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

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

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.

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.

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) $\{f \in C[\mathbb{R}]$ 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 VS 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 $\operatorname{Span}(\varnothing) = \mathbb{0}_V$.

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 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 $\mathrm{Span}(S)=\{(x,y,z):x+y+z=0\}$.

Proposition 1.3 Let V be a VS over a field $\mathbb{F}$, and S be some subspace of it. Then $\mathrm{Span}(S)$ 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 $\mathrm{Span}(S)$ is closed under these operations. Furthermore, $\mathbb{0}_V \in \mathrm{Span}(S)$ by taking a trivial combination of vectors $\implies \mathrm{Span}(S)$ 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. $U \subseteq \mathrm{Span}(S)$ □

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

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

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

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

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

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

$$\mathrm{St_n} := \{(1,\underbrace{0,0,\ldots,0}_{n-1 \text{ times}}),(0,1,0,\ldots,0),\ldots,(0,\ldots,0,0,1)\}$$

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

LINEAR (IN)DEPENDENCE

Let V be a VS 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 VS 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 $v = a_1 v_1 + \ldots + a_n v_n$, where $v_i \in S \setminus v$. It follows that $\mathbb{0}_V = a_1 v_1 + \ldots + a_n v_n - v$. As $-1 \neq 0$, this is non-trivial, and we are done. $\square$

COROLLARY

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 $\text{Span}(S) \neq \text{Span}(S \setminus v)$.

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

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. $av + a_1 v_1 + \ldots + a_n v_n = \mathbb{0}_V$ 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 $v = -a^{-1} a_1 v_1 - \ldots - a^{-1} a_n v_n$, and conclude that $v \in \text{Span}(S)$. Then $V \subseteq \text{Span}(S)$. Clearly, $\text{Span}(S) \subseteq V$, so we conclude that $\text{Span}(S) = V$. $\square$

BASES

1.1 Characterization of a Basis

Let V be a VS 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. $\square$

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 $\mathrm{St_n}$, 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. $\left\{\sum_{n\in\mathbb{F}} a_nt^n : a_n \in \mathbb{F}\right\}$. 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.

$$\begin{tikzcd}
 & \{1,2,3\} \\
\{0\} & \{1,2\} \arrow[u] \\
 & \varnothing \arrow[ul] \arrow[u]
\end{tikzcd} \qquad \text{where } \nearrow \text{ means } \subseteq$$

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:

$$S_1 \subseteq S_2 \subseteq \ldots \subseteq S_n$$

Thus, $v_1, \ldots, v_n \in S_n$, and since $S_n \in C \subseteq \mathcal{I}$ (recall the definition of $\mathcal{I}$), S_n is linearly independent. Thus, $\{v_1, \ldots, v_n\} \subseteq S_n$ 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. $\square$

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 $\mathrm{Span}(B) = V$, or $W = V$. The ($\impliedby$) direction is trivial. $\square$

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 $T : W \to V$ 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 $v = a_1 v_1 + \ldots + a_n v_n$ be the unique representation of v in B. By linearity, $T(v) = a_1 T(v_1) + \ldots + a_n T(v_n) = a_1 w_1 + \ldots + a_n w_n$ 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 linear transformation $\hat{T} : V \to W$.

PROOF.

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

This is essentially the infinite case of theorem 1.2:

Existence: Let $T : \beta \to \gamma$ be (any) function between the bases of V and W. For $v \in V$, let $v = a_1 v_1 + \ldots + a_n v_n$ be its unique representation in β, where $v_i \in \beta$. Define the function

$$\hat{T}(v) = a_1 T(v_1) + \ldots + a_n T(v_n)$$

We'll show that this is linear. Let $x, y \in V$. Without loss of generality, we can write $x = a_1 v_1 + \ldots + a_m v_m$ and $y = b_1 v_1 + \ldots + b_m v_m$ as their unique representations, where a_i, b_i *may* be zero. We thus have

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

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 $a_1 v_1 + \ldots + a_n v_n$ be its unique representation in β.

$$\hat{T}(v) = a_1 T(v_1) + \ldots + a_n T(v_n) = a_1 \tilde{T}(v_1) + \ldots + a_n \tilde{T}(v_n) = \tilde{T}(a_1 v_1 + \ldots + a_n v_n) = \tilde{T}(v)$$

□

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 $V \cong W$.

PROPOSITION 2.1

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

PROOF.

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

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.

PROOF.

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.

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. $\square$

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 $a_1w_1 + a_2w_2 \in \mathrm{Im}(T)$. $\square$

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. $\square$

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 $a_i = 0$ □

PROPOSITION 2.5

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

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 $\dim(V) = \dim(W)$. □

PROPOSITION 2.6

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.

PROOF.

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

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

PROOF.

This follows directly from the 1^{st} 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 $\dim(\mathrm{Im}(T)) = n - k$, then the theorem follows.

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

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

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 $\mathrm{Im}(T)$.

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

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

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. $\operatorname{rank}(T) = n$

PROOF.

T surjective $\iff$ $\operatorname{rank}(T) = \dim(\operatorname{Im}(T)) = \dim(W) = n$
T injective $\implies$ $\operatorname{null}(T) = 0$, so $\operatorname{rank}(T) = \dim(V) = n$
$\operatorname{rank}(T) = n \implies \operatorname{null}(T) = 0$, 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 $\operatorname{Im}(T)$ through the isomorphism $\overline{v} \to T(v)$, where $\overline{v} := v + \ker(T)$ (as in quotient groups).

PROOF.

We know that $\hat{T} : V/\ker(T) \to \operatorname{Im}(T)$ 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 $\operatorname{Hom}(V, W)$ to be the set of all linear transformations from $V \to W$.

PROPOSITION 2.8 $\operatorname{Hom}(V, W)$ 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 $T_0, T_1 \in \operatorname{Hom}(V, W)$.

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:

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

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

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

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 $w_i \in W$.

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

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

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

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

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

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.

PROPOSITION 2.11

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

PROOF.

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

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

$a[T] = [aT]$ 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 $a_1 v_1 + ... + a_n v_n = v$ 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

$$I_\beta(v) = I_\beta(a_1 v_1 + ... + a_n v_n) = a_1 I(v_1) + ... + a_n I(v_n) = a_1 e_1 + ... + a_n e_n = [v]_\beta$$

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:

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

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

$$[T]_\beta^\gamma \cdot [v]_\beta = a_1[T(v_1)]_\gamma + ... + a_n[T(v_n)]_\gamma$$
$$= [a_1T(v_1) + ... + a_nT(v_n)]_\gamma \qquad \text{by linearity of } I_\gamma$$
$$= [T(a_1v_1 + ... + a_nv_n)]_\gamma \qquad \text{by linearity of} T$$
$$= [T(v)]_\gamma \qquad \text{and we are done} \quad \square$$

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

Sometimes called a two-sided 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. $\sum\limits_0^\infty a_n t^t \to \sum\limits_1^\infty a_n t^{n-1}$. This has a right inverse, namely the right shift map of coefficients, $\sum\limits_0^\infty a_n t^n \to \sum\limits_0^\infty a_n t^{n+1}$. 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 $\qquad$ T has a left inverse $\qquad$ T has an inverse

PROOF.

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

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 $B = A^{-1}$.

PROPOSITION 2.17

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 $T(\ker(T)) = \mathbb{0}_V \in \ker(T)$

Proof to come

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.

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).

c.f. Prop. 2.11, 2.14

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 $L_A \cong A$.

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

PROPOSITION 2.19

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

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

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

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

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

Let $U := \ker(T^N)$ and $W := \operatorname{Im}(T^N)$. 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: $T(\operatorname{Im}(T^N)) = \operatorname{Im}(T^{N+1}) = \operatorname{Im}(T^N)$ 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 $v = \{\mathbb{0}\}$.

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

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. $\operatorname{Hom}(V, \mathbb{F})$, is denoted V^*, and is called the *dual space* of V.

For finite dimensional V, we already know that $\dim(V^*) = \dim(\mathrm{Hom}(V, \mathbb{F})) = \dim(V) \cdot \dim(\mathbb{F}) = \dim(V)$. In accordance with our construction of a basis for Hom (pp. 19-20), we let $\beta := \{v_1, \ldots, v_n\}$ be a basis for V and $\gamma = \{1\}$ be the standard basis for $\mathbb{F}$. Then $\beta^* := \{f_1, \ldots, f_n\}$ is a basis for $\mathrm{Hom}(V, \mathbb{F}) = V^*$, 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 $\beta^* := \{f_1, \ldots, f_n\}$, 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 $A_i = [0, \ldots, 0, 1, 0, \ldots, 0]$, 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 $f \to f(x)$. Note that $x \in V^{**}$ by this definition. 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.

PROPOSITION 2.21

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

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. $S^{\perp} = (\mathrm{Span}(S))^{\perp}$

PROOFS.

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 $S^{\perp} \subseteq (\mathrm{Span}(S))^{\perp}$. The converse holds by (2). □

For $S \subseteq V$, we denote $\hat{S} := \{\hat{x} : x \in S\} \subseteq V^{**}$. 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 $(U^{\perp})^{\perp} = \hat{U}$.

PROOF.

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

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 2.23 The transpose has the following properties:

1. $T^t : W^* \to V^*$ is linear
2. $\ker(T^t) = (\mathrm{Im}(T))^\perp$
3. $\mathrm{Im}(T^t) \subseteq (\ker(T))^\perp$, with equality IFF V, W are finite dimensional. In that event, $\dim(\mathrm{Im}(T)) = \dim(\mathrm{Im}(T^t))$.
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), $g \in \ker(T^t) \iff T^t(g) = \mathbb{0} \iff T^t(g)(v) = 0 \ \forall v \in V \iff g(T(v)) = 0 \iff g(w) = 0 \ \forall w \in \mathrm{Im}(T) \iff g \in (\mathrm{Im}(T))^\perp$

For (3), fix $f \in \mathrm{Im}(T^t)$ 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 $\mathrm{Im}(T^t) \subseteq (\ker(T))^\perp$.

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 $B_{ji} = A_{ij}$.

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

□

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

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 $\iff$ $(\mathrm{Im}(T))^\perp = \{0\}$ by part (2) of Prop 2.23. This implies that $\mathrm{Im}(T) = W$, i.e. T is surjective, by Duality (i.e. Prop 2.22). Conversely, if $\mathrm{Im}(T) = W$, then the function which takes all of W to 0 is precisely $\mathbb{0}_{W^*}$, i.e. $(\mathrm{Im}(T))^\perp = 0$. 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 $\mathrm{Im}(T^t) = V^*$. Thus T^t is surjective. Conversely, if $\mathrm{Im}(T^t) = V^*$, 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 $\dim(\mathrm{Im}(T))$. Furthermore, if $\beta = \{v_1, ..., v_n\}$ is a basis for V, then $\mathrm{Im}(T) = \mathrm{Span}(\{T(v_1), ..., T(v_n)\})$. In particular, $\dim(\mathrm{Im}(T))) \leq n$, where $\dim(V) = n$ (see dimension theorem). Thus, we can express $\dim(\mathrm{Im}(T))$ 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 $\mathrm{rank}(A)$, or the *rank* of A, by $\mathrm{rank}(\mathrm{Im}(L_A))$.

Define also the *column rank* of A, denoted 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 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 $\mathrm{rank}(A) = \text{c-rank}(A)$, and this follows from the definitions.

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

Proof.

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

We then conclude that $\text{c-rank}(A) = \text{r-rank}(A) = \mathrm{rank}(A)$ 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:

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

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 $x = \langle x_1, \dots, x_n \rangle$ and $b = \langle b_1, \dots, b_m \rangle \in \mathbb{F}^m$. 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 $b \in \mathrm{Im}(L_A) = \mathrm{Span}(\{A^{(1)}, \dots, A^{(n)}\})$. 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 $b \in \mathrm{Im}(L_A)$, 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. $A \cdot v = b$.

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 $a \in \ker(L_A)$. $\square$

PROPOSITION 2.27

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

PROOF.

$(L_A) = n - \mathrm{rank}(L_A) = n - \dim(\mathrm{Im}(L_A)) > n - m > 0$, so $\ker(L_A)$ is non-trivial. $\square$