiplication. Lastly, if $x_1 + 2x_2 = 0$ and $x_1' + 2x_2' = 0$, then $(x_1 + x_1') + 2(x_2 + x_2') = 0$, so W is closed under addition.
3. *Generally*, though it is not a fact we can prove now, $W \subseteq \mathbb{F}^n :=$

$$\left\{ (x_1, ..., x_n) \in \mathbb{F}^n \text{ s.t. } \begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = 0 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = 0 \end{cases} \right\}$$

   i.e. a subset of $\mathbb{F}^n$ where a system of at least one linear equation is homogeneous. As a counter-example to this construction, see that $\{(x_1, ..., x_n) \in \mathbb{F}^n \ : \ x_1 + x_2 = 1\}$ is not a subspace of $\mathbb{F}^n$ (it violates all 3 conditions).

4. Let $\mathbb{F}[t]_n$ denote the space of polynomials whose degree is at most $n \in \mathbb{N}$. Then $\mathbb{F}[t]_n \subseteq \mathbb{F}[t]$ is a subspace. However, the set of polynomials whose degree is *exactly* n, for some positive $n \in \mathbb{N}$, is *not* a subspace. The following are further subspaces of $\mathbb{F}[t]_n$, where $p''(t)$ is defined as usual (notice the similarity to 3.).
   (a) $\{p(t) \in \mathbb{F}[t]_n : p(1) = 0\}$ or even $\{p(t) \in \mathbb{F}[t]_n : p(\alpha) = 0 \text{ with } \alpha \in \mathbb{F}\}$
   (b) $\{p(t) \in \mathbb{F}[t]_n : p''(t) + 2p'(t) - p(t) = 0\}$
5. For $C[\mathbb{R}]$, which, as noted above, is a vector space, the following are subspaces:
   (a) $\{f \in C[\mathbb{R}] : f(\pi) + 7f(\sqrt{2}) = 0\}$
   (b) $\{f \in C[\mathbb{R}] \text{ differentiable everywhere}\}$
   (c) $\left\{ f \in C[\mathbb{R}] : \int_0^1 f\,dx = 0 \right\}$. The proof of this follows from linearity of the integral (MATH 255). In truth, the bounds for the integral can be arbitrary, though see this integral cannot be set arbitrarily.

PROPOSITION 1.2

If W_1, W_2 are subspaces of some common v.s. over the field $\mathbb{F}$, then

$$W_1 + W_2 := \{w_1 + w_2 : w_1 \in W_1, w_2 \in W_2\} \quad \textbf{and} \quad W_1 \cap W_2$$

are both subspaces. The proofs for these are left to the reader.

LINEAR COMBINATIONS

Define a *linear combination* of vectors $v_1, ..., v_n \in V$, where V is a vector space over $\mathbb{F}$, to be $\sum_{i=1}^{n} a_i v_i$, where $a_i \in \mathbb{F}$. So long as one a_i is non-zero, one calls this a *non-trivial* LC. Otherwise (i.e. all $a_i = 0$), we have a *trivial* LC.

When we deal with a possibly infinite set of vectors, $S \subseteq V$, we will only take *finite* linear combinations, for a subset $\{v_1, ..., v_n\} \subseteq S$. Never will we compute infinite sums in this course.

Define the *span* of $S \subseteq V$ to be the set of all possible linear combinations of S, $\{a_1 v_1 + ... + a_n v_n : a_i \in \mathbb{F}, v_i \in S\}$. By convention, we say that $\text{Span}(\varnothing) = \mathbb{0}_V$.

Example: Let $S := \{(1, 0, -1), (0, 1, -1), (1, 1, -2)\} \subseteq \mathbb{R}^3$. Then $\mathbb{0}_{\mathbb{R}^3} = (0, 0, 0) = 0(1, 0, -1) + 0(0, 1, -1) + 0(1, 1, -2)$ is a trivial linear combination. However, we can get to 0 non-trivially: $(1, 0, -1) + (0, 1, -1) - (1, 1, -2) = 0$.

What about $\text{Span}(S)$? This is the set $\{a(1, 0, -1) + b(0, 1, -1) + c(1, 1, -2)\} = \{(a + c, b + c, -a - b - 2c)\}$. Clearly this is a subset of $\{(a, b, c) : a + b + c = 0\}$, since, indeed,

$a+c+b+c-a-b-2c=0$. The converse is also true: suppose (x,y,z) is such that $x+y+z=0$. Then $z=-x-y$, and one writes $(x,y,-x-y)=x(1,0,-1)+y(0,1,-1)$. It follows that $\text{Span}(S)=\{(x,y,z):x+y+z=0\}$.

Proposition 1.3 Let V be a v.s. over a field $\mathbb{F}$, and S be some subspace of it. Then $\text{Span}(S)$ is a subspace of V containing S, and furthermore is the smallest such subspace containing S.

Proof.

Adding and scalar multiplying a linear combination of vectors produces a further linear combination, so $\text{Span}(S)$ is closed under these operations. Furthermore, $\mathbb{0}_V \in \text{Span}(S)$ by taking a trivial combination of vectors $\implies$ $\text{Span}(S)$ is a subspace.

If $U \supseteq S$ is a subspace, then U is closed under addition and scalar multiplication, so it contains all linear combinations of S, i.e. $U \subseteq \text{Span}(S)$ □

Proposition 1.4 For $S \subseteq V$, $v \in V$, we have that $v \in \text{Span}(S) \iff \text{Span}(S \cup \{v\}) = \text{Span}(S)$.

Proof.

($\implies$) If $v \in \text{Span}(S)$, then v is some linear combination of vectors in S, so $v = a_1v_1+...+a_nv_n$. Let $u \in \text{Span}(S\cup\{v\})$. Then $u = a_1'v_1'+...+a_m'v_m'+av$, where a may be 0, and $v_i' \in S$. One rewrites $u = a_1'v_1' + ... + a_m'v_m' + a(a_1v_1 + ... + a_nv_n)$ from above. Thus, $\text{Span}(S \cup \{v\}) \subseteq \text{Span}(S)$. Trivially, $\text{Span}(S) \subseteq \text{Span}(S \cup \{v\})$, so $\text{Span}(S) = \text{Span}(S \cup \{v\})$.
($\impliedby$) Assume $\text{Span}(S) = \text{Span}(S \cup \{v\})$. Clearly, $v \in \text{Span}(S \cup \{v\})$, so $v \in \text{Span}(S)$ as well. □

For a v.s. over a field $\mathbb{F}$, call $S \subseteq V$ a *spanning set of* V if $\text{Span}(S) = V$. One calls a spanning set *minimal* if no proper subset of S is spanning, i.e. $\text{Span}(S \setminus v) \neq V$ for all $v \in V$.

Example: For $S := \{(1,0,-1),(0,1,-1),(1,1,-2)\}$, we have from Proposition 1.4 that $\text{Span}(S) = \text{Span}(\{(1,0,-1),(0,1,-1)\})$, as $(1,1,-2) \in \text{Span}(\{(1,0,-1),(0,1,-1)\})$.

Thus, it follows that S is not a minimal spanning set over itself.

For the v.s. $\mathbb{F}^n$ over $\mathbb{F}$, define the *standard spanning set*:

$$\text{St}_n := \{(1, \underbrace{0,0,...,0}_{n-1 \text{ times}}), (0,1,0,...,0), ..., (0,...,0,0,1)\}$$

This is indeed spanning for $\mathbb{F}^n$, and is minimal.

LINEAR (IN)DEPENDENCE

Let V be a v.s. and $S \subseteq V$ a subspace. S is called *linearly dependent* if there exists a non-trivial linear combination equal to $\mathbb{0}_V$. Otherwise S is called *linearly independent*.

Examples:

1. The empty set, by vacuous implication, is linearly independent.
2. For $v \in V$, v is linearly dependent $\iff v = \mathbb{0}_V$
3. $S := \{(1,0,-1),(0,1,-1),(1,1,-2)\}$ is linearly dependent
4. $S \subseteq \mathbb{F}^3 = \{(1,0,-1),(0,1,-1),(0,0,,1)\}$ is linearly dependent. We argue by contradiction: let $(0,0,0) = a(1,0,-1)+b(0,1,-1)+c(0,0,1) = (a,b,c-a-b)$. Then $a = b = 0$ by necessity, and it follows that $c - a - b = c = 0$. Thus, only a trivial linear combination equals the zero vector.
5. $\mathrm{St_n} \subseteq \mathbb{F}^n$ is linearly independent

---

Let V be a v.s. over $\mathbb{F}$, $S \subseteq V$ (possibly infinite). Then: PROPOSITION 1.5

(a) S is linearly dependent $\iff$ there exists a finite $S_0 \subseteq S$ which is linearly dependent

(b) S is linearly independent $\iff$ all finite $S_0 \subseteq S$ are linearly independent

PROOF.

Note that (b) is simply the negation of (a), so only (a) requires a proof.

($\implies$) Suppose S is linearly dependent. Then $a_1v_1 + ... + a_nv_n = \mathbb{0}_V$, where, WLOG, we assume that $a_i \neq \mathbb{0}_{\mathbb{F}}$. The set $\{v_1, ..., v_n\} \subseteq S$ is clearly linearly dependent.

($\impliedby$) If $S_0 \subseteq S$ is linearly dependent, then clearly S is too □

For $S \subseteq V$ over $\mathbb{F}$, we have PROPOSITION 1.6

(a) S is linearly dependent $\iff$ there exists $v \in S$ with $v \in \mathrm{Span}\{S \setminus v\}$

(b) S is linearly independent $\iff$ for all $v \in S$, $v \notin \mathrm{Span}\{S \setminus v\}$

PROOF.

Once again, only (a) requires proof.

($\implies$) Let S be linearly dependent. Then $a_1v_1 + ... + a_nv_n = \mathbb{0}_V$, and WLOG we assume all a_i are non-zero. Since $\mathbb{F}$ is a field, we may write $v_1 = -a_1^{-1}a_2v_2 - ... - a_1^{-1}a_nv_n$. Thus, $v_1 \in \mathrm{Span}\{S \setminus v_1\}$, and we are done.

($\Longleftarrow$) Suppose $v \in S$ is such that $v \in \text{Span}(S \setminus v)$. Then $v = a_1 v_1 + ... + a_n v_n$, where $v_i \in S \setminus v$. It follows that $\mathbb{0}_V = a_1 v_1 + ... + a_n v_n - v$. As $-1 \neq 0$, this is non-trivial, and we are done. □

Corollary

$S \subseteq V$ is linearly independent $\iff$ S is a minimal spanning set for $\text{Span}(S)$

Clearly, $\text{Span}(S) = \text{Span}(S)$. However, $v \in S \implies v \notin \text{Span}(S \setminus v)$. We know $v \in \text{Span}(S)$, so $\text{Span}(S) \neq \text{Span}(S \setminus v)$.

For a vector space V over $\mathbb{F}$, $S \subseteq V$ is called *maximally independent* if S is linearly independent AND there does not exist $v \in V \setminus S$ s.t. $S \cup \{v\}$ is linearly independent. In other words, S is independent, and adding *any* new vectors will break this independence.

Proposition 1.7

If S is maximally independent, then S is spanning for V.

Proof.

Let S be maximally independent. Then for any $v \in V \setminus S$, the set $S \cup \{v\}$ is linearly dependent, i.e. $av + a_1 v_1 + ... + a_n v_n = \mathbb{0}_V$ for all non-zero a_i. In particular, $a \neq 0$, or else we would yield a non-trivial linear combination for only vectors in S, which violates our independence condition.

Thus, write $v = -a^{-1} a_1 v_1 - ... - a^{-1} a_n v_n$, and conclude that $v \in \text{Span}(S)$. Then $V \subseteq \text{Span}(S)$. Clearly, $\text{Span}(S) \subseteq V$, so we conclude that $\text{Span}(S) = V$. □

BASES

1.1 Characterization of a Basis

Let V be a v.s. over $\mathbb{F}$ and $S \subseteq V$. The following are then equivalent:

1. S is a minimal spanning set for V
2. S is linearly independent and spanning for V
3. S is maximally independent
4. Every $v \in V$ is equal to a *unique* combination of vectors in S

Proofs.

(1) $\implies$ (2) Let $S \subseteq V$ be a minimal spanning set for V. Then, especially, S is a minimal spanning set for $\text{Span}(S)$, and by the corollary above, S is linearly independent.

(3) $\implies$ (1) Let $S \subseteq V$ be maximally independent. By proposition 1.7, S is spanning for V. By the corollary, S is also minimally spanning for $\text{Span}(S)$. Combining, we see that S is minimally spanning for V.

(2) $\implies$ (4) Let $S \subseteq V$ be linearly independent and spanning for V. Then, clearly, $l \in V \in \text{Span}(S)$ means that it can be written as a linear combination

of vectors in S. We need this combination to be unique: let $a_1v_1 + ... + a_nv_n = l$ and $b_1v_1 + ... + b_nv_n = l$, where $S = \{v_i\}_{1 \leq i \leq n}$ One uses the same vectors, noting that some coefficients may be 0, as needed.

$a_1v_1 + ... + a_nv_n = b_1v_1 + ... + b_nv_n \implies a_1v_1 + ... + a_nv_n - b_1v_1 - ... - b_mv_m = 0$. We can thus combine $a_i - b_i = c_i$, and write $c_1v_1 + ... + c_nv_n = 0$. Since S is linearly independent, we require that all $c_i = 0$, i.e. $a_i = b_i \forall i$.

(4) $\implies$ (2) This result is immediate, as $V \subseteq \text{Span}(S)$, $\text{Span}(S) \subseteq V \implies \text{Span}(S) = V$. Since all vectors in v have a *unique* representation, consider $v = \mathbb{0}_V$. A trivial combination produces the zero vector, and by uniqueness this must be the *only* such combination, and we conclude that S is linearly independent. □

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

With respect to (4), the unique combination is called a *unique representation of v in S*. The associated coefficients are called the *Fourier coefficients of v in S*

Examples:

1. Consider St_n, the standard basis for $\mathbb{F}^n$ (notice the terminology). This is, of course, a basis
2. $\mathbb{F}[t]_n$, the space of polynomials with degree at most n, has a basis $\{1, t, t^2, ..., t^n\}$.
3. In $\mathbb{F}^3$, $\{1, 0, -1), (0, 1, -2), (0, 0, 1)\}$ is a basis.
4. The standard basis of $\mathbb{F}[t]$, the space of *all* polynomials, is $\{1, t, t^2, ...\} = \{t^n : n \in \mathbb{N}\}$. One checks linear independence of this space by considering all finite subsets (remember, we do not take infinite sums). Note: in this couse, $0 \in \mathbb{N}$
5. Define $\mathbb{F}[[t]]$ to be the set of all power series, i.e. $\left\{\sum_{n \in \mathbb{F}} a_n t^n : a_n \in \mathbb{F}\right\}$. In the bullet above, we consider the space of polynomials, i.e. formal power series with *finitely* many non-zero terms. Not so for $\mathbb{F}[[t]]$, in generality. We ask: does this have a basis?

1.2 Every vector space has a basis

PROOF ATTEMPT.

Since V may be infinite, we will have to rely on some non-rigorous notions to prove this in any short form. Suppose V is a vector space over $\mathbb{F}$, and let $S_0 = \varnothing$ be a trivial, independent subspace. If S_0 is maximally independent, then we are done. Otherwise, there exists $v_1 \in V$ such that $S_1 := S_0 \cup \{v_1\}$ is also independent. If S_1 is maximal, then we are done. Otherwise, choose

$v_2 \in V$, define $S_2 := S_1 \cup \{v_2\}$, and so on and so on. This last notion ("and so on and so on") is problematic when V is not finite. To resolve this, we'll need to learn and understand *Zorn's Lemma* □

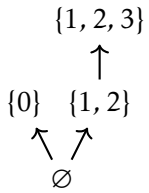
Zorn's Lemma

The definition of ambient sets is not necessary to understand Zorn's lemma, but you can read about it here

Let X be some ambient set and $\mathcal{I}$ be a collection of subsets of X. In other words, $\mathcal{I} \subseteq \mathcal{P}(X)$, the powerset of X. Call a set $S \in \mathcal{I}$ an *inclusion-maximal* element if $\nexists$ any strict superset $S' \supsetneq S$ such that $S' \in \mathcal{I}$. Call a collection of sets $C \subseteq \mathcal{P}(X)$ a *chain* if, for any two sets $A, B \in C$, one has $A \subseteq B$ or $B \subseteq A$.

To demonstrate these definitions, let $X := \mathbb{N}$ and $\mathcal{I} := \{\varnothing, \{0\}, \{1,2\}, \{1,2,3\}\} \subseteq \mathcal{P}$.

Both $\{0\}$ and $\{1,2,3\}$ are inclusion-maximal in $\mathcal{I}$: adding any element to either of these sets will land you outside of $\mathcal{I}$. $C_1 = \{\varnothing, \{1,2\}, \{1,2,3\}\}$ is a chain, but $C_2 = \{\varnothing, \{1,2\}, \{0\}\}$ is *not* a chain.



where ↗ means ⊆

Lastly, define an *upper bound* of $\mathcal{J} \subseteq \mathcal{P}(X)$ to be a set $U \subseteq X$ such that $U \supseteq J$ for all sets $J \in \mathcal{J}$.

1.3 Zorn's Lemma
Let X be a set, $\mathcal{I} \subseteq \mathcal{P}(X)$ non-empty. If every chain $C \subseteq \mathcal{I}$ has an upper bound in $\mathcal{I}$, then $\mathcal{I}$ has a maximal element.

The proof for this is statement beyond this course (see MATH 488).

Let's revisit the statement that every vector space has a basis, now equipped with Zorn's lemma:

Proof.

Let $\mathcal{I}$ be the collection of linearly independent subspaces in V. This is non-empty, since at least the empty set is linearly independent. If one can show that $\mathcal{I}$ has a maximal element, in the sense of Zorn's lemma, then this element is also maximally independent.

Consider a chain $C \subseteq \mathcal{I}$, and let $S := \cup C$ be the union of all sets in C. This is clearly an upper bound of C, so we want to show that it is linearly independent. However, S may be infinite, so consider an arbitrary subset $\{v_1, ..., v_n\} \subseteq S$.

Let $S_i \in C$ be some set that contains v_i, from the set described above. Since C is a chain, for any i, j, we have $S_i \subseteq S_j$ or $S_j \subseteq S_i$, so WLOG we can order these sets as follows:

$$S_1 \subseteq S_2 \subseteq ... \subseteq S_n$$

Thus, $v_1, ..., v_n \in S_n$, and since $S_n \in C \subseteq \mathcal{I}$ (recall the definition of $\mathcal{I}$), S_n is linearly independent. Thus, $\{v_1, ..., v_n\} \subseteq S_n$ is linearly independent $\implies$ S is linearly independent $\implies$ $S \in \mathcal{I}$ is an upper-bound of $\mathcal{I}$.

Zorn's lemma is satisfied, so $\mathcal{I}$ has a maximal element, and we are done. □

Steinitz Substitution Lemma

1.4 Cardinality of Bases
For a vector space V over $\mathbb{F}$, any two bases have the same cardinality.

We'll require another lemma to prove this statement:

1.5 Steinitz Substitution Lemma
Let V be a vector space over $\mathbb{F}$. Let $Y \subseteq V$ be a finite, linearly independent set and $Z \subseteq V$ be a finite spanning set. Then the following hold:

(a) $|Y| \leq |Z|$

(b) $\exists\ Z' \subseteq Z$ such that $Y \cup Z'$ still spans V, where $|Z'| = |Z| - |Y|$

Proof TBD

Now we'll show theorem 1.4:

PROOF.

Let Y and Z be two finite bases for V. Then Y is linearly independent and Z is spanning. Thus, $|Y| \leq |Z|$ be Steinitz. However, Y is also spanning, and Z is linearly independent, so $|Z| \leq |Y| \implies |Y| = |Z|$. □

For a vector space V over $\mathbb{F}$, define the *dimension* of V, denoted by $\dim(V)$, to be the cardinality of its (i.e. any) basis. We call V a *finite dimensional vector space* if $\dim(V)$ is a natural number, otherwise its *infinite dimensional*.

PROPOSITION 1.8

Let V have $\dim(V) = n$. Then the following hold by Steinitz:

(a) For every linearly independent set $I \subseteq V$, $|I| \leq n$. If $|I| = n$, then I is a basis.

(b) For every spanning set $S \subseteq V$, $|S| \geq n$. If $|S| = n$, then S is a basis.

(c) Every linearly independent set I can be completed to a basis for V, i.e. $\exists$ a basis B for V which contains I

PROOFS.

(a) Since a basis B is spanning, one has $|I| \leq |B| = n$

(b) Since a basis B is independent, $|B| \leq |S|$, i.e. $|S| \geq n$

(c) Let I be independent and B be a basis. Then $\exists\ B' \subseteq B$ with $I \cup B'$ spanning. $I \cup B'$ is also independent: we know that $|I \cup B'| \geq n$. However, $|I \cup B'| \leq |I| + |B'| = |B| = n$. Thus, $|I \cup B'| = n$. It follows that this set is minimally spanning, since $|I \cup B'| = n - 1$ is a contradiction of (b). $\implies |I + B'|$ is a basis. □

1.6 Monotonicity of Dimension
Let V be a finite dimensional vector space. Then for any subspace $W \subseteq V$, $\dim(W) \leq \dim(V)$ and $\dim(W) = \dim(V) \iff W = V$.

Proof.

Let B be a basis for W. Since B is independent and $W \subseteq V$, $|B| \leq \dim(V)$ by proposition 1.8, so $\dim(W) \leq \dim(V)$.

($\implies$) If $|B| = \dim(V)$, then B is a basis for V by 1.8, so $\text{Span}(B) = V$, or $W = V$. The ($\impliedby$) direction is trivial. □

II Linear Transformations

AXIOMS AND INITIAL PROPERTIES

Let V, W be vector spaces over $\mathbb{F}$. One calls a mapping $T : V \to W$ a *linear transformation* if it preserves vector space structure, i.e.

1. $T(v_1 + v_2) = T(v_1) + T(v_2)\ \forall v_1, v_2 \in V$
2. $T(\alpha v) = \alpha T(v)\ \forall v \in V, \alpha \in \mathbb{F}$

Immediately, we have that $T(\mathbb{0}_V) = \mathbb{0}_W$ and $T(-v) = -T(v)$.

Examples:

1. Consider $T : \mathbb{F}^2 \to \mathbb{F}^2 : T(a_1, a_2) = (a_1 + 2a_2, a_1)$. This is a linear transformation. Checking the axioms: $T(a_1 + b_1, a_2 + b_2) = (a_1 + b_1 + 2(a_2 + b_2), a_1 + b_1) = (a_1 + 2a_2 + b_1 + 2b_2, a_1 + b_1) = T(a_1, a_2) + T(b_1, b_2)$. Also, $T(\alpha a_1, \alpha a_2) = (\alpha a_1 + 2\alpha a_2, \alpha a_1) = \alpha(a_1 + 2a_2, a_1) = \alpha T(a_1, a_2)$.
2. Let θ be an angle, and $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the rotation of a vector by θ. This is a linear transformation.
3. $T : \mathbb{R}^2 \to \mathbb{R}^2$, the reflection transformation defined by $T(a_1, a_2) = T(a_1, -a_2)$
4. The transpose $M_n(\mathbb{F}) \to M_n(\mathbb{F}) : A \to A^T$
5. $\mathcal{D}$, the derivative of finite polynomials.

2.1 Linear transformations are completely determined by values on a basis

Let $B := v_1, ..., v_n$ be a basis for a vector space V. Let W be a vector space over a common field $\mathbb{F}$, and $w_1, ..., w_n \in W$. *Then* there exists a unique linear transformation $T : V \to W$ which sends $T(v_i) = w_i\ \forall i \in [1, n]$.

PROOF.

Existence: Let $v \in V$, $B \subseteq V$ a basis for V, and consider some transformation $T(v)$. We write $v = a_1 v_1 + ...a_n v_n, v_i \in B$, the unique representation of v in B. Now, define $T(v) = a_1 w_1 + ... + a_n w_n$ for fixed $w_i \in W$. This will indeed send $T(v_i) = w_i$ as desired. To show that T is linear, one checks the axioms:

For $u, v \in V$, let $v = a_1 v_1 + ... + a_n v_n$ and $u = b_1 v_1 + ... + b_n v_n$ be the unique representations of u, v in B. Then $u + v = (a_1 + b_1)v_1 + ... + (a_n + b_n)v_n$, so $T(u+v) = (a_1+b_1)w_1 + ... + (a_n+b_n)w_n = a_1 w_1 + ... + a_n w_n + b_1 w_1 + ... + b_n w_n = T(u) + T(v)$.

$T(\alpha v) = \alpha T(v)$ follows immediately from its definition.

Uniqueness: Suppose T_1, T_2 are both such that $T_1(v_i) = w_i = T_2(v_i)$ for all i. One shows that $T_1(v) = T_2(v)\ \forall v \in V$. Let $v = a_1v_1 + ... + a_nv_n$ be the unique representation of v in B. By linearity, $T(v) = a_1T(v_1) + ... + a_nT(v_n) = a_1w_1 + ... + a_nw_n$ for both T_1 and T_2. Since a_i and w_i are all fixed, we see that $T_1(v) = T_2(v)$. □

2.2 Extension of Functions on Basis

Let V, W be vector spaces, possibly infinite, over $\mathbb{F}$, and let β be a basis for V. Every function $T : \beta \to W$ can be extended to a unique linear transformation $\hat{T} : V \to W$.

Proof.

This is essentially the infinite case of theorem 2.1:

Existence: Let $T : \beta \to \gamma$ be (any) function between the bases of V and W. For $v \in V$, let $v = a_1v_1 + ... + a_nv_n$ be its unique representation in β, where $v_i \in \beta$. Define the function

$\hat{T}$ is indeed an extension of T. See that $\hat{T}(v_i) = T(v_i)$, since v_i is its own representation in β.

$$\hat{T}(v) = a_1T(v_1) + ... + a_nT(v_n)$$

We'll show that this is linear. Let $x, y \in V$. Without loss of generality, we can write $x = a_1v_1 + ... + a_mv_m$ and $y = b_1v_1 + ... + b_mv_m$ as their unique representations, where a_i, b_i *may* be zero. We thus have

$$\begin{aligned}\hat{T}(x+y) &= (a_1+b_1)T(v_1) + ... + (a_m+b_m)T(v_m)\\ &= a_1T(v_1) + b_1T(v_1) + ... + a_mT(v_m) + b_mT(v_m)\\ &= \hat{T}(x) + \hat{T}(y)\\ \hat{T}(\alpha x) &= \alpha a_1T(v_1) + ... + \alpha a_mT(v_m)\\ &= \alpha[a_1T(v_1) + ... + a_mT(v_m)] = \alpha\hat{T}(x)\end{aligned}$$

Uniqueness: Let $\hat{T}$ be as defined, and let $\tilde{T} : V \to W$ be another transformation which *also* sends $\beta \to \gamma$ according to $T : \beta \to \gamma$. Fix $v \in V$, and let $a_1v_1 + ... + a_nv_n$ be its unique representation in β.

$$\hat{T}(v) = a_1T(v_1) + ... + a_nT(v_n) = a_1\tilde{T}(v_1) + ... + a_n\tilde{T}(v_n) = \tilde{T}(a_1v_1 + ... + a_nv_n) = \tilde{T}(v)$$

□

ISOMORPHISMS

Define an *isomorphism* $T : V \to W$, for two vector space V, W over $\mathbb{F}$, to be a linear transformation which admits a linear inverse.

If there exists an isomorphism between V and W, one says that V and W are *isomorphic* (to eachother). Write $V \cong W$.

$T : V \to W$ is an isomorphism $\iff$ T is linear and bijective. PROPOSITION 2.1

This may seem trivial, and the ($\implies$) direction is. However, we need to show that, for T linear and bijective, its inverse is linear: PROOF.

We know T^{-1} exists, since T is bijective. Let $w_1, w_2 \in W$ and $a_1, a_2 \in \mathbb{F}$:

$$\begin{aligned} T^{-1}(a_1 w_1 + a_2 w_2) &= T^{-1}[a_1 T(T^{-1}(w_1)) + a_2 T(T^{-1}(w_2))] \\ &= T^{-1}[T(a_1 T^{-1}(w_1)) + T(a_2 T^{-1}(w_2))] \\ &= T^{-1}[T(a_1 T^{-1}(w_1) + a_2 T^{-1}(w_2))] \\ &= a_1 T^{-1}(w_1) + a_2 T^{-1}(w_2) \quad \square \end{aligned}$$

2.3 Freeness of Vector Spaces
All bijections from $\beta \to \gamma$ can be extended to a unique isomorphism between V and W. This follows from Theorem 2.2.

2.4 Isomorphism with Same Dimension
For $n \in \mathbb{N}$, a vector space V over $\mathbb{F}$ with $\dim(V) = n$ is isomorphic to $\mathbb{F}^n$. In particular, all n-dimensional vector spaces over $\mathbb{F}$ are isomorphic to eachother.

Fix a basis $B := \{v_1, ..., v_n\}$ for V. Let $V \to \mathbb{F}^n$ be the unique transformation which sends $T(v_i) = e_i$, where $e_i = \{0, ..., 0, 1, 0, ..., 0\}$, with 1 in the i^{th} position. PROOF.

T is injective: let $T(x) = T(y)$ for $x, y \in V$, and write $x = a_1 v_1 + ... + a_n v_n$, $y = b_1 v_1 + ... + b_n v_n$, the unique representations of x, y in B.

Then $T(x) = T(y) \implies a_1 e_1 + ... + a_n e_n = b_1 e_1 + ... + b_n e_n$, since T sends $v_i \to e_i$. By uniqueness of representation in a basis, one has $a_i = b_i$.

T is surjective: let $w \in \mathbb{F}^n$. Then let $w = a_1 e_1 + ... + a_n e_n$ be its unique representation in $\mathrm{St_n}$. Then $T(a_1 v_1 + ... + a_n v_n) = w$.

Recall that $\{e_i\}_{i\leq n}$ is the standard basis, $\mathrm{St_n}$, of $\mathbb{F}^n$.

Thus, $V \cong \mathbb{F}^n$, and so all n-dim vector spaces are isomorphic to eachother. □

For a linear transformation $T : V \to W$, define its *image*, notated $\mathrm{Im}(T)$ or $T(V)$, to be the set $\{T(v) : v \in V\}$. Similarly, define its *kernel*, notated $\ker(T)$ or $T^{-1}(\mathbb{0}_W)$, to be $\{v \in V : T(v) = \mathbb{0}_W\}$.

Proposition 2.2 $\ker(T)$ and $\mathrm{Im}(T)$ are subspaces of V and W, respectively.

Proof. *For* $\ker(T)$: Let $v_1, v_2 \in \ker(T)$ and $a_1, a_2 \in \mathbb{F}$. Then $T(a_1v_1 + a_2v_2) = a_1T(v_1) + a_2T(v_2) = a_1\mathbb{0}_W + a_2\mathbb{0}_W = \mathbb{0}_W$, so $a_1v_1 + a_2v_2 \in \ker(T)$.

For $\mathrm{Im}(T)$: Let $w_1, w_2 \in \mathrm{Im}(T)$. Then $w_i = T(v_i)$ for some $v_i \in V$, so $a_1w_1 + a_2w_2 = a_1T(v_1) + a_2T(v_2) = T(a_1v_1 + a_2v_2)$, so $a_1w_1 + a_2w_2 \in \mathrm{Im}(T)$. □

Proposition 2.3 Let $T : V \to W$ be a linear transformation. Let $B \subseteq V$ be a basis for V. Then $T(B)$ spans $\mathrm{Im}(T)$. In particular, T is surjective $\iff$ $T(B)$ spans W.

Proof. Let $w \in \mathrm{Im}(T)$, so $w = T(v)$ for some $v \in V$. Write $v = a_1v_1 + ... + a_nv_n$ to be the unique representation of v in B. Then $w = T(v) = a_1T(v_1) + ... + a_nT(v_n) \in \mathrm{Span}(\{T(v_1), ..., T(v_n)\})$, so $T(B)$ spans $\mathrm{Im}(T)$

If T is surjective, then $\mathrm{Im}(T) = W$, and vice-versa. □

Proposition 2.4 Let $T : V \to W$ be a linear transformation. Then the following are equivalent:

1. T is injective
2. $\ker(T) = \{0_V\}$
3. $T(B)$ is independent for all bases $B \subseteq V$
4. $T(B)$ is independent *for some* basis $B \subseteq V$

Proof. (1) $\iff$ (2). $\implies$ direction trivial. ($\impliedby$) Let $\ker(T) = \{0_V\}$, and $T(x) = T(y)$ for some $x, y \in V$. Then $T(x) - T(y) = \mathbb{0}_W = T(x - y)$, so $x - y \in \ker(T)$. But then $\mathbb{0}_V = x - y$, so $x = y$.

(2) $\implies$ (3) Fix a basis $B := \{v_1, ..., v_n\} \subseteq V$. To show that $T(B)$ is linearly independent, take a combination $a_1w_1 + ... + a_nw_n$, where $T(v_i) = w_i$. These w_i are distinct, since T is injective by (2) $\implies$ (1).

Suppose $a_1w_1 + ... + a_nw_n = 0$. Then $T(a_1v_1 + ... + a_nv_n) = 0$, so $a_1v_1 + ... + a_nv_n \in \ker(T)$. Thus, by (2), $a_1v_1 + ... + a_nv_n = 0$, but $v_i \in B$ are linearly independent, so $a_i = 0$.

(3) $\implies$ (4) trivial.

(4) $\implies$ (2) Fix $B \subseteq V$ and let $T(B)$ be linearly independent. Suppose $T(v) = 0$, and write $v = a_1 v_1 + ... + a_n v_n$ for $v_i \in B$. Then $a_1 T(v_1) + ... + a_n T(v_n) = 0$, but $T(B)$ is linearly independent, so $a_i = 0$ □

If V and W are isomorphic, they have the same dimension. PROPOSITION 2.5

PROOF. This follows directly from propositions 2.3 and 2.4: if V and W are isomorphic, then $\exists T : V \to W$ which is bijective. Let B be a basis for V. Then T surjective $\implies$ $T(B)$ is spanning for W by 2.3. T injective $\implies$ $T(B)$ independent by 2.4. Thus, $T(B)$ is a basis for W. But T is a bijection, so $|T(B)| = |B|$, and we conclude that $\dim(V) = \dim(W)$. □

If $T : V \to W$ is an injective linear transformation, then $\dim(W) \geq \dim(V)$. This is something along the lines of a pigeonhole principle for vector spaces. PROPOSITION 2.6

PROOF. Since $\text{Im}(T) \subseteq W$, we know $\dim(\text{Im}(T)) \leq \dim(W)$. Thus, we show that $\dim(\text{Im}(T)) = \dim(V)$. But since T is injective, it is an extension of $\hat{T} : V \to \text{Im}(T)$ which *is* surjective, and thus bijective. We conclude that V and $\text{Im}(T)$ are isomorphic to eachother, so they have the same dimension. □

For vector spaces V, W over $\mathbb{F}$, define the *rank* of T, denoted $\text{rank}(T)$, to be $\dim(\text{Im}(T))$. Similarly, define the *nullity* of T, denoted $\text{null}(T)$, to be $\dim(\ker(T))$.

2.6 Rank-Nullity (or Dimension) Theorem
Let V be finite dimensional, and W any v.s. over a common field $\mathbb{F}$. If $T : V \to W$ is a linear transformation, then $\text{null}(T) + \text{rank}(T) = \dim(V)$

PROOF. This follows directly from the 1st isomorphism theorem for vector space (to be seen), along with the fact that $\dim(V/\ker(T)) = \dim(V) - \dim(\ker(T))$. A more manual proof is as follows:

Let $\{v_1, ..., v_k\}$ be a basis for $\ker(T)$. By Steinitz' Lemma, this can be completed to a basis for V, say $\{v_1, ..., v_k, u_1, ..., u_{n-k}\}$, where $n = \dim(V)$. If we show $\dim(\text{Im}(T)) = n - k$, then the theorem follows.

Recall that $T(B)$ spans $\text{Im}(T)$ for any basis $B \subseteq V$. Thus,

$$\text{Span}(\{T(v_1), ..., T(v_k), T(u_1), ..., T(u_{n-k})\}) = \text{Im}(T)$$

However, $v_i \in \ker(T)$, so $T(v_i) = 0$, and we conclude that $\{T(u_1), ..., T(u_{n-k})\}$ is spanning for $\text{Im}(T)$.

This set (of $n-k$ elements) is linearly independent as well:

$$\begin{aligned} & a_1 T(u_1) + ... + a_{n-k} T(u_{n-k}) = \mathbb{0}_W \\ \implies & a_1 u_1 + ... + a_{n-k} u_{n-k} \in \ker(T) \\ \implies & a_1 u_1 + ... + a_{n-k} u_{n-k} = b_1 v_1 + ... + b_n v_n \quad \text{(uniquely)} \\ \implies & a_1 u_1 + ... + a_{n-k} u_{n-k} - b_1 v_1 - ... - b_n v_n = \mathbb{0}_V \\ \implies & a_i = 0 \ \forall i \quad \text{by linear independence of basis of } V \quad \square \end{aligned}$$

PROPOSITION 2.7 Let V, W be n-dimensional vector spaces over $\mathbb{F}$. For a linear transformation $T : V \to W$, the following are equivalent:

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

PROOF. T surjective $\iff \text{rank}(T) = \dim(\text{Im}(T)) = \dim(W) = n$
T injective $\implies \text{null}(T) = 0$, so $\text{rank}(T) = \dim(V) = n$
$\text{rank}(T) = n \implies \text{null}(T) = 0$, so $\ker(T) = \{0_V\}$, so T injective $\square$

2.7 First Isomorphism Theorem
Let V, W be vector spaces over $\mathbb{F}$. Let $T : V \to W$ be a linear transformation. Then $V/\ker(T)$ is isomorphic to $\text{Im}(T)$ through the isomorphism $\overline{v} \to T(v)$, where $\overline{v} := v + \ker(T)$ (as in quotient groups).

PROOF. We know that $\hat{T} : V/\ker(T) \to \text{Im}(T)$ given by $\hat{T}(\overline{v}) = T(v)$ is a well-defined group isomorphism. Thus, we need to check that $\hat{T}$ is linear. In particular, we need to check scalar multiplication, since group homomorphisms obey $T(x+y) = T(x) + T(y)$.

$$\hat{T}(\alpha\overline{v}) = \hat{T}(\overline{av}) = T(av) = aT(v) = a\hat{T}(\overline{v}) \quad \square$$

THE SPACE HOM(V, W)

For vector spaces V, W over $\mathbb{F}$, define $\text{Hom}(V, W)$ to be the set of all linear transformations from $V \to W$.

PROPOSITION 2.8 $\text{Hom}(V, W)$ is a vector space over $\mathbb{F}$, equipped with the following:

Addition $T_0 + T_1$ defines the function $T_0 + T_1 : V \to W$, where $(T_0 + T_1)(v) = T_0(v) + T_1(v)$, where $T_0, T_1 \in \text{Hom}(V, W)$.

Scalar Multiplication For $T \in \text{Hom}(V, W)$ and $a \in \mathbb{F}$, $a \times T$ defines the function $(aT) : V \to W$, where $(aT)(v) = aT(v)$.

Zero Vector $\mathbb{0}_{\text{Hom}(V,W)}$ is the function which takes $v \to \mathbb{0}_W$

2.8 Basis for Hom
Let V, W be vector spaces over $\mathbb{F}$, which have bases β, γ, respectively, where β is finite. The set $\tau = \{T_{v,w} : v \in \beta, w \in \gamma\}$ is a basis for $\text{Hom}(V, W)$, where $T_{v,w}$ is the unique transformation such that $T_{v,w}(v) = w$ and $T_{v,w}(v') = \mathbb{0}_W$ for all $v' \in \beta \setminus \{v\}$.

PROOF.

Independence To consider a truly arbitrary subset of τ, we need to represent all $T_{v_i,\times}$ and, for $T_{v_i,\times}$, any number of $\times = w_i$. Thus, we form the following combination:

$$\star \quad a_{11}T_{v_1,w_1} + ... + a_{1k}T_{v_1,w_k} + ... + a_{nl}T_{v_n,w_l} + ... + a_{nm}T_{v_n,w_m} = \mathbb{0}$$

where $\mathbb{0}$ is the transformation that sends $v \to \mathbb{0}_W$.

This must hold for all $v_i \in \beta$, so we can evaluate the combination at $v = v_1$. Since $T_{v_1,w}(w) = w$ and $T_{v_i}(w) = 0$ for $i \neq j$, $w \in \gamma$, we have

$$a_{11}w_1 + ... + a_{1k}w_k = 0 \implies a_{11} = ... = a_{1k} = 0$$

as $w_i \in \gamma$ are members of a basis. Similarly, evaluating $\star$ at any v_j will imply that $a_{v_j,w} = 0$, $w \in \gamma$. These are all our coefficients, so $\star$ is a trivial combination, and τ is linearly independent.

Spanning Consider a transformation $T : V \to W$, which sends $v_i \to w_i$ for $w_i \in W$.

$$\begin{aligned} T(v) &= T(a_1v_1 + ... + a_nv_n) \quad \text{for constants } a_i \in \mathbb{F} \\ &= a_1T(v_1) + ... + a_nT(v_n) = a_1w_1 + ... + a_nw_n \\ &= T_{v_1,w_1}(v) + ... + T_{v_n,w_n}(v) \quad \spadesuit \end{aligned}$$

where T_{v_i,w_i} sends $v_i \to w_i$ and $v_j \to 0$ for $j \neq i$. For this last step, see that

$$\begin{aligned} T_{v_i,w_i}(v) &= T_{v_i,w_i}(a_1v_1 + ... + a_nv_n) \\ &= a_1T_{v_i,w_i}(v_1) + ... + a_iT_{v_i,w_i}(v_i) + ... + a_nT_{v_i,w_i}(v_n) = a_iw_i \end{aligned}$$

Thus, it only remains to show that $T_{v_i,w_i} \in \text{Span}(\tau)$, but

$$T_{v_i,w_i}(v) = a_i w_i = a_i[b_1 w_1^* + ... + b_n w_n^*] \quad w_i^* \in \gamma, b_i \in \mathbb{F}$$
$$= a_i \left[\frac{b_1}{a_1} T_{v_1,w_1^*}(v) + ... + \frac{b_n}{a_n} T_{v_n,w_n^*}(v) \right]$$

where $w_i^* \in \gamma$. The second line requires the following justification:

$$T_{v_1,w_1^*}(v) = T_{v_1,w_1^*}(a_1 v_1 + ... + a_n v_n) = a_1 w_1^*$$

Since $w_i^* \in \gamma$, $T_{v_i,w_i^*} \in \tau$, so $T_{v_i,w_i} \in \text{Span}(\tau)$. Thus, ♠, i.e. $T(v), \in \text{Span}(\tau)$. Clearly $\text{Span}(\tau) \subseteq \text{Hom}(V, W)$, so $\text{Span}(\tau) = \text{Hom}(V, W)$, and τ is a basis. □

Proposition 2.9 If V, W are finite dimensional, then $\dim(\text{Hom}(V, W)) = \dim(V) \cdot \dim(W)$.

Proof. Let $\beta = \{v_1, ..., v_n\}$, $\gamma = \{w_1, ..., w_m\}$. Then $\{T_{v_i,w_j} : i \in [1, n], j \in [1, m]\}$ is a basis for $\text{Hom}(V, W)$ by the theorem above. This has $n \cdot m$ elements. □

MATRICES

We wish to characterize a transformation $T : \mathbb{F}^n \to \mathbb{F}^m \in \text{Hom}(\mathbb{F}^n, \mathbb{F}^m)$ in matrix form. Given T, we know it's uniquely determined by its values on the standard basis for $\mathbb{F}^n$, $\text{St}_\text{n} = \{e_1, ..., e_n\}$. Thus, T is uniquely determined by the ordered set

$$\{T(e_1), ..., T(e_n)\} \subseteq \mathbb{F}^m$$

Each $T(e_i)$ is a vector in $\mathbb{F}^m$, so we can represent it as $\langle a_{1i}, ..., a_{mi} \rangle$, where $a_{ij} \in \mathbb{F}$. Thus, form the following matrix of column vectors:

$$[T] := \begin{bmatrix} | & | & & | \\ T(e_1) & T(e_2) & \cdots & T(e_n) \\ | & | & & | \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

We call this the *matrix representation* of T in the standard basis.

Proposition 2.10 $T(v) = [T] \cdot v$, where v is represented as a column vector, $\langle b_1, ..., b_n \rangle$, for $b_i \in \mathbb{F}$.

Proof. We have $v = b_1 e_1 + ... + b_n e_n$ in the standard basis. Then, $T(v) = b_1 T(e_1) + ... +$

$b_n T(e_n)$, where $T(e_i) = \langle a_{1i}, ..., a_{mi} \rangle \subseteq \mathbb{F}^m$. In column-vector form, this is:

$$T(v) = \begin{bmatrix} b_1 a_{11} + ... + b_n a_{1n} \\ b_1 a_{21} + ... + b_n a_{2n} \\ \vdots \\ b_1 a_{m1} + ... + b_n a_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = [T]v$$

□

In this way, matrices can act as linear transformations, but we would like to formalize this idea.

For a given $m \times n$ matrix A, define the function $L_A : \mathbb{F}^n \to \mathbb{F}^m$ by $v \to A \cdot v$, where $v \in \mathbb{F}^n$. This is a linear transformation by the proposition above.

Proposition 2.10 established that every transformation T can be represented in matrix form. One can work backwards, too: given a matrix A, one forms the unique tranformation that sends $e_i \to A^{(j)}$, the j^{th} column of A.

The function from $\text{Hom}(\mathbb{F}^n, \mathbb{F}^m) \to M_{m\times n}(\mathbb{F})$ defined by $T \to [T]$ is an isomorphism. Furthermore, its inverse $M_{m\times n}(\mathbb{F}) \to \text{Hom}(\mathbb{F}^n, \mathbb{F}^m)$ is $A \to L_A$.

PROPOSITION 2.11

Linearity: We need to first show $[T_1 + T_2] = [T_1] + [T_2]$ and $[aT] = a[T]$ for $a \in \mathbb{F}$, $T \in \text{Hom}(\mathbb{F}^n, \mathbb{F}^m)$. Consider the standard basis for $\mathbb{F}^n$, $\text{St}_\text{n} = \{e_1, ..., e_n\}$. We have that

PROOF.

$$[T_1 + T_2] = \begin{bmatrix} | & | & & | \\ (T_1 + T_2)(e_1) & (T_1 + T_2)(e_2) & \cdots & (T_1 + T_2)(e_n) \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} | & | & & | \\ T_1(e_1) + T_2(e_1) & T_1(e_2) + T_2(e_2) & \cdots & T_1(e_n) + T_2(e_n) \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} | & | & & | \\ T_1(e_1) & T_1(e_2) & \cdots & T_1(e_n) \\ | & | & & | \end{bmatrix} + \begin{bmatrix} | & | & & | \\ T_2(e_1) & T_2(e_2) & \cdots & T_2(e_n) \\ | & | & & | \end{bmatrix}$$

$$= [T_1] + [T_2]$$

$a[T] = [aT]$ is shown similarly.

Inverse: If an inverse exists for $T \to [T]$, then it is bijective; as linearity has

been shown, this is sufficient to show isomorphism by Prop. 2.1.

Consider the composition $T \to [T] \to L_{[T]}$. One sees $L_{[T]}(v) = [T] \cdot v = T(v)$ by definition, so this is precisely the identity on $\text{Hom}(\mathbb{F}^n, \mathbb{F}^m)$.

Now we need to show $A \to L_A \to [L_A]$ is the identity on $M_{m\times n}(\mathbb{F})$. Consider the j^{th} column of $[L_A]$. This is the result of $L_A(e_j)$, which is $A \cdot e_j$. Thus:

$$[L_A] = \begin{bmatrix} | & | & & | \\ A\cdot e_1 & A\cdot e_2 & \cdots & A\cdot e_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ A^{(1)} & A^{(2)} & \cdots & A^{(n)} \\ | & | & & | \end{bmatrix} = A$$

□

PROPOSITION 2.12 As a corollary, we get that $\dim(\text{Hom}(\mathbb{F}^n, \mathbb{F}^m)) = \dim(M_{m\times n}(\mathbb{F}))$

Matrix Representations in Generality

Thus far we've considered matrix representations in $\mathbb{F}^n, \mathbb{F}^m$, but we can do so for general vector spaces V, W.

Let V be finite dimensional over $\mathbb{F}$, and $\beta = \{v_1, ..., v_n\}$ be a basis for V. Recall the set $\{a_1, ..., a_n\}$ for which $a_1v_1 + ... + a_nv_n = v$ is the unique representation of V in β. We call this set the *coordinates* of v in β. Represented as a column vector, define

$$[v]_\beta = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{F}^n$$

to be the *coordinate vector* of v in β.

Recall that, in the proof that all n-dimensional vector spaces V are isomorphic to $\mathbb{F}^n$, we used the transformation $T(v_i) = e_i$. We denote this function by $I_\beta : V \to \mathbb{F}^n$. For any $v \in V$, we have

$$I_\beta(v) = I_\beta(a_1v_1 + ... + a_nv_n) = a_1I(v_1) + ... + a_nI(v_n) = a_1e_1 + ... + a_ne_n = [v]_\beta$$

Thus, $I_\beta : V \to \mathbb{F}^n$ which sends $v \to [v]_\beta$ is an isomorphism.

Suppose we are given $T : V \to W$, where V and W are both finite dimensional. Let $\beta = \{v_1, ..., v_n\}$ and $\gamma = \{w_1, ..., w_m\}$ be bases of V and W, respectively. We know that T is determined by its values on β. Thus, we can encode T in matrix-form,

where the i^{th} column corresponds to $[T(v_i)]_\gamma \in \mathbb{F}^m$, as follows:

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

We call this the *matrix representation* of T from $\beta \to \gamma$.

2.9 Relation Between V, W, $\mathbb{F}^n$, and $\mathbb{F}^m$
Let V, W be of dimension n and m with bases β and γ, respectively. Let $T : V \to W$ be a linear transformation. Then the following diagram commutes:

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ {\scriptstyle I_\beta}\downarrow & & \downarrow{\scriptstyle I_\gamma} \\ \mathbb{F}^n & \xrightarrow[L_{[T]_\beta^\gamma}]{} & \mathbb{F}^m \end{array}$$

Furthermore, the function $\text{Hom}(V, W) \to M_{m\times n}(\mathbb{F})$ that maps $T \to [T]_\beta^\gamma$ is an isomorphism whose inverse is the map $M_{m\times n}(\mathbb{F}) \to \text{Hom}(V, W)$ which maps $A \to I_\gamma^{-1} \circ L_A \circ I_\beta$

PROOF. To show the diagram commutes, we essentially prove $I_\gamma \circ T = L_{[T]_\beta^\gamma} \circ I_\beta$. We have $I_\gamma \circ T(v) = [T(v)]_\gamma$, applying definitions. On the other hand,

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

Thus, we need to show that $[T]_\beta^\gamma \cdot [v]_\beta = [T(v)]_\gamma$. To do so, write $[v]_\beta = \langle a_1, ..., a_n \rangle \in \mathbb{F}^n$, and recall that

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

Then we can write

$$\begin{aligned}
[T]_\beta^\gamma \cdot [v]_\beta &= a_1[T(v_1)]_\gamma + ... + a_n[T(v_n)]_\gamma \\
&= [a_1 T(v_1) + ... + a_n T(v_n)]_\gamma && \text{by linearity of } I_\gamma \\
&= [T(a_1 v_1 + ... + a_n v_n)]_\gamma && \text{by linearity of } T \\
&= [T(v)]_\gamma && \text{and we are done } \square
\end{aligned}$$

Compositions and Matrix Multiplication

By function, we don't just mean linear transformations

Recall the composition of functions $T : V \to W$, $S : W \to X$, written as $S \circ T(v) = S(T(v))$. Compositions are associative: for functions $T \to S \to R$, we have $(R \circ S) \circ T = (R \circ S)(T(v)) = R(S(T(v))) = R(S \circ T(v)) = R \circ (S \circ T(v))$.

Consider the two linear maps $L_A : \mathbb{F}^n \to \mathbb{F}^m$, $L_B : \mathbb{F}^m \to \mathbb{F}^l$. Then the composition $L_B \circ L_A$ is itself a linear transformation, and is equal to $L_C : \mathbb{F}^n \to \mathbb{F}^l$ for some matrix $C \in M_{l\times n}(\mathbb{F})$. This unknown C is precisely $[L_B \circ L_A]$, by definition.

One can work out, explicitly, what $[L_B \circ L_A]$ is, and see that it agrees with our usual notion for $B \cdot A$

For two matrices A and B, define their product $B \cdot A$ to be $[L_B \circ L_A]$.

PROPOSITION 2.13 $L_B \circ L_A = L_{B\cdot A}$. The proof for this follows immediately from $[L_B \circ L_A] = B \cdot A$.

PROPOSITION 2.14 Matrix multiplication is associative.

PROOF. $C(BA) = C \cdot [L_B \circ L_A] = [L_C \circ (L_B \circ L_A)] = [(L_C \circ L_B) \circ L_A] = (CB)A$ $\square$

PROPOSITION 2.15 For V, W, U finite-dimensional, with bases α, β, γ, respectively, and transformations $T : V \to W$, $S : W \to U$, we have the similar statement $[S \circ T]_\alpha^\gamma = [S]_\beta^\gamma \cdot [T]_\alpha^\beta$.

Where $T := L_A$ and $S := L_B$ as above, this is equivalent to saying $[L_B \circ L_A] = B \cdot A$, which has been shown.

INVARIANTS AND NILPOTENT TRANSFORMATIONS

Preliminaries

For a function $f : X \to Y$, we call $g : Y \to X$

i.e. takes $x \to x$

1. a *left inverse* if $g \circ f = I_X$, the identity on X
2. a *right inverse* if $f \circ g = I_Y$, the identity on Y

Sometimes called a two-sided inverse

3. an *inverse* if g is both a left and right inverse

Also consider the following facts, whose proofs are good exercise:

1. f has a left inverse $\iff$ f is injective
2. f has a right inverse $\iff$ f is surjective

3. f has an inverse $\iff$ f is bijective

Examples:

1. $\delta : \mathbb{F}[t]_{n+1} \to \mathbb{F}[t]_n$, the derivative of polynomials, has a right inverse, namely the anti-derivative.
2. Let $f : \mathbb{F}[[t]] \to \mathbb{F}[[t]]$ be the left shift map of coefficients, i.e. $\sum_0^\infty a_n t^t \to \sum_1^\infty a_n t^{n-1}$. This has a right inverse, namely the right shift map of coefficients, $\sum_0^\infty a_n t^n \to \sum_0^\infty a_n t^{n+1}$. Recall that $\mathbb{F}[[t]]$ is the set of formal power series.

PROPOSITION 2.16

Let $T : V \to W$ be a transformation of vector spaces of the same (finite) dimension. Then TFAE:

T has a right inverse $\quad$ T has a left inverse $\quad$ T has an inverse

PROOF.

This follows directly from Prop. 2.7, which states that transformations over n dimensional spaces are surjective IFF injective □

PROPOSITION 2.17

Recall that an $n \times n$ dimensional matrix A is called invertible IFF there exists B such that $A \cdot B = B \cdot A = I$, the identity matrix. One notates $B = A^{-1}$.

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 $\iff$ it has a right inverse.

PROOF.

L_A is invertible $\iff$ there exists $T : \mathbb{F}^n \to \mathbb{F}^n$ such that $L_A \circ T = T \circ L_A = I_{\mathbb{F}^n} \iff \exists B \in M_n(\mathbb{F})$ with $L_A \circ L_B = L_B \circ L_A = I_{\mathbb{F}^n} \iff \exists B$ s.t. $L_{AB} = L_{BA} = I_{\mathbb{F}^n} \iff \exists B$ s.t. $AB = BA = [I]$, and $[I]$ is the identity matrix (this last bit has not been previously shown, but the verification is easy).

This shows (1), and (2) follows directly. □

T-Invariants

Let $T : V \to V$ be a linear transformation over a vector space V. Transformations of this form are sometimes called *linear operators*. A subspace $W \subseteq V$ is called T *invariant* if $T(W) \subseteq W$.

i.e., you can apply T to W an indeterminate amount of times, and it will always remain as a subset of itself.

Examples:

1. For $T : V \to V$, both $\ker(T)$ and $\mathrm{Im}(T)$ are T-invariant

For (1), note that $T(\mathrm{Im}(T)) \subseteq \mathrm{Im}(T)$ by definition, and $T(\ker(T)) = \mathbb{0}_V \in \ker(T)$

2. For any $n \in \mathbb{N}$, where $T^n := \underbrace{T \circ T \circ ... \circ T}_{n \text{ times}}$, $\ker(T^n)$ is T-invariant.

Proof to come

3. For $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(x, y, z) = \langle 2x + y, 3x - y, 7z \rangle$, both the xy-plane and z-axis are T-invariant. As proof, observe $T(x, y, 0) = \langle 2x + y, 3x - y, 0 \rangle \subseteq$ xy-plane, and also $T(0, 0, z) = \langle 0, 0, 7z \rangle \subseteq z$-axis. In fact, $\mathbb{R}^3$ always decomposes into a direct sum of 2 T–invariant subspaces, the xy-plane and z-axis.

Proposition 2.18 For $T : V \to V$, and any n, we have

1. $V \supseteq \text{Im}(T) \supseteq \text{Im}(T^2) \supseteq ...$, and $\text{Im}(T^n)$ is T-invariant.
2. $\{\mathbb{0}_V\} \subseteq \ker(T) \subseteq \ker(T^2) \subseteq ...$, and $\ker(T^n)$ is T-invariant.

Proof. (1): Let $x \in \text{Im}(T^{n+1})$. Then $x = T^{n+1}(y) = T^n(T(y)) \in \text{Im}(T^n)$ for some y, so $\text{Im}(T^n) \supseteq \text{Im}(T^{n+1})$. Now let $x \in \text{Im}(T^n)$. Then $x = T^n(y)$, so $T(x) = T(T^n(y)) = T^n(T(y))$, and we conclude $T(x) \in \text{Im}(T^n)$, i.e. $T(\text{Im}(T^n)) \subseteq \text{Im}(T^n)$, and $\text{Im}(T^n)$ is T-invariant.

(2): Let $x \in \ker(T^n)$. Then $T^{n+1}(x) = T(T^n(x)) = T(\mathbb{0}) = \mathbb{0}$, so $x \in \ker(T^{n+1})$, and $\ker(T^n) \subseteq \ker(T^{n+1})$. We also see that $T(x) \in \ker(T^n)$, since $T(T^n(x)) = T^n(T(x)) = \mathbb{0}$, from before. Thus, $\ker(T^n)$ is T-invariant. □

Nilpotent Transformations

Nilpotency has varying definitions in mathematics: for a ring R, $r \in R$ is called nilpotent if $r^n = 0$ for some n. In our study, a linear transformation $T : V \to V$ is called *nilpotent* if $T^n = 0$ for some n, and a matrix $A \in M_n(\mathbb{F})$ is *nilpotent* if $A^n = 0$ for some n.

Examples:

1. Let V be an n-dimensional vector space over $\mathbb{F}$ with a basis $\beta = \{v_1, ..., v_n\}$, and let $T : V \to V$ be the unique transformation that "shifts" basis members, i.e. $T(v_1) = \mathbb{0}_V$, $T(v_2) = v_1$, $T(v_3) = v_2$, etc. Then T^n sends $v_i \to v_{i-n} = 0$ for $i \leq n$, which is all vectors on the basis, so T is nilpotent.
2. $\delta : \mathbb{F}[t]_n \to \mathbb{F}[t]_n$, the differentiation function on polynomials, is nilpotent, since $\delta^{n+1} = 0$ (the $n + 1^{\text{th}}$ derivative of $\leq n$-degree polynomials is 0).
3. For $A \in M_n(\mathbb{F})$, A is nilpotent $\iff$ $L_A : \mathbb{F}^n \to \mathbb{F}^n$ is nilpotent. As proof, recall that $L_{[A^k]} = L^k_{[A]}$, so $L^k_{[A]} = 0 \iff L_{[A^k]} = 0 \iff A^k = 0$, since $L_A \cong A$.

c.f. Prop. 2.11, 2.14

4. Matrices which are strictly upper triangle (i.e. 0s on $i \leq j$) are nilpotent.

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

PROPOSITION 2.19

NOTATION
For $f : X \to Y$, $A \subseteq X$, define the *restriction* of f to A, $f_A : A \to Y$, taking $a \to f(a)$

2.10 Fitting's Theorem
For an n-dimensional vector space V over $\mathbb{F}$ and $T : V \to V$, there exists a decomposition $V = U \oplus W$, where $U, W \subseteq V$ are such that $T_U : U \to U$ is nilpotent and $T_W : W \to W$ is an isomorphism.

PROOF.

Recall that

$$V \supseteq \mathrm{Im}(T) \supseteq \mathrm{Im}(T^2) \supseteq \ldots \text{ and } \{\mathbb{0}_V\} \subseteq \ker(T) \subseteq \ker(T^2) \subseteq \ldots$$

$$\implies n \geq \dim(\mathrm{Im}(T)) \geq \ldots \text{ and } 0 \leq \dim(\ker(T)) \leq \ldots$$

Since both $\dim \ker(T^k)$ and $\dim \mathrm{Im}(T^k)$ are bound by $[0, n]$, these inequalities may be strict at most n times, so $\exists N \in \mathbb{N}$ such that $\forall\ k \geq N$, $\dim(\mathrm{Im}(T^{k+N})) = \dim(\mathrm{Im}(T^N))$. Note that $\mathrm{Im}(T^{k+N}) \subseteq \mathrm{Im}(T^N)$, so this necessarily means that $\mathrm{Im}(T^{k+N}) = \mathrm{Im}(T^N)$ (c.f. Thm. 1.6). Similarly, $\ker(T^{k+N}) = \ker(T^N)$.

Let $U := \ker(T^N)$ and $W := \mathrm{Im}(T^N)$. We know that these sets are T-invariant.

$T|_U$ is nilpotent: $T^N(\ker(T^N)) = \mathbb{0}$ by definition. We also see that $T|_U$ maps to U as claimed, since $\ker(T^N)$ is T-invariant.

$T|_W$ is an isomorphism: $T(\mathrm{Im}(T^N)) = \mathrm{Im}(T^{N+1}) = \mathrm{Im}(T^N)$ by assumption, so $T|_W$ is surjective. Thus, $T|_W$ is also injective, by Prop. 2.7., and is an isomorphism.

Lastly, we need to show that $U \oplus W = V$ and $U \cap W = \{\mathbb{0}\}$. For the latter, suppose $v \in U \cap W$. Then $T^N(v) = \mathbb{0}$ as shown, and T is an isomorphism over W, so $v = \{\mathbb{0}\}$.

$\dim(U \oplus W) = \dim(U) + \dim(W) - \dim(U \cap W) = \dim(U) + \dim(W) = \dim(\ker(T^N)) + \dim(\mathrm{Im}(T^N)) = \dim(V)$, which means $U \oplus W = V$ again by Thm 1.6. □

DUAL SPACES

For a vector space V over $\mathbb{F}$, we call a linear transformation $V \to \mathbb{F}$ a *linear functional*. The space of linear functionals, i.e. $\mathrm{Hom}(V, \mathbb{F})$, is denoted V^*, and is called the *dual space* of V.

For finite dimensional V, we already know that $\dim(V^*) = \dim(\text{Hom}(V, \mathbb{F})) = \dim(V) \cdot \dim(\mathbb{F}) = \dim(V)$. In accordance with our construction of a basis for Hom (pp. 19-20), we let $\beta := \{v_1, ..., v_n\}$ be a basis for V and $\gamma = \{1\}$ be the standard basis for $\mathbb{F}$. Then $\beta^* := \{f_1, ..., f_n\}$ is a basis for $\text{Hom}(V, \mathbb{F}) = V^*$, where $f_i : V \to \mathbb{F}$ are precisely $T_{v_i,1}$ in our previous notation, i.e. $f_i(v_i) = 1$ and $f_i(v_j) = 0$ when $i \neq j$. We call the set β^* the *dual basis* for β.

Proposition 2.20

β^* is a basis for V^*, and every $f \in V^*$ has the unique representation

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

Proof.

The first part of this proposition is just a special case of Theorem 2.8, as discussed above. f thus *does* have a unique representation in β^*, so if $f = \sum_{i=1}^{n} f(v_i) f_i = f$, then this is indeed unique. It is enough to show that these functions agree on $v_i \in \beta$, as any $v \in V$ could be representation by linearity.

$$\sum_{i=1}^{n} f(v_i) f_i(v_j) = f(v_i) f_i(v_i) = f(v_i) \quad \square$$

Remark: we will use the *Kronecker delta* function in the future. It is defined to be

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Note that $f_i(v_j) = \delta_{ij}$

Example: Let $V := \mathbb{F}^n$ be viewed as a vector space over $\mathbb{F}$. Then V^* has a basis $\beta^* := \{f_1, ..., f_n\}$, where $f_i(e_j) = \delta_{ij}$. Since f_i are linear transformations, they can be represented as L_{A_i}, where $A_i \in M_{1\times n}(\mathbb{F})$. We can then deduce that $A_i = [0, ..., 0, 1, 0, ..., 0]$, the row vector with a 1 in the i^{th} position.

Just as we took a dual space of V, we can take a dual space *of* the dual space, and denote it V^{**}. Since $\dim(V) = \dim(V^*)$ in finite dimensions, we know $\dim(V^*) = \dim(V^{**})$, and conclude that $\dim(V) = \dim(V^{**})$. From this statement arises an abstract notion of isomorphism between V and V^{**}.

It can be shown as exercise that the natural map from $V \to V^*$ which takes $v_i \to f_i$ is a vector space isomorphism. We'll try to form a similar natural map between V and V^{**} to strengthen notations of their isomorphism.

Let V be an arbitrary vector space over $\mathbb{F}$. For each $x \in V$, define $\hat{x} \in V^{**}$ to be a function from $V^* \to \mathbb{F}$ that takes $f \to f(x)$. Another way of writing this is: $\hat{x} = f(x)$, where $f \in V^*$.

2.11 The function $x \to \hat{x}$ is an isomorphism from $V \to V^{**}$.

Proof.

If $x \to \hat{x}$ is injective, it will follow immediately that, if $\dim(V) < \infty$, then $x \to \hat{x}$ is an isomorphism, as it must also be surjective (recall that $\dim(V) = \dim(V^{**})$).

Let $x \in V$, and let $\hat{x} = \mathbb{0}_{V^{**}}$. We have a unique representation $a_1 v_1 + ... + a_n v_n = x$ in a basis $\beta = \{v_1, ..., v_n\}$ for V. Then $\hat{x}$ takes $f \to f(x)$ for $f \in V^*$, so $\hat{x}(f_i) = f_i(x) = f_i(a_1 v_1 + ... + a_n v_n) = a_i$. But $\hat{x} = \mathbb{0}$, so $a_i = 0$. Now, since $\hat{x}(f_i) = a_i$ in generality, all $a_i = 0$, so $x = 0$. □

Let V be a vector space and $S \subseteq V$ some sub**set**. Then we call the set

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

the *annihilator* of S.

We observe the following facts about the annihilator of $S \subseteq V$: PROPOSITION 2.21

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

For (1), we have $(af_1 + f_2)(u) = af_1(u) + f_2(u) = 0$ for any $u \in S$, so then $af_1 + f_2 \in S^\perp$. (2)'s proof is just an observation: if $S_1 \subseteq S_2$, then we will find more $f \in V^*$ which map to 0 on S_1 than those which map to 0 on S_2, as the latter is just a more restrictive condition. For (3), note that, if $f \in V^*$ takes all $u \in S$ to 0, then it must also take all linear combinations of $u \in S$ to 0, so $S^\perp \subseteq (\text{Span}(S))^\perp$. The converse holds by (2). □ PROOFS.

For $S \subseteq V$, we denote $\hat{S} := \{\hat{x} : x \in S\} \subseteq V^{**}$ in the finite-dimensional case. From Theorem 2.11, we have $\hat{V} = V^{**}$. Some texts will refer to V^{**} explicitly as $\hat{V}$, but this is a notational preference that we will not indulge.

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

$\hat{x} \in (U^\perp)^\perp \iff \hat{x}(f) = f(x) = 0 \ \forall f \in U^\perp$. Hence, if $x \in U$, then $\hat{x} \in (U^\perp)^\perp$, and we conclude that $\hat{U} \subseteq (U^\perp)^\perp$. PROOF.

That was the easy direction. For the converse, if $\hat{x} \in (U^\perp)^\perp$, then we know $f(x) = 0 \ \forall f \in U^\perp$. We want to show that $x \in U$. Suppose otherwise. Then we define $f \in U^\perp$ such that $f(x) = 1$, by which a contradiction arises.

Let $\{u_1, ...u_k\}$ be a basis for U. Note that, since $x \notin U$, the set $\{u_1, ..., u_k, x\}$ is still linearly independent. We can thus extend this to a basis for U, i.e. $\{u_1, ..., u_k, x, v_1, ..., v_m\}$. Define $f \in V^*$ that takes all elements of this basis to 0 except x, which is mapped to 1. Observe, then, that $f(u) = 0 \forall u \in U$, so $f \in U^\perp$. But $f(x) = 1$ ↯

$\implies x \in U$, and thus $\hat{x} \in \hat{U} \implies \hat{U} = (U^\perp)^\perp$ □

Proposition 2.22 For a finite dimensional vector space V and subspace $U \subseteq V$, we have

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

Proof. We know the $\subseteq$ direction holds trivially. Suppose $x \in V$ is such that $f(x) = 0$ for any $f \in U^\perp$. Then $\hat{x} \in (U^\perp)^\perp$, and from above, $\hat{x} \in \hat{U}$, so $x \in U$. □

Let V, W be vector spaces over $\mathbb{F}$ and $T : V \to W$ be a linear transformation. The *dual* or *transpose* of T is the map $T^t : W^* \to V^*$ which takes $g \to g \circ T$.

Proposition 2.22a If $\dim(V) = n$ and $U \subseteq V$, then $\dim(U^\perp) + \dim(U) = \dim(V) = n$. As proof, we let $\{v_1, ..., v_k\}$ be a basis for U, $\{v_1, ..., v_k, v_{k+1}, ..., v_n\}$ be a basis for V, and notice that $\{f_{k+1}, ..., f_n\}$ is a basis for $U^\perp$.

Proposition 2.23 The transpose has the following properties:

1. $T^t : W^* \to V^*$ is linear
2. $\ker(T^t) = (\text{Im}(T))^\perp$
3. $\text{Im}(T^t) = (\ker(T))^\perp$. If V, W are finite dimensional, we also have $\dim(\text{Im}(T)) = \dim(\text{Im}(T^t))$.
4. If V, W are finite dimensional with bases β, γ, respectively, the

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

Proofs. For (1), $T^t(ag_1 + g_2) = (ag_1 + g_2) \circ T = a(g_1 \circ T) + (g_2 \circ T) = aT^t(g_1) + T^t(g_2)$.

For (2), $g \in \ker(T^t) \iff T^t(g) = \mathbb{0} \iff T^t(g)(v) = 0 \ \forall v \in V \iff g(T(v)) = 0 \iff g(w) = 0 \ \forall w \in \text{Im}(T) \iff g \in (\text{Im}(T))^\perp$

For (3), fix $f \in \text{Im}(T^t)$ and $u \in \ker(T)$. Then note that that $f(u) = T^t(g)(u)$ for some $g \in W^*$. Then $T^t(g)(u) = g(T(u)) = g(\mathbb{0}_W) = 0$, so $f \in (\ker(T))^\perp$. We conclude that $\text{Im}(T^t) \subseteq (\ker(T))^\perp$.

Now suppose that V, W are both finite dimensional. The obvious roadmap to

showing equality, since we've shown inclusion, is showing equal dimensionality between $\ker(T^t)$ and $(\mathrm{Im}(T))^\perp$.

$\dim(\mathrm{Im}(T^t)) = \dim(W^*) - \dim(\ker(T^t))$ by rank-nullity. But the dimension of W^* is the same as that of W. Furthermore, we know $\dim(\ker(T^t)) = \dim((\mathrm{Im}(T))^\perp)$ by (2), so $\dim(\mathrm{Im}(T^t)) = \dim(W) - \dim((\mathrm{Im}(T))^\perp)$. Then $\dim((\mathrm{Im}(T))^\perp) = \dim(W) - \dim(\mathrm{Im}(T))$, so we conclude that $\dim(\mathrm{Im}(T^t)) = \dim(\mathrm{Im}(T))$.

On the other hand, $\dim((\ker(T))^\perp) = \dim(V) - \dim(\ker(T))$, which, by rank-nullity, is $\dim(\mathrm{Im}(T))$. Thus, $\dim(\mathrm{Im}(T^t)) = \dim((\ker(T))^\perp)$

For (4), let β, γ be finite bases for V and W, respectively, and recall that $A := [T]_\beta^\gamma$ is the matrix

$$\begin{bmatrix} | & | & & | \\ [T(v_1)]_\gamma & [T(v_2)]_\gamma & \cdots & [T(v_n)]_\gamma \\ | & | & & | \end{bmatrix}$$

where $\beta = \{v_1, ..., v_n\}$ and $\gamma = \{w_1, ..., w_m\}$. Then $A^{(j)} = [T(v_j)]_\gamma$, and hence $T(v_j) = \sum_{k=1}^{m} A_{kj} w_k$. Similarly, we express $B := [T^t]_{\gamma^*}^{\beta^*}$ as the matrix

$$\begin{bmatrix} | & | & & | \\ [T^t(g_1)]_{\beta^*} & [T^t(g_2)]_{\beta^*} & \cdots & [T^t(g_m)]_{\beta^*} \\ | & | & & | \end{bmatrix}$$

where $\gamma^* = \{g_1, ..., g_m\}$ and $\beta^* = \{f_1, ..., f_n\}$. Then $T^t(g_i) = \sum_{j=1}^{n} B_{ji} f_j = \sum_{j=1}^{n} T^t(g_i)(v_j) \cdot f_j$, so $B_{ji} = T^t(g_i)(v_j)$. It remains to show that $B_{ji} = A_{ij}$.

$$\begin{aligned} B_{ji} &= T^t(g_i)(v_j) = g_i(T(v_j)) \\ &= g_i\left(\sum_{k=1}^{m} A_{kj} w_k\right) = \sum_{k=1}^{m} A_{kj} g_i(w_k) \\ &= \sum_{k=1}^{m} A_{kj} \delta_{ik} = A_{ij} \end{aligned}$$

□

Proposition 2.24 Let V, W be vector spaces over $\mathbb{F}$ and $T : V \to W$ be some linear transformation. Then

1. T^t is injective $\iff$ T is surjective
2. T is injective $\iff$ T^t is surjective, provided that V, W finite dimensional.

Proof. For (1): we know that T^t is injective IFF $\ker(T^t) = \{0\}$, which happens $\iff$ $(\text{Im}(T))^\perp = \{0\}$ by part (2) of Prop 2.23. This implies that $\text{Im}(T) = W$, i.e. T is surjective, by Duality (i.e. Prop 2.22). Conversely, if $\text{Im}(T) = W$, then the function which takes all of W to 0 is precisely $\mathbb{0}_{W^*}$, i.e. $(\text{Im}(T))^\perp = 0$. Then part (2) from Prop 2.23 says $\ker(T^t) = 0$, i.e. T^t is injective.

Similarly for (2), if T is injective, then $\ker(T) = \{0\}$, so $(\ker(T))^\perp = V^*$. Then part (3) of Prop 2.23 says that $\text{Im}(T^t) = V^*$. Thus T^t is surjective. Conversely, if $\text{Im}(T^t) = V^*$, then $(\ker(T))^\perp = V^*$. The only element which is *always* taken to 0 is 0, so $\ker(T) = \{0\}$, i.e. T is injective. □

Applications of Dual Spaces on Matrices

Recall that, for $T : V \to W$, the rank of T is $\dim(\text{Im}(T))$. Furthermore, if $\beta = \{v_1, ..., v_n\}$ is a basis for V, then $\text{Im}(T) = \text{Span}(\{T(v_1), ..., T(v_n)\})$. In particular, $\dim(\text{Im}(T))) \leq n$, where $\dim(V) = n$ (see dimension theorem). Thus, we can express $\dim(\text{Im}(T))$ as the size of a maximally independent subset of $\{T(v_1), ..., T(v_n)\}$.

For an $m \times n$ matrix $A \in M_{m\times n}(\mathbb{F})$, define $\text{rank}(A)$, or the *rank* of A, by $\text{rank}(\text{Im}(L_A))$.

Define also the *column rank* of A, denoted $\text{c-rank}(A)$, to be the size of a maximally independent subset of $\{A^{(1)}, ..., A^{(n)}\}$, where $A^{(j)}$ denotes the j^{th} column of A.

Finally, we define the *row rank*, or $\text{r-rank}(A)$, to be the size of a maximally independent subset of $\{A_{(1)}, ..., A_{(m)}\}$, where $A_{(i)}$ denotes the i^{th} row of A.

Proposition 2.25 $\text{rank}(A) = \text{c-rank}(A)$, and this follows from the definitions.

Proposition 2.26 $\text{rank}(A) = \text{rank}(A^t) = \text{r-rank}(A)$.

Proof. We know that $\text{rank}(A^t) = \text{c-rank}(A^t) = \text{r-rank}(A)$, and thus we only need to show that $\text{rank}(A) = \text{rank}(A^t)$. But we've seen that $\dim(\text{Im}(T)) = \dim(\text{Im}(T^t))$ from above, so $\text{rank}(A) = \text{rank}(L_A) = \text{rank}(L_A^t)$. Then $\text{rank}(A) = \text{rank}(A^t)$ by part (4) of the same proposition (one should ponder about what $\beta, \gamma, \beta^*, \gamma^*$ are). □

We then conclude that $\text{c-rank}(A) = \text{r-rank}(A) = \text{rank}(A)$ for all $A \in M_{m\times n}(\mathbb{F})$.

SYSTEM OF LINEAR EQUATIONS

A *system of linear equations* over some field $\mathbb{F}$ is as follows:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n &= b_m \end{aligned}$$

where $a_{ij}, b_i \in \mathbb{F}$ and x_j are variables. We can re-write this as follows: $A \cdot x = b$, where $x = \langle x_1, ..., x_n \rangle$ and $b = \langle b_1, ..., b_m \rangle \in \mathbb{F}^m$. Thus, x is a solution to $Ax = b$ IFF $L_A(x) = b$ IFF $x \in L_A^{-1}(b)$ (reads: x is in the preimage of $L_A(b)$).

Hence, $Ax = b$ has a solution IFF $b \in \text{Im}(L_A) = \text{Span}(\{A^{(1)}, ..., A^{(n)}\})$. In particular, if $b = 0$, we always have a solution, namely $x = 0$. There may also be non-zero solutions: call $Ax = 0$ the *homogeneous system of equations* for A. We observe that the homogeneous system has non-zero solutions exactly when $\ker(L_A)$ is non-trivial.

Note that, if y is a solution to a homogeneous system, and $Ax = b$, then $A(x+y) = b$ by linearity. Thus, for $A \in M_{m\times n}(\mathbb{F})$ and $b \in \text{Im}(L_A)$, the set of solutions to $Ax = b$ is precisely the coset $v + \ker(L_A)$, where v is a particular solution to $Ax = b$, i.e. $A \cdot v = b$.

PROPOSITION 2.27

PROOF.

Indeed, $v + a$, where $a \in \ker(L_A)$ and v is a solution to $Ax = b$, is also a solution to $Ax = b$. Conversely, if v and w are solutions to $Ax = b$, then $A(w - v) = b - b = 0$, so $w - v \in \ker(L_A)$. We then write $w = v + (w - v) = v + a$ for some $a \in \ker(L_A)$. □

If $m < n$, and $A \in M_{m\times n}(\mathbb{F})$, then there exists a non-zero solution to $Ax = 0$.

PROPOSITION 2.28

PROOF.

$\text{null}(L_A) = n - \text{rank}(L_A) = n - \dim(\text{Im}(L_A)) > n - m > 0$, so $\ker(L_A) \neq \{\mathbb{0}\}$ □

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

PROPOSITION 2.29

1. $\ker(A)$ is trivial $\iff$ $Ax = b$ has *at most* 1 solution for each $b \in \mathbb{F}^m$
2. If $n = m$, then A is invertible $\iff$ $Ax = b$ has *exactly* one solution.

PROOF.

Part (a) follows from our statement about the coset representation of the solution set (Prop 2.27), and part (b) follows from (a), with consideration of

the fact that $\ker(A) = \{\mathbb{0}\} \iff L_A$ injective $\iff L_A$ surjective. □

Elementary Operations

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

Type I Interchanging any 2 rows/columns

Type II Multiplying a row/column by some non-zero scalar in $\mathbb{F}$

Type III Adding to a row/column a scalar multiple of some other row/column

We refer to these operations by their type. Observe that these operations are invertible linear transformations, and therefore have a matrix representation. In particular, one can invert an operation by performing one of the same type.

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

Example:

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

are obtained from I_3 via Type I, II, and III elementary row operations, respectively.

PROPOSITION 2.30 Each elementary matrix $E \in M_n(\mathbb{F})$ obtained by I_n via a row operation may be obtained from I_n via a column operation of the same type, and vice-versa.

PROOF. Proof left as observational exercise. □

2.13 Consistency of Elementary Operations
For $A, B \in M_{m\times n}(\mathbb{F})$, if B is obtained from A by applying an elementary row operation, then $B = EA$, where $E \in M_m(\mathbb{F})$ is obtained from I_m via the same operation. If B is instead obtained via a *column* operation, then $B = AE$, $E \in M_n(\mathbb{F})$, and E is obtained from I_n.

PROPOSITION 2.31 Conversely, if E is an elementary matrix obtained via some row or column operation, then EA or AE is obtained from A by the same row or column operation, respectively. Proofs of this and the theorem above are observational.

PROPOSITION 2.32 Elementary matrices are invertible.

PROOF.

This just follows from the fact that elementary operations are invertible by one of the same type. □

If $A \in M_{m\times n}$, $P \in GL_m(\mathbb{F})$, and $Q \in GL_n(\mathbb{F})$, then $\operatorname{rank}(PA) = \operatorname{rank}(A) = \operatorname{rank}(AQ)$. More generally, if $T : V \to W$ is linear with V, W finite dimensional, and $S : W \to W$, $R : V \to V$ are linear and invertible, then $\operatorname{rank}(S \circ T) = \operatorname{rank}(T) = \operatorname{rank}(T \circ R)$. PROPOSITION 2.33

PROOF.

Recall that $GL_n(\mathbb{F})$ is the space of invertible square matrices of size n. We only need to prove the latter part of this statement, as the claim about matrices is just a special case of the more general claim.

We have $\operatorname{rank}(T) = \dim(\operatorname{Im}(T))$. Since $S : W \to W$ is an isomorphism, $S|_{\operatorname{Im}(T)}$ is injective, and thus $\dim(\ker(S|_{\operatorname{Im}(T)})) = \{\mathbb{0}\}$, i.e. $\dim(\operatorname{Im}(T)) = \dim(\operatorname{Im}(S|_{\operatorname{Im}(T)}))$. But $S|_{\operatorname{Im}(T)} = S \circ T$, so indeed $\operatorname{rank}(T) = \operatorname{rank}(S \circ T)$.

For $\operatorname{rank}(T \circ R)$, observe that $\operatorname{Im}(T(R(v))) = \operatorname{Im}(T)$, since R is bijective. □

We immediately see that elementary row/column operations are rank-preserving, as they are invertible, i.e. $B = E_R A E_C \implies \operatorname{rank}(A) = \operatorname{rank}(B)$, where E_R, E_C are elementary row/column operations, respectively.

2.14 Reduction of Square Matrices

Every square matrix $A \in M_n(\mathbb{F})$ can be transformed into a matrix B of the following form, using row or column operations. In particular, $r = \operatorname{rank}(A)$:

$$B = \underbrace{\begin{pmatrix} I_r & \overbrace{0}^{r\times n-r} \\ \underbrace{0}_{n-r\times r} & \underbrace{0}_{n-r\times n-r} \end{pmatrix}}_{n\times n}$$

PROOF.

We'll show this by induction on n. This clearly holds for $n = 1$ (just multiply by a_{11}^{-1}). For $n - 1 \to n$, if $A = 0$, then we have nothing to show. Otherwise, assume A has a non-zero element, and swap at most two rows and two columns to get this element in the a_{11} position (in practice, this is swapping columns 1 and n', and rows 1 and m', where the non-zero element is $a_{m'n'}$). Thus, we assume that $a_{11} \neq 1$ by taking a scalar inverse.

We perform repeated type III row operations that take $A_i \to A_{(i)} - A_{(1)}a_{i1}$ (i.e. Gaussian elimination) to get all elements $a_{i1} = 0$. Similarly, we perform the

column operation $A^{(j)} \to A^{(j)} - A^{(1)}a_{1j}$ to get all elements $a_{j1} = 0$. We end up with the following matrix after all aforementioned operations:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & & & & \\ \vdots & & \boxed{\text{A'}} & & \\ 0 & & & & \\ 0 & & & & \end{pmatrix}$$

Now, by our induction hypothesis, we know there exists operations E_R, E_C such that $E_R A' E_C$ is of the desired form. Note that these operations may still be performed on the larger matrix above, and, since they effect only rows and columns ≥ 2, the zero entries a_{i1} and a_{1j} will remain 0.

Thus, we have transformed A into $\begin{pmatrix} \boxed{I} & \boxed{0} \\ \boxed{0} & \boxed{0} \end{pmatrix}$ as desired □

Proposition 2.34 For each $A \in M_n(\mathbb{F})$, there exists invertible matrices $P, Q \in GL_n(\mathbb{F})$ such that $B = PAQ$ is of the form $\begin{pmatrix} \boxed{I} & \boxed{0} \\ \boxed{0} & \boxed{0} \end{pmatrix}$ as above. Moreover, P and Q are products of elementary matrices.

Proof. This follows immediately from the theorem above, observing that P are row operations and Q are column operations. □

Proposition 2.35 Every invertible matrix $A \in GL_n(\mathbb{F})$ is a product of invertible matrices.

Proof. There exist $P, Q \in GL_n(\mathbb{F})$ such that $PAQ = \begin{pmatrix} \boxed{I} & \boxed{0} \\ \boxed{0} & \boxed{0} \end{pmatrix}$ with $r = n$, i.e. $PAQ = I_n$, since A is invertible. Then $A = P^{-1}Q^{-1}$, but P and Q are themselves products of elementary matrices, so $A = [R_1 \cdot ... \cdot R_k]^{-1}[C_1 \cdot ... \cdot C_l]^{-1} = R_k^{-1} \cdot ... \cdot R_1^{-1} C_l^{-1} \cdot ... \cdot C_1^{-1}$, where R_i and C_i are row and column operations. □

Proposition 2.36 The transpose of an invertible matrix is invertible, and $(A^t)^{-1} = (A^{-1})^t$. This does not follow from anything given above, but is useful nonetheless.

Proof. $AA^{-1} = I = A^{-1}A$, so $(A^{-1})^t A^t = I_n^t = A^t (A^{-1})^t$, but $I_n^t = I_n$, so in particular $(A^t)^{-1} := (A^{-1})^t$ yields $(A^t)^{-1} A^t = I_n = A^t (A^t)^{-1}$, as desired. □

From Prop 2.35, we noted that an invertible matrix A can be written as $A = E_1 \cdot ... \cdot E_k$ for elementary matrices E_i, and thus $A^{-1} = E_k^{-1} \cdot ... \cdot E_1^{-1}$. Thus, treating E_i^{-1} as row operations, we have $E_k^{-1} \cdot ... \cdot E_1^{-1} A = I_n$. We conclude that the same operations which turn A into I_n also turn I_n into A^{-1}.

Define now the *augmented matrix* $(A|B)$ to be the matrix whose first columns are that of A, and last columns that of B. Note that A and B must have the same number of rows.

Observe now that $B(A|I_n) = (BA|BI_n) = (BA|B)$. In particular $E_k^{-1} \cdot ... \cdot E_1^{-1}(A|I_n) = (E_k^{-1} \cdot ... \cdot E_1^{-1} A|E_k^{-1} \cdot ... \cdot E_1^{-1}) = (I_n|A^{-1})$. Thus, there exist elementary row operations which turn $(A|I_n)$ to $(I_n|A^{-1})$.

Let $A \in M_n(\mathbb{F})$ be invertible. If row operations turn $(A|I_n)$ into $(I_n|B)$, then $B = A^{-1}$. PROPOSITION 2.37

Solving Linear Systems with Row Operations

For matrices $A_1, A_2 \in M_{m\times n}(\mathbb{F})$ and $b_1, b_2 \in \mathbb{F}^m$, then the systems $A_1 x = b_1$ and $A_1 x = b_2$ are called *equivalent* if their solution sets are equal. In particular, any two systems with no solutions are equivalent.

Observe that, for $G \in GL_m(\mathbb{F})$, $A \in M_{m\times n}(\mathbb{F})$, then the system $GAx = Gb$ is equivalent to the system $Ax = b$. As we know, multiplying on the left by an elementary matrix corresponds to a row operation, so clearly $EAx = Eb$ and $Ax = b$ are equivalent, meaning:

Just multiply by G^{-1} or G to see this!

The system encoded in $E(A|b)$ is equivalent to that of $(A|b)$, where $A \in M_{m\times n}$, $b \in \mathbb{F}^m$, and E is an elementary row operation. This follows from the observations above. PROPOSITION 2.38

Let $B \in M_n(\mathbb{F})$. We say that B is in *row echelon form* if:

(*i*) For each row $B_{(i)}$, the first non-zero entry (i.e. pivot) occurs at a column j, where j is strictly larger than the column in which $B^{(i-1)}$ has its first non-zero entry.

(*ii*) All rows with only zero entries are at the bottom of the matrix

If B is in row echelon form and all pivots (i.e. first non-zero entry of a given row) are equal to 1, then we say that B is in *reduced row echelon form*, or RREF.

2.15 Existence of Gaussian Elimination
There exist elementary operations of types I and III which transform $A \in M_{m\times n}(\mathbb{F})$ into reduced echelon form. Moreover, the addition of type II operations can yield RREF. This procedure is called *Gaussian elimination*.

The proof of this is tedious and observational, but consider the following example:

$$A = \begin{pmatrix} 3 & 2 & 3 & -2 \\ 1 & 1 & 1 & 0 \\ 1 & 2 & 1 & -1 \end{pmatrix} \xrightarrow{\text{I}} \begin{pmatrix} 1 & 2 & 1 & -1 \\ 1 & 1 & 1 & 0 \\ 3 & 2 & 3 & -2 \end{pmatrix} \xrightarrow{\text{III}} \begin{pmatrix} 1 & 2 & 1 & -1 \\ 0 & -1 & 0 & 1 \\ 0 & -4 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow{\text{III}} \begin{pmatrix} 1 & 2 & 1 & -1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & -3 \end{pmatrix} \xrightarrow{\text{III}} \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & -3 \end{pmatrix} \xrightarrow{\text{III}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix} \xrightarrow{\text{II}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2.16 Characterization of Solution Set

For a system $Ax = b$, we form $A'x = b'$ via Gaussian elimination on $(A|b)$.

1. $Ax = b$ has a solution $\iff$ $\text{rank}((A'|b')) = \text{rank}(A')$ = the # of non-zero rows of A'.
2. If $Ax = b$ has a solution, then its solution set is $v + t_1u_1 + ... + t_{n-r}u_{n-r}$, where v is a particular solution, $t_i \in \mathbb{F}$, and $\{u_1, ..., u_{n-r}\}$ is a basis for $\ker(L_A)$, where $r = \text{rank}(A)$ and n is the number of columns of A.

Proof. We will only show (1), since (2) follows directly from previous theory (see Prop 2.27). Recall that $Ax = b$ has a solution $\iff$ $A'x = b'$ has a solution $\iff$ $b' \in \text{Im}(L'_A) = \text{Span}(\text{col's of } A')$, $\iff$ $\text{Span}(\text{col's of } A') = \text{Span}(\text{col's of } (A'|b'))$ (Prop 1.4), and this holds IFF $\text{rank}(A') = \text{rank}(A'|b')$. □

As a corollary, we see that $Ax = b$ has a solution $\iff$ the RREF does *not* have a pivot in the last column.

To show that the RREF of a particular matrix A is uniquely determined, we observe the following 2 lemmas:

Lemma 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 any chosen constants $a_i \in \mathbb{F}$, we have

$$a_1A^{(1)} + ... + a_nA^{(n)} = 0 \iff a_1B^{(1)} + ... + a_nB^{(n)} = 0$$

Lemma 2 Let B be the RREF of $A \in M_{m\times n}(\mathbb{F})$. Then:

1. The number of non-zero rows of $B = \text{rank}(B) = \text{rank}(A) =: r$
2. For each $i = 1, ..., r$, denote j_i to be the column of the pivot contained in the i^{th} row. Then $B^{(j_i)} = e_i \in \text{St}_n$. In particular, $\{B^{(j_1)}, ..., B^{(j_r)}\}$ is linearly independent.

3. Each column of B *without* a pivot is in the span of the previous columns.

As a corollary, we get that the RREF of A is unique.

DETERMINATE

Define the *determinate*, notated $\det(A)$, of a square matrix $A \in M_n(\mathbb{F})$, to be a scalar in $\mathbb{F}$ that is $0 \iff A$ is not invertible.

To define this, we note that $A \in M_n(\mathbb{F})$ is invertible $\iff L_A$ is invertible $\iff L_A$ is bijective, which occurs $\iff \ker(L_A) = \{0\}$, i.e. $\text{rank}(L_A) = \text{rank}(A) = n \iff$ the columns of A are linearly independent.

Example: Let $A \in M_3(\mathbb{R})$, and let $A = \begin{pmatrix} - - v_1 - - \\ - - v_2 - - \\ - - v_3 - - \end{pmatrix}$, where $v_i \in \mathbb{R}^3$. We see that, if $\{v_1, v_2, v_3\}$ were linearly dependent, then $\dim(\text{Span}(v_1, v_2, v_3))$ would be at most 2. This is equivalent to saying that (at least) one of v_i lies in the span of the other 2 v_j's. Visually, we have that the parallelepiped composed of v_1, v_2, v_3 has "volume 0." Well, we wanted the determinate to be 0 when A is not invertible, i.e. the rows are linearly dependent. Thus, we may want to generalize the notion of volume to define the determinate.

A function $\delta : M_n(\mathbb{F}) \to \mathbb{F}$ is called (row) *n-linear* (also called an *n-linear form* or *multilinear form*) if it is linear in every row, i.e.

$$\delta \begin{pmatrix} \text{——} & v_1 & \text{——} \\ & \vdots & \\ \text{——} & v_{i-1} & \text{——} \\ \text{——} & cx + y & \text{——} \\ \text{——} & v_{i+1} & \text{——} \\ & \vdots & \\ \text{——} & v_n & \text{——} \end{pmatrix} = c\delta \begin{pmatrix} \text{——} & v_1 & \text{——} \\ & \vdots & \\ \text{——} & v_{i-1} & \text{——} \\ \text{——} & x & \text{——} \\ \text{——} & v_{i+1} & \text{——} \\ & \vdots & \\ \text{——} & v_n & \text{——} \end{pmatrix} + \delta \begin{pmatrix} \text{——} & v_1 & \text{——} \\ & \vdots & \\ \text{——} & v_{i-1} & \text{——} \\ \text{——} & y & \text{——} \\ \text{——} & v_{i+1} & \text{——} \\ & \vdots & \\ \text{——} & v_n & \text{——} \end{pmatrix}$$

where all rows $1, ..., i-1, i+1, ..., n$ remain constant. For example, one can show that $\delta(A) := a_{11} \cdot ... \cdot a_{nn}$ is n-linear, but $\text{tr}(A)$, which *sums* diagonal elements, is not n-linear.

PROPOSITION 2.39

For an n-linear form $\delta : M_n(\mathbb{F}) \to \mathbb{F}$, if a matrix $A \in M_n(\mathbb{F})$ has a zero row, then $\delta(A) = 0$. To show this, observe that $\delta(A) = \delta(A) + \delta(A)$ in this case.

Note that, in the parallelepiped example described above, if two sides are equal, then the volume is 0. This motivates the following definition:

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

Proposition 2.40 Let $\delta : M_n(\mathbb{F}) \to \mathbb{F}$ be an alternating n-linear form. If B is obtained from A by a row operation of type I (i.e. swapping rows), then $\delta(B) = -\delta(A)$.

Proof. For (1), it is enough to show that swapping the first two rows changes the sign of $\delta(A)$. Suppose B is obtained by swapping rows 1 and 2 of A. Then $\delta(A + B) = 0$, since then the first two rows will be equal. On the other hand, $\delta(A + B) = \delta(A) + \delta(B)$ by n-linearity, so in particular $\delta(A) = -\delta(B)$. □

2.17 Existence of the Determinate
There exists a unique alternating linear form δ such that $\delta(I_n) = 1$, and this is called the *determinate*.

Before we prove this, we'll need to study permutations for a minute. For $\pi \in S_n$, the symmetric group, we let $\#\pi$ denote the number of pairs $i, j \in [1, n]$ such that $i < j$ but $\pi(i) > \pi(k)$. Such pairs of i, j are called *inversions*. We say that π is *even* or *odd* if $\#\pi$ is even or odd, respectively. Furthermore, we can express the sign of π as $\operatorname{sgn}(\pi) := (-1)^{\#\pi}$.

Observe now that $\operatorname{sgn} : S_n \to (\{-1, 1\}, \times)$ is a group homomorphism, which is -1 on transpositions, i.e. $\pi \in S_n$ which swap any two elements. In particular, we have

1. $\operatorname{sgn}(\pi^{-1}) = \operatorname{sgn}(\pi)$.
2. If π is a composition of k transpositions, $\tau_1, ..., \tau_k$, then $\operatorname{sgn}(\pi) = (-1)^k$.

For a proof of (1), we observe that $\operatorname{sgn}(\pi^{-1}) = (\operatorname{sgn}(\pi))^{-1} = \operatorname{sgn}(\pi)$. For (2), we have $\operatorname{sgn}(\pi) = \operatorname{sgn}(\tau_1 \cdot ... \cdot \tau_k) = \operatorname{sgn}(\tau_1) \cdot ... \cdot \operatorname{sgn}(\tau_k) = (-1)^k$.

We first note that any alternating multilinear form δ can be written as

$$\delta(A) = \sum_{\pi \in S_n} a_{1\pi(1)} \cdot ... \cdot a_{n\pi(n)} \delta(\pi I_n) \qquad \pi I_n := (e_{\pi(1)}, ..., e_{\pi(n)})$$

and get as a corollary to the properties of sgn described above:

Proposition 2.41 For $\delta : M_n(\mathbb{F}) \to \mathbb{F}$ alternating, $A \in M_n(\mathbb{F})$, we have

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

Proof.

We just need to show $\delta(\pi I_n) = \operatorname{sgn}(\pi)\delta(I_n)$. But $\delta = \tau_1 \cdot ... \cdot \tau_n$, so $(-1)^k = \operatorname{sgn}(\pi)$, and we conclude $\delta(\pi I_n) = (-1)^k \delta(I_n) = \operatorname{sgn}(\pi)\delta(I_n)$. □

We now turn to the main proof of the theorem:

PROOF.

Existence: We write $\det(A) := \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot ... \cdot a_{n\pi(n)}$

Normalized: $\det(I_n) = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot ... \cdot a_{n\pi(n)} = (-1)^0 \cdot \underbrace{1 \cdot ... \cdot 1}_{n \text{ times}}$.

Multilinear: Note that any linear combination of a MLF is a MLF, so we just need to show that $\delta(A) := a_{1\pi(1)} \cdot ... \cdot a_{n\pi(n)}$ for some fixed $\pi \in S_n$ is multilinear, but we have shown this before.

Alternating: Suppose A has $A_{(1)} = A_{(2)}$. We partition S_n into the disjoint union of even and odd permutations, denoting the set of even ones as A_n, and noting that $S_n \setminus A_n = (12) \times A_n = \overline{(12)}$. Thus $\pi' : A_n \to \overline{(12)}$ that takes $\pi \to \pi(12)$ is a bijection, and we write

$$\begin{aligned}
\det(A) &= \sum_{S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot ... \cdot a_{n\pi(n)} \\
&= \sum_{A_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot ... \cdot a_{n\pi(n)} + \underbrace{\operatorname{sgn}(\pi')}_{-\operatorname{sgn}(\pi)} \underbrace{a_{1\pi'(1)}}_{a_{1\pi(2)}} \cdot \underbrace{a_{2\pi'(2)}}_{a_{2\pi(2)}} \cdot ... \cdot a_{n\pi'(n)} \\
&= 0 \quad \square
\end{aligned}$$

We then get the following corollaries: PROPOSITION 2.42

(1) $\det(A) = 0 \iff A$ is non-invertible

(2) $\det(AB) = \det(A)\det(B)$

(3) $\det(A^{-1}) = (\det(A))^{-1}$

(4) $\det(A^t) = \det(A)$

III Diagonalization

We will now begin to study the decomposition of $T : V \to V$ into a direct sum of simpler operators, where $\dim(V) < \infty$. In this chapter, all vector spaces V are finite dimensional, unless otherwise stated. The "simplest" linear operator is scalar multiplication, so, ideally, we'd like to decompose V into a direct sum $V_1 \oplus ... \oplus V_k$ of T-invariant subspaces, where $T_{V_i} : V_i \to V_i$ is scalar multiplication.

For $V_1, ..., V_k \subseteq V$, we say that $\{V_1, ..., V_k\}$ are *linearly independent* if $V_i \cap (V_1 + ... + V_{i-1} + V_{i+1} + ... + V_k) = \{\mathbb{0}\}\ \forall i$. If this is linearly independent, we call $V_1 + ... + V_k$ *direct*, and write $V_1 \oplus ... \oplus V_k$ instead.

Call $T : V \to V$ *diagonalizable* if it admits a diagonalization, i.e. $V = V_1 \oplus ... \oplus V_k$, where $V_i \subseteq V$ are subspaces, and $T|_{V_i}$ defines scalar multiplication by $\lambda_i \in \mathbb{F}$.

Examples: If A is diagonal, i.e. $A = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$, then A is diagonalizable: take $V_i := \text{Span}(e_i)$. Then $\mathbb{F}^n = V_1 \oplus ... \oplus V_n$, and see that $L_A(v) = \lambda_i v\ \forall v \in V_i$.

Now let B be *similar* to the diagonal matrix A as above. Then $B = QAQ^{-1}$ for some $Q \in GL_n(\mathbb{F})$. But all invertible matrices are change of basis matrices, i.e. $[I]^{\beta}_{\alpha}$, where $\alpha := \text{St}_\text{n}$, and $\beta = \{v_1, ..., v_n\}$ is some other basis for $\mathbb{F}^n$. Letting $V_i := \text{Span}(v_i)$, we have $\mathbb{F}^n = V_1 \oplus ... \oplus V_n$ and $L_B(v) = \lambda_i v\ \forall v \in V_i$.

Proposition 3.1 Let $\dim(V) < \infty$. A linear operator $T : V \to V$ is diagonalizable $\iff \exists$ a basis $\beta \subseteq V$ such that $[T]_\beta$ is diagonal.

Proof.

($\implies$) Let $V = V_1 \oplus ... \oplus V_k$ be such that T_{V_i} defined scalar multiplication by λ_i. Let $\beta_i \subseteq V_i$ be a basis for V_i. It is easy to verify that $\beta := \beta_1 \cup ... \cup \beta_k$ is a basis for V. But for each $v \in \beta$, $v \in \beta_i$ for some i, so $T(v) = \lambda_i v$, and $[T(v)]_\beta = \langle 0, ..., 0, \lambda_i, 0, ..., 0\rangle$. Thus, we can order β such that $[T]_\beta$ is diagonal, and we are done.

($\impliedby$) Let $[T]_\beta$ be diagonal with elements $\lambda_1, ..., \lambda_n$ for $\beta = \{v_1, ..., v_n\} \subseteq V$, a basis for V. Then, taking $V_i := \text{Span}(v_i)$, $[T(v)]_\beta = \lambda_i e_i = \lambda[v]_\beta = [\lambda_i v]_\beta\ \forall v \in V_i$. Since $v \to [v]_\beta$ is injective, we conclude that $T(v) = \lambda_i v\ \forall v \in V_i$. Remark also that $\bigoplus_{i=1}^{n} \text{Span}(v_i) = V$ for basis vectors v_i. □

For $T : V \to V$, $\lambda \in \mathbb{F}$, λ is called an *eigenvalue* if T if $\exists v \in V$ such that $T(v) = \lambda v$, where $v \neq \mathbb{0}$. In this event, v is called an *eigenvector* corresponding to λ.

For a finite dimensional vector space V and a linear transformation $T : V \to V$, the following are equivalent: PROPOSITION 3.2

1. T is diagonalizable, i.e. $V = V_1 \oplus ... \oplus V_k$ where $T|_{V_i}$ is scalar multiplication
2. There exists a basis $\beta \subseteq V$ such that $[T]_\beta$ is diagonal
3. There exists a basis $\beta \subseteq V$ containing eigenvectors of T

PROOF.

$2 \implies 3$. Let $\beta := \{v_1, ..., v_n\}$ be a basis for V such that $[T]_\beta = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$. Then $[T(v_j)]_\beta = \lambda_j e_j$, so $T(v_j) = \lambda_j v_j$, and hence v_j is an eigenvector.

$3 \implies 2$. Let $\beta := \{v_1, ..., v_n\}$ be a basis for V where $T(v_j) = \lambda_j v_j$ for $\lambda_j \in \mathbb{F}$. Then $[T(v_j)]_\beta = [\lambda_j v_j]_\beta = \lambda_j e_j$, so in particular $[T]_\beta = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$ □

For $A \in M_n(\mathbb{F})$, A is diagonalizable $\iff$ $\exists Q \in GL_n(\mathbb{F})$ such that $Q^{-1}AQ$ is diagonal. The columns of Q are then eigenvectors which form a basis for $\mathbb{F}^n$. PROPOSITION 3.3

PROOF.

We know that A is diagonalizable $\iff$ there is a basis $\beta \subseteq \mathbb{F}^n$ such that $[L_A]_\beta$ is diagonal. But, letting $\alpha := \mathrm{St_n}$, we have that $A = [L_A]_\alpha = [I]^\alpha_\beta \cdot [L_A]_\beta \cdot [I]^\beta_\alpha = [I]^\alpha_\beta \cdot [L_A]_\beta \cdot ([I]^\alpha_\beta)^{-1}$, so $[L_A]_\beta = ([I]^\alpha_\beta)^{-1} A [I]^\alpha_\beta$, so denoting $Q := [I]^\alpha_\beta$, and we get $Q^{-1}AQ$ is diagonal.

Note that the columns of Q are exactly the vectors in β, which are hence eigenvectors, as previously shown. □

For an eigenvalue $\lambda \in \mathbb{F}$ for a linear operator $T : V \to V$, let

$$\mathrm{Eig}_T(\lambda) := \{v \in V : T(v) = \lambda v\}$$

be called the *eigenspace* of T corresponding to λ. Observe that this is a subspace of V, and that all non-zero vectors in it are exactly the eigenvectors of T corresponding to λ.

It is true that, if T is diagonalizable, then V decomposes into a direct sum $V = \bigoplus_{i=1}^n V_i$ of subspaces of eigenspaces. Does each V_i have to be an eigenspace itself? How many eigenvalues and eigenspaces might T have? Since diagonalizability is conjugate-invariant (i.e. if $A \sim B$ and A is diagonalizable, then so is B), it makes sense to study other conjugation-invariant functions on matrices.

The trace and determinate functions $M_n(\mathbb{F}) \to \mathbb{F}$ are conjugation-invariant. PROPOSITION 3.4

This allows us to contextualize the trace and determinate for transformations. Let V be n-dimensional and $T : V \to V$ be linear. Define $\text{tr}(T)$, the *trace* of T, to be $\text{tr}([T]_\beta)$ for some, or *any*, basis $\beta \subseteq V$. This is well-defined, as we can shift between bases using the change of bases matrices, i.e. $[T]_\beta \sim [T]_\alpha$ for bases α, β. Similarly, the *determinate* of T is $\det([T]_\beta)$ for any basis β, and this is well-defined for the same reasons.

Observe now that T is invertible $\iff \det(T) \neq 0$, where $T : V \to V$ is a linear operator over n-dimensional V.

Proof.

T invertible $\iff [T]_\beta$ is invertible $\iff \det([T]_\beta) \neq 0 \iff \det(T) \neq 0$. □

Proposition 3.5

Let $T : V \to V$ be a linear operator on finite dimensional V. Then

1. $v \in V$ is an eigenvector of T corresponding to $\lambda \in \mathbb{F} \iff v \in \ker(\lambda I - T)$.
2. $\lambda \in \mathbb{F}$ is an eigenvalue of $T \iff \lambda I - T$ is non-invertible, i.e. $\det(T) \neq 0$.

Proof.

$T(v) = \lambda v \iff T(v) - \lambda v = 0 \iff (T - \lambda I)v = 0 \iff v \in \ker(T - \lambda I)$

Now, λ is an eigenvalue $\iff \ker(\lambda I - T) \neq \{0\} \iff \lambda I - T$ is not injective. Since $\lambda I - T$ is a linear operator, $\lambda I - T$ is not injective $\iff \lambda I - T$ is not surjective, i.e. non-invertible. □

As a corollary, we see that $\lambda \in \mathbb{F}$ is an eigenvalue of $A \in M_n(\mathbb{F}) \iff \det(\lambda I_n - A) = 0$. Thus, to find the eigenvalues, we need to find the roots of the function $p(t) = \det(tI_n - A)$. This can be written as $p_A(t) := t^n - \text{tr}(A)t^{n-1} + ... + (-1)^n \det(A)$ (some messy form). We define $p_T(t) = \det(tI - T)$ similarly.

Since we've shown that the eigenvalues of T are precisely the roots of $p_T(t)$, which has degree n, there can be at most n eigenvalues of T for $T : V \to V$ and $\dim(V) = n$.

Proposition 3.6

Let $\lambda_1, ..., \lambda_k$ be distinct eigenvalues for $T : V \to V$. Then, if $v_i \in \text{Eig}_T(\lambda_i)$, $\{v_1, ..., v_k\}$ is linearly independent. In particular, $k \leq n$.

Proof.

We'll show this by induction. This clearly holds for the base case, $k = 1$, since $v_1 \neq 0$ is an eigenvalue. Let $k \to k + 1$, i.e. $\{\lambda_1, ..., \lambda_k\}$ is linearly independent.

We then let $a_i \neq 0$, and write

$$\begin{aligned} & a_1 v_1 + ... + a_k v_k + a_{k+1} v_{k+1} = 0 \\ \implies & T(a_1 v_1 + ... + a_k v_k + a_{k+1} v_{k+1}) = 0 \\ \implies & a_1 T(v_1) + ... + a_k T(v_k) + a_{k+1} T(v_{k+1}) = 0 \\ \implies & \lambda a_1 v_1 + ... + \lambda_k a_k v_k + \lambda_{k+1} a_{k+1} v_{k+1} = 0 \\ \implies & (\lambda_1 - \lambda_{k+1}) a_1 v_1 + ... + (\lambda_k - \lambda_{k+1}) a_k v_k = 0 \end{aligned}$$

By our ind. hyp., $\lambda_i = \lambda_{k+1}\ \forall i$, but we chose λ_i to be distinct, so $\lightning$, and $a_i = 0$ □

PROPOSITION 3.7

For distinct $\lambda_1, ..., \lambda_k$ and $T : V \to V$, $\{\text{Eig}(\lambda_1), ..., \text{Eig}(\lambda_n)\}$ are linearly independent.

PROOF.

Let $v_1 \in \text{Eig}(\lambda_1) \cap (\text{Eig}(\lambda_2) + ... + \text{Eig}(\lambda_k))$. Then $v_1 = v_2 + ... + v_k$ for $v_i \in \text{Eig}(v_i)$ $i \neq 1$. However, if v_i are all nontrivial, then we'd violate Prop. 3.6, and so $v_1 = 0$. □

For an eigenvalue λ for $T : V \to V$, denote by $m_g(\lambda) := \dim(\text{Eig}_T(\lambda))$, and call this the *geometric multiplicity*.

PROPOSITION 3.8

For $T : V \to V$ with eigenvalues $\lambda_1, ..., \lambda_k$, we have

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

PROOF.

$$\sum_{i=1}^{k} m_g(\lambda_i) = \dim\left(\bigoplus_{i=1}^{k} \text{Eig}_T(\lambda_i)\right) \leq n$$ □

3.1 Diagonalizability Condition
$T : V \to V$ is diagonalizable if and only if $\sum_{i \in I} m_g(\lambda_i) = n$, where $\{\lambda_i\}_{i \in I}$ represents all eigenvalues of T.

PROOF.

($\implies$) Recall that $T : V \to V$ is diagonalizable $\iff$ $\exists$ a basis of eigenvectors for V. Let $\beta := \{v_1, ..., v_k\}$ be such a basis. Then $v_i \in \text{Eig}(\lambda_j)$ for some j, so $\beta \subseteq \cup_{i=1}^{k} \text{Eig}(\lambda_i)$. Furthermore, we know that $\beta \cap \text{Eig}(\lambda_j)$ is linearly independent, since β is. Thus, $|\beta \cap \text{Eig}(\lambda_j)| \leq \dim(\text{Eig}(\lambda_j)) = m_g(\lambda_j) \implies n = |\beta| = \sum_{i=1}^{k} |\beta \cap \text{Eig}(\lambda_i)| \leq \sum_{i=1}^{k} m_g(\lambda_i)$.

But we've seen in Prop 3.8 that $n \geq \sum_{i=1}^{k} m_g(\lambda_i)$, so in fact $n = \sum_{i=1}^{k} m_g(\lambda_i)$

($\Longleftarrow$) Suppose $n = \sum_{i=1}^{k} m_g(\lambda_i)$, and let β_i be a basis for $\text{Eig}(\lambda_i)$. By the linear independence of eigenspaces (Prop 3.7), $\beta := \cup_{i=1}^{k} \beta_i$ is linearly independent. This has n elements, so it is a basis for V. In particular, β_i are made up of eigenvectors for T. □

For $T : V \to V$ and an eigenvalue λ, define the *algebraic multiplicity*, denoted $m_a(\lambda)$, to be the largest k such that $(t-\lambda)^k | p_T(t)$, i.e. the multiplicity of λ.

Proposition 3.9 Let $T : V \to V$. For each T-invariant $W \subseteq V$, let $T_W := T|_W : W \to W$. Then p_{T_W} divides $p_T(t)$.

Proposition 3.10 For each eigenvalue λ of $T : V \to V$, $m_g(\lambda) \leq m_a(\lambda)$.

Proof. Let $W := \text{Eig}(\lambda)$. This is T-invariant, since T_W is just scalar multiplication by λ. Thus, by the previous proposition, $p_T(t) = p_{T_W}(t)q(t)$ for some $q(t) \in \mathbb{F}[t]$. We fix a basis $\alpha = \{v_1, ..., v_l\}$ for W, and observe that

$$[T_W]_\alpha = \begin{pmatrix} | & & | \\ \lambda e_1 & \cdots & \lambda e_l \\ | & & | \end{pmatrix} = \begin{pmatrix} \lambda & & \\ & \ddots & \\ & & \lambda \end{pmatrix}$$

Thus, $p_{T_W}(t) = \det(tI_l - [T_W]_\alpha) = \det\begin{pmatrix} t-\lambda & & \\ & \ddots & \\ & & t-\lambda \end{pmatrix} = (t-\lambda)^l$. Hence, $p_T(t) = (t-\lambda)^l q(t)$, so $m_a(\lambda) \geq l = \dim(W) = m_g(\lambda)$. □

For a polynomial $p(t) \in \mathbb{F}[t]$, we say that p *splits over* $\mathbb{F}$ if $p(t) = a(t-r_1)\cdot ... \cdot (t-r_n)$ for $a, r_i \in \mathbb{F}$. For an eigenvalues $\lambda_1, ..., \lambda_k$ of T, we see that the characteristic polynomial $p_T(t)$ splits $\iff \sum_{i=1}^{k} m_a(\lambda_k) = n$. Thus:

3.2 Main Criterion
$T : V \to V$ is diagonalizable if and only if $p_T(t)$ splits and $m_a(\lambda) = m_g(\lambda)$ for each eigenvalue λ of T.

T CYCLIC SPACES

Let V be a vector space, $T : V \to V$ linear, $v \in V$. The *T-cyclic subspace* generated by v is defined to be $\text{Span}(\{v, T(v), ...\}) = \text{Span}(\{T^n(v) : n \in \mathbb{N}\})$. Observe that T-cyclic subspaces are T-invariant.

Lemma (C-H) Let V be s.t. $\dim(V) = k$, $T : V \to V$, $v \in V$. Let W be the T-cyclic subspaces generated by v. Then

(a) $\{v, ..., T^{k-1}\}$ is a basis for W

(b) $T^k(v) = a_0 v + ... + a_{k-1} T^{k-1}(v)$ $a_i \in \mathbb{F}$, so we conclude

$$p_{T_W}(t) = t^k - a_{k-1} t^{k-1} - ... - a_1 t - a_0$$

3.3 Caley-Hamilton Theorem
Let $\dim(V) \leq \infty$, $T : V \to V$ linear. Then

$$p_T(T) = T^n - a_{n-1} T^{n-1} - ... - a_0 I = \mathbb{0}_V$$

i.e. $p_T(T)$ is the zero operator on V.

For $A \in M_n(\mathbb{F})$, $p_A(A) = \mathbb{0}_M$. PROPOSITION 3.11

IV Orthogonality

INNER PRODUCTS

For a vector space V over $\mathbb{F}$, an *inner product* is a binary function $V \times V \to \mathbb{F}$ which sends $(u, v) \to \langle u, v \rangle$. Accordingly, we define an *inner product space* V to be a vector space over $\mathbb{F}$ equipped with the $\langle \cdot, \cdot \rangle$ operation. Typically, and for the remainder of these notes, $\mathbb{F}$ will be understood to be $\mathbb{R}$ or $\mathbb{C}$.

Unless otherwise specified, all sets V are understood to be inner product spaces.

Let $u, v, w \in V$, $\alpha \in \mathbb{F}$. Inner products satisfy the following axioms:

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

(ii) $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle$

(iii) $\langle u, v \rangle = \overline{\langle v, u \rangle}$, where $\overline{\cdot}$ is the complex conjugate

(iv) $\langle u, u \rangle \geq 0$, and $\langle u, u \rangle = 0 \iff u = \mathbb{0}$

Let $\langle \cdot, \cdot \rangle$ be an inner product on V. The *norm* associated with $\langle \cdot, \cdot \rangle$ is defined to be $\|v\| = \sqrt{\langle v, v \rangle}$ for each $v \in V$. We call $v \in V$ a *unit vector* if $\|v\| = 1$. For $v \neq 0$, we call $\|v\|^{-1} v$ the *normalization* of v.

PROPOSITION 4.1

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

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

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

(c) $\|\alpha v\| = |\alpha| \|v\|$

(d) $\langle v, \mathbb{0} \rangle = \langle \mathbb{0}, v \rangle = \mathbb{0}$

(a) and (b) follow from axiom (iii), and one shows (c) by writing $\|\alpha v\|^2 = |\alpha|^2 \|v\|^2$.

Examples:

(a) For $V := \mathbb{F}^n$, the standard $\langle \cdot, \cdot \rangle$ is the *dot product*, which, for $\vec{x}, \vec{y} \in \mathbb{F}^n$, is defined to be $\vec{x} \cdot \vec{y} = \sum_{i=1}^{n} x_i \overline{y_i}$. One shows that $\langle \vec{x}, \vec{y} \rangle := \vec{x} \cdot \vec{y}$ indeed defines an inner product. Its associated norm is thus $\|\vec{x}\| = \sqrt{\sum_{i=1}^{n} |x_i|^2}$, which is the Euclidean norm!

(b) For $\mathbb{F} = \mathbb{R}$ and $V = \mathbb{F}^n$, we have $\vec{x} \cdot \vec{y} = \|x\| \|y\| \cos(\alpha)$, where α is the angle from $\vec{x}$ to $\vec{y}$.

(c) If $\langle \cdot, \cdot \rangle$ is an inner product, then $\langle \cdot, \cdot \rangle_r = r \langle \cdot, \cdot \rangle$ is also a valid inner product.

(d) For $V = C[0,1]$ and $f, g \in V$, we define $\langle f, g\rangle = \int_0^1 f(t)\overline{g(t)}dt$. This is an inner product.

(e) For $A \in M_{m\times n}(\mathbb{F})$, let $\overline{A} = (\overline{a_{ij}})$. $A^* = \overline{A}^t$ is called the *conjugate-transpose* of A. Now, for $V = M_n(\mathbb{F})$, $A, B \in V$, $\langle A, B\rangle := \operatorname{tr}(B^*A)$ is an inner product.

Projections and Cauchy-Schwarz

Let V be an inner product space. Call $u, v \in V$ *orthogonal*, and write $u \perp v$, if $\langle u, v\rangle = 0$. For example, in $\mathbb{R}^3$ under the dot product, $(1, 0, -1) \perp (1, 0, 1)$.

4.1 Pythagoras
Let V be an IPS and $u, v \in V$ be s.t. $u \perp v$. Then

$$\|u\|^2 + \|v\|^2 = \|u + v\|^2$$

and $\|u\|, \|v\| \le \|u + v\|$.

PROOF. $\|u + v\|^2 = \langle u + v, u + v\rangle = \underbrace{\langle u, v\rangle}_{0} + \underbrace{\langle v, u\rangle}_{0} + \langle u, u\rangle + \langle v, v\rangle = \|u\|^2 + \|v\|^2$ □

For $u, v \in V$ with $\|u\| = 1$, define the *projection* of v onto u to be $\operatorname{proj}_u(v) = \langle v, u\rangle u$.

PROPOSITION 4.2 For $u \in V$ with $\|u\| = 1$, $v - \operatorname{proj}_u(v) \perp u \ \forall v \in V$. In particular, $v = \operatorname{proj}_u(v) + w$, where $w := v - \operatorname{proj}_u(v)$ and $w \perp u$.

PROOF. $\langle v - \operatorname{proj}_u(v), u\rangle = \langle v, u\rangle - \langle \operatorname{proj}_u(v), u\rangle = \langle v, u\rangle - \langle v, u\rangle\langle u, u\rangle = \langle v, u\rangle - \langle v, u\rangle$ (the last term $\to 0$). □

For $u \in V$ with $\|u\| = 1$, $\|\operatorname{proj}_u(v)\| \le \|v\| \ \forall v \in V$.

PROOF. $\operatorname{proj}_u(v) \perp w$, where $w = v - \operatorname{proj}_u(v)$, and hence $\|\operatorname{proj}_u(v)\| \le \|\operatorname{proj}_u(v) + w\| = \|v\|$ by Pythagoras, as desired. □

4.2 Cauchy-Schwarz

$$|\langle x, y\rangle| \le \|x\|\|y\| \quad x, y \in V$$

PROOF.

If $\|y\| = 0$, then we are done, so suppose otherwise, and divide both sides. We then need to show $|\left\langle x, \|y\|^{-1} y \right\rangle| \leq \|x\|$, but $|\langle x, u\rangle| = \|\langle x, u\rangle u\| = \|\text{proj}_u(x)\| \leq \|x\|$, where $u := \|y\|^{-1} y$ is a unit vector. □

4.3 Triangle Inequality

$$\|x + y\| \leq \|x\| + \|y\| \quad x, y \in V$$

Proof. (T. Tao) $\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle$ $leq\|x\|^2 + \|y\|^2 + 2|\langle x, y\rangle| \leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2$, so especially $\|x + y\| \leq \|x\| + \|y\|$. □

Examples: For $V = \mathbb{F}^n$ under the dot product, Cauchy-Schwarz says that

$$\left|\sum_{i=1}^{n} x_i y_i\right| \leq \sqrt{\sum_{i=1}^{n} |x_i|^2}\sqrt{\sum_{i=1}^{n} |y_i|^2}$$

For $f, g \in C[0, 1] := V$, we similarly find

$$\int_0^1 f(t)g(t) \leq \sqrt{\int_0^1 |f(t)|^2}\sqrt{\int_0^1 |g(t)|^2}$$

Proposition 4.3 Define $d(u, v) := \|u - v\| = \|v - u\|$ to be the *distance metric* on V, i.e. $d : V \times V \to [0, \infty)$ is s.t.

(i) $d(u, v) \geq 0$, and $d(u, v) = 0 \iff u = v$

(ii) $d(u, v) = d(v, u)$

(iii) $d(u, w) \leq d(u, v) + d(v, w)$

Proof. Axioms (i) and (ii) are trivial to verify. For (iii), note that $d(u, w) = \|u - w\| = \|u - v + v - w\| \leq \|u - v\| + \|v - w\| = d(u, w) + d(v, w)$ (i.e., we use the triangle inequality to prove the triangle inequality!). □

Proposition 4.4 (Parallelogram Law) For $u, v \in V$, we have

(a) $2\|u\|^2 + 2\|v\|^2 = \|u + v\|^2 + \|v - u\|^2$

(b) $\mathfrak{Re}(\langle u, v\rangle) = \frac{1}{2}\left(\|u\|^2 + \|v\|^2 - \|v - u\|^2\right)$

Orthonormality

Call $S \subseteq V$ *orthogonal* if $v \in S$ are pairwise orthogonal to eachother. If $\|v\| = 1 \; \forall v \in S$ and S is orthogonal, we say that S is *orthonormal*.

Orthonormal sets are linearly independent. PROPOSITION 4.5

PROOF. Let S be orthonormal, and $a_1 v_1 + ... + a_n v_n = 0$ for $v_i \in S$, as usual. Then $\langle a_1 v_1 + ... + a_n v_n, v_i \rangle = \langle 0, x_i \rangle = 0$. But, by linearity of the first coordinate of inner products, we write

$$\sum_{j=1}^{n} a_j \left\langle v_j, v_i \right\rangle = a_i \langle v_i, v_i \rangle = a_i \|v_i\|^2 = 0$$

by the orthogonality of S. Since $\|v_i\|^2 = 1$, this means $a_i = 0$. □

A basis for V is called an *orthonormal basis* if it is orthonormal.

Example: For $V = \mathbb{F}^n$, $\mathrm{St_n}$ is orthonormal w.r.t. the dot product. Indeed, $\left\langle e_i, e_j \right\rangle = \delta_{ij}$, and $\|e_i\| = 1$. For $V = \mathbb{F}^4$, see that $\alpha := \{(1,0,1,0), (1,0,-1,0), (0,1,0,1), (0,1,0,-1)\}$ is an orthogonal basis. We normalize each vector (i.e. multiply by $\frac{1}{\sqrt{2}}$) to create an orthonormal basis.

Let $\beta := \{u_1, ..., u_n\}$ be an orthonormal basis for V. Then PROPOSITION 4.5

(a) For every $v \in V$, the coordinates of $v \in \beta$ are $\langle v, u_i \rangle$, i.e.

$$v = \langle v, u_1 \rangle u_1 + ... + \langle v, u_n \rangle u_n = \mathrm{proj}_{u_1}(v_1) + ... + \mathrm{proj}_{u_n}(v_n)$$

These coordinates are called *Fourier coefficients*

(b) For $T : V \to V$, we have $[T]_\beta = \left(\left\langle T(u_i), u_j \right\rangle \right)_{i,j}$, i.e.

$$[T]_\beta = \begin{bmatrix} \langle T(u_1), u_1 \rangle & \cdots & \langle T(u_n), u_1 \rangle \\ \vdots & \ddots & \vdots \\ \langle T(u_1), u_n \rangle & \cdots & \langle T(u_n), u_n \rangle \end{bmatrix}$$

PROOF. For (a), let $v = a_1 u_1 + ... + a_n u_n$ be the representation of $v \in \beta$. Then observe that $\langle v, u_i \rangle = \sum_{j=1}^{n} a_j \left\langle u_j, u_i \right\rangle = \sum_{j=1}^{n} a_j \delta_{ij} = a_i$.

For (b), notice that the j^{th} column of $[T]_\beta$ is

$$[T(u_j)]_\beta = \left(\left\langle T(u_j), u_1\right\rangle, ..., \left\langle T(u_j), u_n\right\rangle\right) \quad \square$$

For $S \subseteq V$, $v \in V$, we say that v is *orthogonal* to S if $v \perp s \ \forall s \in S$. Remark that $v \perp V \iff v = \mathbb{0}$.

Lemma (G-S) Let $\alpha := \{u_1, ..., u_k\}$ be orthonormal. For each $v \in V$, the vector $\text{proj}_\alpha(v) := \sum_{i=1}^n \text{proj}_{u_i}(v) = \sum_{i=k}^n \langle v, u_i\rangle u_i$ has the property that $v - \text{proj}_\alpha(v) \perp \alpha$. In particular, $v - \text{proj}_\alpha(v) \perp \text{proj}_\alpha(v)$. Thus, we can decompose v into $\text{proj}_\alpha(v) + \text{orth}_\alpha(v)$, where $\text{orth}_\alpha(v) := v - \text{proj}_\alpha(v)$, and $\text{proj}_\alpha(v) \perp \text{orth}_\alpha(v)$.

Proof. Let $u_j \in \alpha$. Then

$$\begin{aligned}\left\langle v - \text{proj}_\alpha(v), u_j\right\rangle &= \left\langle v, u_j\right\rangle - \left\langle \text{proj}_\alpha(v), u_j\right\rangle \\ &= \left\langle v, u_j\right\rangle - \left\langle \sum_{i=1}^k \langle v, u_i\rangle u_i, u_j\right\rangle = \left\langle v, u_j\right\rangle - \sum_{i=1}^k \langle v, u_i\rangle \left\langle u_i, u_j\right\rangle \\ &= \left\langle v, u_j\right\rangle - \sum_{i=1}^k \left\langle v, u_j\right\rangle \delta_{ij} = \left\langle v, u_j\right\rangle - \left\langle v, u_j\right\rangle = 0 \quad \square\end{aligned}$$

4.4 Gram-Schmidt Process

There exists and algorithm that takes in an input $\beta := \{v_1, ..., v_k\}$ of linearly independent vectors, and outputs an orthonormal set $\alpha := \{u_1, ..., u_k\}$ such that $\text{Span}(\beta) = \text{Span}(\alpha)$. It's l^{th} step is as follows:

$\{u_1, ..., u_{l-1}\}$ is orthonormal. Then let $v_l' := v_l - \text{proj}_{\{u_1,...,u_{l-1}\}}(v_l)$, and $u_l := \frac{v_l'}{\|v_l'\|}$.

$\implies \{u_1, ..., u_l\}$ is orthonormal.

Proof. By the lemma above, we know that $v_l' := v_l - \text{proj}_{\{u_1,...,u_{l-1}\}}(v_l) \perp \{u_1, ..., u_{l-1}\}$. We normalize v_l' as above to conclude that $\|u_l'\| = 1$ and $u_l \perp \{u_1, ..., u_{l-1}\}$. Thus, $\{u_1, ..., u_l\}$ is orthonormal, as desired. $\square$

Proposition 4.6 As a corollary, we see that a basis $\{v_1, ..., v_n\}$ is turned into an orthonormal basis $\{u_1, ..., u_n\}$. Thus, every finite-dimensional vector spaces has an orthonormal basis.

Proof.

$\{u_1, ..., u_n\}$ is orthonormal and thus linearly independent. Since it has n elements, i.e. $|\{v_1, ..., v_n\}|$, it is a basis. □

For an inner product space V, $S \subseteq V$, the *orthogonal compliment* of S is the subspace $S^\perp := \{v \in V : v \perp S\} = \{v \in V : v \perp s \ \forall s \in S\}$.

It is easy to check that $S^\perp$ is a subspace of V (even if S isn't): $\langle v + \alpha w, s\rangle = \langle v, s\rangle + \alpha \langle w, s\rangle = 0 \implies v + \alpha w \in S^\perp$.

Previously, we denoted by $S^\perp$ the annihilator of S, i.e. $S^\perp \subseteq V^*$. It is tempting to call our new definition an abuse of notation. However, this is standard, as (later we'll see) $\{v \in V : v \perp S\}$ and $\{f \in V^* : f|_S = 0\}$ are analogs in many respects.

4.5 Orthogonal Decomposition of V
Let V be an inner product space, and let $W \subseteq V$ be finite-dimensional.

(a) For $v \in V$, there exists a unique decomposition $v = w + w_\perp$, where $w \in W$, $w_\perp \in W^\perp$. We call w the *orthogonal projection* of v onto W, and write $w = \text{proj}_W(v)$.

(b) $V = W \oplus W^\perp$. In particular, if $\dim(V) < \infty$, then $\dim(V) = \dim(W) + \dim(W^\perp)$.

PROOF.

For (a), let $\alpha := \{w_1, ..., w_k\}$ be an orthonormal basis for W, and let $w = \text{proj}_\alpha(v)$. Then $w_\perp := v - w$ is orthogonal to α, and hence orthogonal to $\text{Span}(\alpha) = W$. *Existence //*

Suppose $w + w_\perp = v = w' + w'_\perp$. Since $v - w$ and $v - w' \in W^\perp$, so does their difference $\implies v - w - v + w' = w' - w \in W^\perp$, as $W^\perp$ is a subspace. But $w' - w \in W$ as well, so $w' - w = 0 \implies w' = w$. Since $v = w' + w'_\perp = w + w_\perp$, we conclude $w'_\perp = w_\perp$ as well. *Uniqueness //*

For (b), we already know $V = W + W'$ by (a). It remains to show that $W \cap W' = \{0\}$, but clearly if $w \in W, w \in W'$, then $w = 0$. □

As an immediate corollary, we see that if α, β are two orthonormal basis for finite-dimensional $W \subseteq V$, then $\text{proj}_\alpha(v) = \text{proj}_\beta(v)\ \forall v \in V$, since $\text{proj}_W(v)$ is unique.

4.6 $\text{proj}_W(v)$ minimizes distance to W
For finite-dimensional $W \subseteq V$ and some $v \in V$, $\text{proj}_W(v)$ is the *unique*, closest vector to $v \in W$.

PROPOSITION 4.7

(a) $\text{proj}_W(v) : V \to V$ is a linear operator.

(b) A linear operator $T : V \to V$ is a projection onto $\text{Im}(T) \iff \ker(T) = \text{Im}(T)^\perp$.

(c) For an inner product space V and $W \subseteq V$, $(W^\perp)^\perp = W$.

PROOF. We'll only show part (c). By definition, we know $W \subseteq (W^\perp)^\perp$. Now let $v \in (W^\perp)^\perp$. Then $v = w + w_\perp$ for some $w \in W$, $w_\perp \in W^\perp$ by Thm 4.5. Then

$$\|v\|^2 = \langle v, v\rangle = \langle v, w + w_\perp\rangle = \langle v, w\rangle + \cancelto{0}{\langle v, w_\perp\rangle}$$

$$= \langle v, w\rangle = \langle w_\perp + w, w\rangle = \cancelto{0}{\langle w_\perp, w\rangle} + \langle w, w\rangle = \langle w, w\rangle = \|w\|^2$$

By Pythagoras, $\|v\|^2 = \|w\|^2 + \|w_\perp\|^2$, so $\|w_\perp\|^2 = 0 \implies w_\perp = 0$, so $v = w$, and especially $v \in W$. □

Let V be an inner product space. For $w \in V$, we define the linear functional $f_w \in V^*$ s.t. $f_w(v) = \langle v, w\rangle$. One verifies that this is linear.

4.7 Riesz Representation Theorem
Let V be finite-dimensional. Then for each $f \in V^*$ $\exists$ a unique $w \in V$ s.t. $f = f_w$, i.e. $f(v) = \langle v, w\rangle$ $\forall v \in V$.

PROOF. *Existence:* fix $f \in V^*$, and let $\beta := \{v_1, ..., v_n\}$ be an orthonormal basis for V. For each $v \in V$, we know that

$$v = \langle v, v_1\rangle v_1 + ... + \langle v, v_n\rangle v_n$$

Hence, $f(v) = \langle v, v_1\rangle f(v_1) + ... + \langle v, v_n\rangle f(v_n) = \left\langle v, \overline{f(v_1)}v_1 + ... + \overline{f(v_n)}v_n\right\rangle$.

Independence: Suppose $f_{w_1} = f_{w_2} = f$. Then $f_{w_1 - w_2} = f_{w_1} - f_{w_2} = \mathbb{0}$, so $f_{w_1 - w_2} = \langle v, w_1 - w_2\rangle = \mathbb{0} \implies w_1 - w_2 = 0 \implies w_1 = w_2$. □

4.8 Adjoint Existence
Let $\dim(V) < \infty$. For $T : V \to V$, $\exists$ a unique linear operator $T^* : V \to V$, called the *adjoint*, such that $\langle T(v), w\rangle = \langle v, T^*(w)\rangle$ $\forall v, w \in V$.

PROOF. For $w \in V$, define $\tilde{f_w} \in V^*$ as $\tilde{f_w}(v) = \langle T(v), w\rangle$. This is indeed a linear functional on V. By Riesz representation, $\exists$ a unique $\tilde{w} \in V$ s.t. $\tilde{f_w} = f_{\tilde{w}}$, i.e. $\tilde{f_w}(v) = \langle T(v), w\rangle = \langle v, \tilde{w}\rangle$ $\forall v \in V$. Set $T^*(w) = \tilde{w}$. Then T^* meets our condition, and it remains to show that T^* is linear.

For $w_1, w_2 \in V$, $a \in \mathbb{F}$, $T^*(aw_1 + w_2)$ is the unique vector s.t. $\langle T(v), aw_1 + w_2\rangle = \langle v, T^*(aw_1 + w_2)\rangle$. It is enough to show that $aT^*(w_1) + T^*(w_2)$ also satisfies our condition. But $\langle T(v), aw_1 + w_2\rangle = \overline{a}\langle T(v), w_1\rangle + \langle T(v), w_2\rangle = \overline{a}\langle v, T^*(w_1)\rangle +$

$\langle v, T^*(w_2)\rangle = \langle v, aT^*(w_1) + T^*(w_2)\rangle$. □

Let $T : V \to V$ be a linear operator on a finite-dimensional V. Let β be an orthonormal basis for V. Then PROPOSITION 4.8

$$[T^*]_\beta = [T]_\beta^*$$

where $A^* = \overline{A}^t$ (conjugate-transpose, sometimes called the *adjoint* of A). Furthermore, for $A \in M_n(\mathbb{F})$, $L_A^* = L_{A^*}$.

PROOF.

Recall that, for $[T^*]_\beta$, where $\beta = \{v_1, ..., v_n\}$, the $(ij)^{th}$ entry is $\left\langle T^*(v_j), v_i \right\rangle = \overline{\left\langle v_i, T^*(v_j) \right\rangle} = \overline{\left\langle T(v_i), v_j \right\rangle}$, which is the conjugate of the $(ji)^{th}$ entry of $[T]_\beta$. Hence, $[T^*]_\beta = \overline{[T]_\beta^t} = [T]_\beta^*$, as desired.

For the last statement, let β be the standard orthonormal basis. Then $[L_A^*]_\beta$ is $B \in M_n(\mathbb{F})$ with $L_B = L_A^*$. By (a), $B = [L_A]_\beta^* = A^*$. □

We observe the following properties of the adjoint: PROPOSITION 4.9

(a) The function which sends $T \to T^*$ is conjugate-linear.

  (a) $(T_1 + T_2)^* = T_1^* + T_2^*$
  (b) $(aT)^* = \overline{a}T^*$

(b) $(T_1 \circ T_2)^* = T_1^* \circ T_2^*$

(c) $I_V^* = I_V$

(d) If T is invertible, then so is T^*, and $(T^*)^{-1} = (T^{-1})^*$.

Let $T : V \to V$ be a linear operator and $\dim(V) < \infty$. Then $\mathrm{Im}(T^*) = \ker(T)^\perp$ and $\ker(T^*) = \mathrm{Im}(T)^\perp$. PROPOSITION 4.10

For $T : V \to V$, $\dim(V) < \infty$, we have $\mathrm{null}(T) = \mathrm{null}(T^*)$ and $\mathrm{rank}(T) = \mathrm{rank}(T^*)$. PROPOSITION 4.11

PROOF.

$\mathrm{rank}(T^*) = \dim(\mathrm{Im}(T^*)) = \dim(\ker(T)^\perp) = n - \dim(\ker(T)) = \mathrm{rank}(T)$

$\mathrm{null}(T^*) = \dim(\ker(T^*)) = \dim(\mathrm{Im}(T)^\perp) = n - \dim(\mathrm{Im}(T)) = \mathrm{null}(T)$ □

Let $T : V \to V$ be a linear operator for $\dim(V) < \infty$. For $\lambda \in \mathbb{F}$, λ is an eigenvalue of $T \iff \overline{\lambda}$ is an eigenvalue of T^*. PROPOSITION 4.12

PROOF.

λ is an eigenvalue of $T \iff \text{null}(T-\lambda I) > 0 \iff \text{null}(T^* - \lambda I) > 0 \iff \lambda$ is an eigenvalue of T^*. □

4.9 Schur's Lemma
Let $T : V \to V$ be a linear operator, $\dim(V) < \infty$. Then, if $p_T(t)$ splits, $\exists$ an orthonormal basis β for V such that $[T]_\beta$ is upper triangular.

Proof.

Since $p_T(t)$ splits, T and T^* have eigenvalues. We will show the above by induction on $\dim(V) =: n$. Let $n = 1$. Then clearly $[T]_\beta$ is upper triangular.

$n \to n+1$: Let λ be an eigenvalue of T^* and its corresponding eigenvector be v_n, i.e. $T^*(v_n) = \lambda v_n$. Let $W := \text{Span}(v_n)$. Then $W^\perp$ is T-invariant: indeed, if $v \perp W$, then $w \perp v_n$, i.e. $\langle v, v_n\rangle = 0$. Then $\langle T(v), v_n\rangle = \langle v, T^*(v_n)\rangle = \langle v, \lambda v_n\rangle = \overline{\lambda}\langle v, v_n\rangle = 0$, so $T(v) \perp v_n$.

$\dim(W^\perp) = n - \dim(W) = n-1$, and $T_{W^\perp} : W^\perp \to W^\perp$, so by induction, $\exists$ an orthonormal basis $\alpha := \{v_1, ..., v_{n-1}\}$ of $W^\perp$ s.t. $[T_{W^\perp}]_\alpha$ is upper triangular.

Let $\beta := \{v_1, ..., v_n\}$. This is an orthonormal basis for V, since $V = W \oplus W^\perp$. Notice also that

$$[T]_\beta = \begin{bmatrix} | & & | & | \\ [T_{W^\perp}(v_1)]_\alpha & \cdots & [T_{W^\perp}(v_{n-1})]_\alpha & [T_{W^\perp}(v_n)]_\alpha \\ | & & | & | \\ 0 & & 0 & \end{bmatrix}$$

is upper triangular, since the first $n-1$ columns are upper triangular by assumption. □

A linear operator $T : V \to V$, where $\dim(V) < \infty$, is called *normal* if T and T^* commute, i.e. $T \circ T^* = T^* \circ T$. T is called *self-adjoint* if $T = T^*$.

Examples:

1. Orthogonal projections are self-adjoint. Let $W \subseteq V$, and p be the projection onto W. Fix $u, v \in V$. Then $u = p(u) + u_\perp$ and $v = p(v) + v_\perp$. Then observe that $\langle p(u), v\rangle = \langle p(u), p(v) + v_\perp\rangle = \langle p(u), p(v)\rangle + \cancel{\langle p(u), v_\perp\rangle} = \langle p(u), p(v)\rangle$.

   Similarly, we find that $\langle u, p(v)\rangle = \langle p(u), p(v)\rangle$, and so $\langle p(u), v\rangle = \langle u, p(v)\rangle$, i.e. $p = p^*$, as desired.

2. If $p : V \to V$ is an orthogonal projection, and $\lambda \in \mathbb{C} \setminus \mathbb{R}$, then $(\lambda p)^* = \overline{\lambda} p \neq \lambda p$, so λp is *not* self-adjoint. However, λp is still normal: $(\lambda p) \circ (\lambda p)^* = (\lambda p) \circ (\overline{\lambda} p) = \lambda^2 p^2 = (\overline{p}) \circ (\lambda p) = (\lambda p)^*(\lambda p)$.

3. Let $V = W_1 \oplus ... W_k$, where $W_i \perp W_j \ \forall i \neq j$. Then $\forall \lambda_1, ..., \lambda_k \in \mathbb{F}$, $T := \lambda_1 \text{proj}_{W_1} + ... + \lambda_k \text{proj}_{W_k}$ is normal.

Let $T : V \to V$ be normal, and $\dim(V) < \infty$. Then PROPOSITION 4.13

(a) $\|Tv\| = \|T^* v\| \ \forall v$

(b) $p(T)$ for any polynomial p is normal

(c) $\forall v \in V$, v is an eigenvector of T corresponding to $\lambda \iff v$ is an eigenvector of T^* corresponding to $\overline{\lambda}$

(d) For eigenvalues $\lambda_1 \neq \lambda_2$, $\text{Eig}(\lambda_1) \perp \text{Eig}(\lambda_2)$

A basis of V consisting of eigenvectors of T is called an *eigenbasis* of T.

4.10 Diagonalizability of Normal Operators on $\mathbb{C}$
Let $T : V \to V$ be a linear operator over a finite-dimensional inner product space V, where $\mathbb{F} := \mathbb{C}$. Then T is normal $\iff$ $\exists$ an orthonormal eigenbasis of T.

PROOF.

($\implies$) Suppose $T \circ T^* = T^* \circ T$. Then T and T^* have eigenvalues (since p_T splits over $\mathbb{C}$), and they are the same. As in Schur's Lemma, let v_n be an eigenvalue of T/T^*. By putting $W := \text{Span}(v_n)$, we have that $W^\perp$ is T-invariant, so one writes $T_{W^\perp} : W^\perp \to W^\perp$. Furthermore, one can show that $T_{W^\perp}$ is normal if T is normal. Then our induction hypothesis has that $\beta := \{v_1, ..., v_{n-1}\}$ is an orthonormal eigenbasis for $W^\perp \implies \beta' := \{v_1, ..., v_n\}$ is an orthonormal eigenbasis for V, since $V = W \oplus W^\perp$.

($\impliedby$) Let β be an orthonormal eigenbasis for T. Then $[T]_\beta$ is diagonal (each v_i would map to $e_i \lambda_i$). Then $[T^*]_\beta = [T]_\beta^*$ is also diagonal. Thus, since diagonal matrices commute, $[T \circ T^*]_\beta = [T]_\beta [T^*]_\beta = [T^*]_\beta [T]_\beta = [T^* \circ T]_\beta$, and, as I_β is a linear isomorphism, this mean $T^* \circ T = T \circ T^*$, as desired □

The eigenvalues of self-adjoint operators are always real. PROPOSITION 4.13

PROOF.

Let T be self-adjoint, and λ be an eigenvalue. Then $T(v) = \lambda v$. Since T is normal, $T^*(v) = \overline{\lambda} v$. But $T^* = T$, so $\overline{\lambda} = \lambda$, i.e. $\lambda \in \mathbb{R}$. □

Characteristic polynomials of real, symmetric matrices split over $\mathbb{R}$. PROPOSITION 4.14

PROOF.

Let $A \in M_n(\mathbb{R})$ and $A = A^t$. Thus, $A = A^*$, since conjugations are irrelevant in $\mathbb{R}$. Let $L_A : \mathbb{C}^n \to \mathbb{C}^n$ by $L_A(v) = Av$, as usual. Then $L_A^* = L_{A^*} = L_A$, so L_A is self-adjoint. Thus, p_{L_A} splits over $\mathbb{C}$. But the roots of p_{L_A} are all real, so p_{L_A} splits over $\mathbb{R}$! Then we just observe that $p_{L_A} = p_{[L_A]} = p_A$. □

As an immediate corollary, we see that, for self-adjoint $T : V \to V$ over $\mathbb{R}$, p_T splits over $\mathbb{R}$. As proof, we let β be an orthonormal basis. Then $A := [T]_\beta$ is real. Furthermore, $A^t = A^* = [T]_\beta^* = [T^*]_\beta = [T]_\beta$, so $A = A^t$. From the proposition above, $p_A = p_{[T]_\beta} = p_T$ splits over $\mathbb{R}$.

4.11 Diagonalizability of Self-Adjoint Operators on $\mathbb{R}$
Let $T : V \to V$ be a linear operator, where $\mathbb{F} = \mathbb{R}$. Then T is self-adjoint $\iff$ $\exists$ an orthonormal eigenbasis for T.

Proof.

($\implies$) Let $T : V \to V$ be self-adjoint. Then p_T splits over $\mathbb{R}$ by the above. Then, by Schur's Lemma, $\exists$ an orthonormal basis β for V such that $[T]_\beta$ is upper triangular. Thus, $[T]_\beta^t$ is lower triangular. But $[T]_\beta^t = [T]_\beta^* = [T^*]_\beta = [T]_\beta$, so $[T]_\beta$ is actually *diagonal*, i.e. β is made up of eigenvectors.

($\impliedby$) Suppose $\exists$ an orthonormal eigenbasis β for T. Then $[T^*]_\beta$ is real and diagonal, and hence $[T^*]_\beta = [T]_\beta^t = [T]_\beta$, so $T = T^*$, as desired. □

The following theorem is the finale, in effect, of this course:

4.12 Spectral Theorem
Let V be finite-dimensional inner product space, and $T : V \to V$ be a linear operator. If $\mathbb{F} = \mathbb{C}$, let T be normal. If $\mathbb{F} = \mathbb{R}$, let T be self-adjoint. Then T admits a unique (up to reorderings) spectral decomposition, i.e.

$$T = \lambda_1 p_1 + ... + \lambda_k p_k$$

where p_i are orthogonal projections such that $I_V = p_1 + ... + p_k$, and $p_i \circ p_j = \delta_{ij} p_j$. In other words, $V = \bigoplus_{i=1}^n \mathrm{Im}(p_i)$, with $\mathrm{Im}(p_i) \perp \mathrm{Im}(p_j)\ \forall i \neq j$.

Proof of Spectral Theorem

Lemma 1

Let V be finite-dimensional, and $p_1, ..., p_k$ be orthogonal projections on V. TFAE:

1. $p_i \circ p_j = \mathbb{0}\ \forall i \neq j$ and $I_V = p_1 + ... + p_k$
2. $\mathrm{Im}(p_i) \perp \mathrm{Im}(p_j)$ and $\bigoplus_{i=1}^n \mathrm{Im}(p_i)$

Proof.

For $i \neq j$, $p_i \circ p_j = 0 \iff p_i \circ p_j(v) = \{0\}\ \forall v \iff p_i(\text{Im}(p_j)) = \{0\} \iff \text{Im}(p_j) \subseteq \ker(p_i) = \text{Im}(p_i)^\perp \iff \text{Im}(p_i) \perp \text{Im}(p_j)$ //

($\implies$) $I_V = p_1 + ... + p_k \iff v = p_1 v + ... + p_k v\ \forall v \implies V = \bigoplus_{i=1}^n \text{Im}(p_i)$, since $\text{Im}(p_i) \perp \text{Im}(p_j) \implies \text{Im}(p_i) \cap \text{Im}(p_j) = \{0\}$.

($\impliedby$) Take $v \in \bigoplus_{i=1}^n \text{Im}(p_i)$. Then $v = w_1 + ... + w_k$, where $w_i \in \text{Im}(p_i)$, so $p_i(v) = p_i(w_1) + ... + p_i(w_k) = \sum_{j=1}^k \delta_{ij} w_j = w_i$, hence $v = p_1(v_1) + ... + p_k(v_k)$, i.e. $I_V = p_1 + ... + p_k$. □

Spectral Decomposition via Eigenvalues:

LEMMA 2 Let $T : V \to V$, $\dim(V) < \infty$, $p_1, ..., p_k : V \to V$ be orthogonal projections, and $\lambda_1, ..., \lambda_k \in \mathbb{F}$. Then TFAE:

1. $T = \lambda_1 p_1 + ... + \lambda_k p_k$ is a spectral decomposition.
2. $\{\lambda_1, ..., \lambda_k\}$ are all distinct eigenvalues of T **AND** $\text{Im}(p_i) = \text{Eig}_T(\lambda_i)$ **AND** $\text{Eig}_T(\lambda_i) \perp \text{Eig}_T(\lambda_j)\ \forall i \neq j$ **AND** $V = \bigoplus_{i=1}^k \text{Eig}_T(\lambda_i)$.

PROOF. (1 $\implies$ 2) Denote $W_i = \text{Im}(p_i)$ and remark that $W_i \subseteq \text{Eig}_T(\lambda_i)$. Indeed, if $w_i \in W_i$, then $T(w_i) = \lambda_i p_i(w_i) = \lambda_i w_i$, as $p_i(w_j) = 0\ \forall i \neq j$. We can also write $V = \bigoplus_{i=1}^n \text{Im}(p_i)$, so $\sum \dim(W_i) = \dim(V) = n$. Since $W_i \subseteq \text{Eig}_T(\lambda_i)$, this means $\sum \dim(\text{Eig}(\lambda_i)) \geq n$, i.e. $= n$, as well. We conclude that $W_i = \text{Eig}(\lambda_i)$. Furthermore, since $\sum m_g(\lambda_i) = n$, we conclude that these are *all* the eigenvalues of T.

(2 $\implies$ 1) Suppose $p_i \circ p_j = 0\ \forall i \neq j$ and $V = \bigoplus_{i=1}^k \text{Eig}_T(\lambda_i)$. Since $\bigoplus_{i=1}^k \text{Im}(p_i)$ as well, we have $I_V = p_1 + ... + p_k$ by Lemma 1.

Since $v = w_1 + ... + w_k\ \forall v$, where $w_i \in \text{Eig}(\lambda_i)$, we have $p_i(v) = w_i\ \forall i \neq j$, since $\text{Eig}(\lambda_i) \perp \text{Eig}(\lambda_j)$. Then $T(v) = T(w_1) + ... + T(w_k) = \lambda_1 w_1 + ... + \lambda_k w_k = \lambda_1 p_1(v) + ... + \lambda_k p_k(v)$, i.e. $T = \lambda_1 p_1 + ... + \lambda_k p_k$ □

We have thus proven uniqueness of the spectral decomposition, since $\lambda_1, ..., \lambda_k$ must all be the unique eigenvalues of T, and $\text{Im}(p_i) = \text{Eig}_T(\lambda_i)$.

As for existence, we have shown, for normal T, $\text{Eig}(\lambda_i) \perp \text{Eig}(\lambda_j)\ \forall i \neq j$. And, if ($\mathbb{F} = \mathbb{C}$) or ($\mathbb{F} = \mathbb{R} \wedge T = T^*$), then T admits and orthonormal eigenbasis, i.e. $V = \bigoplus_{i=1}^k \text{Eig}_T(\lambda_i)$. Then the conditions for Lemma 2 kick in, and hence $T = \lambda_1 p_1 + ... + \lambda_k p_k$. *Remark:* The set $\{\lambda_1, ..., \lambda_k\}$ of $T : V \to V$ is called the *spectrum* of T.