Notes for Math 110: Linear Algebra
UC Berkeley Fall 2019

Aditya Sengupta

December 9, 2019

Contents

Math 110: Linear Algebra **Fall 2019**

Lecture 1: Introduction

Lecturer: Edward Frenkel *28 August* *Aditya Sengupta*

Note: *LaTeX format adapted from template for lecture notes from CS 267, Applications of Parallel Computing, UC Berkeley EECS department.*

1.1 Vector spaces and linear transformations

These are the main objects in linear algebra.

Definition 1.1. *A vector space over a field F is a set $\mathbb{V}$ with two operations,*

- *addition: given $x, y \in V$, we get $x + y \in V$.*
- *scalar multiplication: given $c \in F, x \in V$, we get $cx \in V$.*

satisfying the following axioms,

VS1 $\forall x, y \in V, x + y = y + x$

VS2 $\forall x, y, z \in V, (x + y) + z = x + (y + z)$

VS3 $\exists 0 \in V \text{ s.t. } \forall x \in V, 0 + x = x$

VS4 $\forall x \in V, \exists y \in V \text{ s.t. } x + y = 0$

VS5 $\forall x \in V, 1 \cdot x = x$

VS6 $\forall x \in V, a, b \in F, a(bx) = (ab)x$

VS7 $\forall x, y \in V, a \in F, a(x + y) = ax + ay$

VS8 $\forall x \in V, a, b \in F, (a + b)x = ax + bx$

Example 1.1. The canonical example of a vector space is $V = \mathbb{R}^n = \{ \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \mid a_i \in \mathbb{R}\}$. The corresponding field is $F = \mathbb{R}$.

We can verify that this satisfies the vector space axioms. For example, it satisfies VS1:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix} \tag{1.1}$$

VS2:

$$c \in \mathbb{R}, x = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}^T \implies c \cdot x = \begin{bmatrix} ca_1 & ca_2 & \dots & ca_n \end{bmatrix}^T \tag{1.2}$$

VS3: define $\underline{0} = \begin{bmatrix} 0 & 0 & \dots & 0 \end{bmatrix}$. VS4: given $x = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}^T$, choose $y = \begin{bmatrix} -a_1 & -a_2 & \dots & -a_n \end{bmatrix}^T$.

□

Proposition 1.1. *Given a vector space V over a field F such that $\underline{0}$ and $\underline{0'}$ both satisfy VS3, $\underline{0} = \underline{0'}$.*

Proof. By VS3, $\forall x \in V, x + \underline{0'} = x$ and also $x = x + \underline{0'}$, so we can choose $x = \underline{0}$. Then we get $\underline{0} = \underline{0} + \underline{0'} = \underline{0'} + \underline{0}$ by VS1. Then, we apply VS3 to $\underline{0}$, choosing $x = \underline{0'}$. This gives us $\underline{0'} + \underline{0} = \underline{0'}$, therefore $\underline{0} = \underline{0'}$. □

Math 110: Linear Algebra **Fall 2019**

Lecture 2: Vector space isomorphisms, subspaces

Lecturer: Edward Frenkel *4 September* *Aditya Sengupta*

Fix a field F. Then, we can define a vector space over F as a set with two operators $(+, \cdot)$ satisfying 8 axioms. Now, we can derive some properties of vector spaces from these axioms. We do this by looking at subsets of a given vector space.

Example 2.1. Fix a point in $\mathbb{R}^2$ and consider the set of all vectors whose tails are at that point. Every point in the plane is associated with a corresponding vector in this set. Let $F = \mathbb{R}$ and define two operations on these vectors:

1. scalar multiplication; for $c \in \mathbb{R}$, $c\vec{v}$ points in the same direction as $\vec{v}$ if $c > 0$, and in the opposite direction if $c < 0$. The vector has a length of $|c||\vec{v}|$.
2. addition: translate the tail of one vector to the tip of the other, forming a parallelogram. The vector sum is the vector from the fixed point to the new tip of the translated vector.

Denote this vector space Vec_2. There is a one-to-one mapping between a vector $\vec{v_1} \in Vec_2$ and $\begin{bmatrix} x & y \end{bmatrix}^T \in \mathbb{R}^2$. This mapping is called an *isomorphism*.

□

Definition 2.1. *Let V be a vector space over F and $S \subset V$ a subset. We say that S is closed under $+$ if $\forall x, y \in S, x + y \in S$. We say that S is closed under $\cdot$ if $\forall x \in S, c \in F, c \cdot x \in S$.*

Example 2.2. Consider the following subsets of Vec_2:

$$\mathcal{S} = \{\vec{v_1}, \vec{v_2}, \vec{v_1} + \vec{v_2}\} \tag{2.1}$$
$$\mathcal{L} = \{c \cdot \vec{v_1} \mid c \in \mathbb{R}\} \tag{2.2}$$

$\mathcal{S}$ is not closed under addition, but $\mathcal{L}$ is. $\mathcal{L}$ inherits two operations from Vec_2, i.e. addition and multiplication are well defined on $\mathcal{L} \subset Vec_2$. We want to check if it is a vector space. If it turns out that it does satisfy the axioms of a vector space, we will call it a subspace of our vector space.

□

However, we do not have to verify that every axiom is satisfied; it is sufficient to check only one of the axioms, provided that closure (which we just checked above) is satisfied.

Theorem 2.1. *Let V be a vector space over F, and W is a subset of V. Then W is a subspace of V if and only if the following are true:*

1. *W is closed under addition.*
2. *W is closed under scalar multiplication.*
3. *The zero vector of V is in W.*

An 'if and only if' statement means that an implication goes both ways: A if and only if B means $A \implies B$ and $B \implies A$.

Before we prove the theorem, we introduce a lemma and prove it.

Lemma 2.2. *Let V be a vector space over F and let $x, y, z \in V$ such that $x + z = y + z$. Then $x = y$.*

Proof. (of the lemma) We start with

$$x + z = y + z \tag{2.3}$$

We know that $\exists - z \in V$, therefore we add that to both sides:

$$(x + z) + (-z) = (y + z) + (-z) \tag{2.4}$$

Then, we use associativity to regroup this:

$$x + (z + (-z)) = y + (z + (-z)) \tag{2.5}$$

Simplifying, we get

$$x + \underline{0}_V = y + \underline{0}_V \tag{2.6}$$

$$x = y \tag{2.7}$$

□

Proof. (of the theorem) In the forward direction, we let W be a subspace of V and derive the required properties. (1) and (2) follow from the definition of a subspace. Since W is a vector space, according to (VS 3) there is an element $\underline{0}_W \in W$ such that $x + \underline{0}_W = x \ \forall x \in W$. Let $x = \underline{0}_W$:

$$\underline{0}_W + \underline{0}_W = \underline{0}_W \tag{2.8}$$

Also, these are all elements of V, so

$$\underline{0}_W + \underline{0}_V = \underline{0}_W \tag{2.9}$$

Addition commutes in vector spaces, so we get

$$\underline{0}_V + \underline{0}_W = \underline{0}_W \tag{2.10}$$

Also, in $\underline{0}_W$, we can add the zero element, so we get

$$\underline{0}_V + \underline{0}_W = \underline{0}_W + \underline{0}_W \tag{2.11}$$

Finally, we apply the lemma to get

$$\underline{0}_V = \underline{0}_W. \tag{2.12}$$

The backward direction was (I'm not lying) left as an exercise for the reader. $\square$

Example 2.3. Let W be the set of all continuous functions $[0, 1] \to \mathbb{R}$. Prove that this set is a vector space.

Proof. Given a set T and a field F, let $\mathcal{F}(T, F)$ be the set of all functions $f : T \to F$. Define $+$ and $\cdot$ pointwise by

$$(f + g)(t) = f(t) + g(t) \tag{2.13}$$

$$(cf)(t) = c \cdot f(t) \tag{2.14}$$

It can be verified (handwavy again) that this is a vector space. Then, $C([0, 1]) \subset \mathcal{F}(T, F)$. If we can show that $C([0, 1])$ is closed under the two operations (it is) and that the zero function is in it (it is), we can show that it is a subspace and therefore a vector space. $\square$

$\square$

Math 110: Linear Algebra **Fall 2019**

Lecture 3: Span, linear dependence

Lecturer: Edward Frenkel *9 September* *Aditya Sengupta*

tbd: bring tikz back in from notes when you've recovered from the shame

3.1 Linear combinations

Even though the vector space Vec_2 is infinite, it turns out that we can generate it from just two vectors. Call them $\vec{u_1}$ and $\vec{u_2}$, and suppose they do not point in the same direction. By the property of scalar multiplication, we can create $a_1\vec{u_1}, a_1 \in \mathbb{R}$ and similarly $a_2\vec{u_2}, a_2 \in \mathbb{R}$; by the property of superposition, we can sum them to obtain an arbitrary vector $a_1\vec{u_1} + a_2\vec{u_2} \in Vec_2$. Therefore, finitely many elements of the vector space are sufficient to generate the entire infinite space.

$$test \tag{3.1}$$

The expression representing an arbitrary element of the vector space is called a *linear combination* of $\vec{u_1}$ and $\vec{u_2}$.

Definition 3.1. *Let V be a vector space over F and S a nonempty subset of V. An element $v \in V$ is called a linear combination of elements of S if there is a finite subset $\{u_1, \dots, u_n\} \subseteq S$ and a collection $a_1, \dots, a_n \in F$ such that*

$$v = a_1u_1 + a_2u_2 + \cdots + a_nu_n = \sum_{i=1}^{n} a_iu_i \tag{3.2}$$

Note that we have not technically defined vector addition for more than two elements, but because of associativity, we can iteratively compute pairwise sums, and the result is guaranteed to be in the vector space.

Remark 3.1. *S can be finite or infinite. However, we will only consider finite linear combinations.*

Example 3.1. Every vector in Vec_2 is a linear combination of elements of a subset $S = \{\vec{u_1}, \vec{u_2}\}$. □

3.2 Span

Definition 3.2. *Given a nonempty subset S in a vector space V over F, we define* $\text{span}(S)$ *as the set of all linear combinations of finitely many elements $u_1, \dots, u_n \in S$. We say that V is generated by S if* $\text{span}(S) = V$.

Example 3.2. Consider

$$V = F^n = \{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mid x_i \in F\} \tag{3.3}$$

and the finite n-dimensional subset,

$$S = \{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \} \tag{3.4}$$

Then,

$$\text{span}(S) = \{a_1u_1 + a_2u_2 + \cdots + a_nu_n\} = \{ \begin{bmatrix} a_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ a_2 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ a_n \end{bmatrix} \} = \{ \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \} \tag{3.5}$$

□

This allows us to reduce general problems about vector spaces to systems of linear equations.

Example 3.3. Define $v = \begin{bmatrix} -2 & 0 & 3 \end{bmatrix}^T, u_1 = \begin{bmatrix} 1 & 3 & 0 \end{bmatrix}^T, u_2 = \begin{bmatrix} 2 & 4 & -1 \end{bmatrix}$. Is v a linear combination of u_1 and u_2?

This would be the case if we could find $a_1, a_2 \in \mathbb{R}$ such that $v = a_1u_1 + a_2u_2$.

$$\begin{bmatrix} -2 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} a_1 \\ 3a_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 2a_2 \\ 4a_2 \\ -a_2 \end{bmatrix} = \begin{bmatrix} a_1 + 2a_2 \\ 3a_1 + 4a_2 \\ -a_2 \end{bmatrix} \tag{3.6}$$

This is a system of 3 linear equations in 2 variables, which can in general be solved (will either be overdetermined or inconsistent).

$$a_1 + 2a_2 = 2 \tag{3.7}$$
$$3a_1 + 4a_2 = 0 \tag{3.8}$$
$$0a_1 - a_2 = 3 \tag{3.9}$$

This gives us $a_2 = -3$ and from that, we can substitute back into either of the other equations and get $a_1 = 4$.

□

Proposition 3.2. *Let V be a vector space over F and S a nonempty subset of V. Then* $\operatorname{span}(S)$ *is a subspace of V.*

Proof. We know that $W \subset V$ is a subspace if and only if it is closed under addition and scalar multiplication, and $\underline{0}_V \in W$, so it suffices to show that these hold for $\operatorname{span}(S)$. Consider $v_1, v_2 \in \operatorname{span}(S) \subset V$. We know that there exist a_i, b_j such that

$$v_1 = \sum_{i=1}^{n} a_i u_i \tag{3.10}$$
$$v_2 = \sum_{i=1}^{m} b_j w_j \tag{3.11}$$

for $U = \{u_i\}_{i=1}^n$ and $W = \{w_j\}_{j=1}^m$ finite subsets of S. Then, consider the set $\{u_1, \ldots, u_n, w_1, \ldots, w_m\}$ and the corresponding scalars

$$c_k = \begin{cases} a_k & 1 \le k \le n \\ b_{k-n} & k > n \end{cases} \tag{3.12}$$

Similarly, $\operatorname{span}(S)$ is closed under scalar multiplication by choosing $a_i' = c \cdot a_i$ for each element of the finite subset $\{u_i\}_{i=1}^r$ to scalar multiply any element of the span of S keeping it within S.

Finally, we can show that $\underline{0}_V \in \operatorname{span}(S)$. Take any $u_1 \in S$, and take $v = 0 \cdot u_1 = \underline{0}_V$, so $\underline{0}_V \in \operatorname{span}(V)$. □

Proposition 3.3. *If W is a subspace of V containing S, then W contains* $\operatorname{span}(S)$.

Remark 3.4. *This means that* $\operatorname{span}(S)$ *is the smallest subspace of V containing S.*

Consider the span of the empty set, $\operatorname{span}(\varnothing)$. To figure out what this should be, we view it as a subset of some vector space V. We define it to be $\{\underline{0}_V\}$. With this definition, we can establish a surjection between the subsets of V and the subspaces of V, namely $S \subset V \to \operatorname{span}(S)$. This is not complete unless we set this as the definition for what happens to the empty set under this bijection.

3.3 Linear Independence

Definition 3.3. *Let V be a vector space over a field F, and let S be a subset of V. S is called linearly dependent if there exists a finite subset $\{u_1, \dots, u_n\}$ of S and $a_1, \dots, a_n \in F$ not all zero such that*

$$a_1 u_1 + a_2 u_2 + \cdots + a_n u_n = \underline{0}_V \tag{3.13}$$

This essentially means that we can obtain the zero vector in more than one way (beyond the trivial solution.)

Example 3.4. Let $S = \{u_1, u_2, u_3\}$, where $u_3 = u_1 + u_2$. Then, if we select $a_1 = a_2 = 1, a_3 = -1$, we get the following linear combination:

$$1 \cdot u_1 + 1 \cdot u_2 + (-1) \cdot u_3 = (u_1 + u_2) + (-u_3) = \underline{0} \tag{3.14}$$

Since there is a nontrivial linear combination that is equal to zero, we conclude that S is linearly dependent.

□

Conversely, start with the nontrivial linear combination $\sum_j a_j u_j = 0$. We know that $\exists i$ such that $a_i \neq 0$, so $\exists a_i^{-1} \in F$ as it is nonzero. We first rearrange the linear combination:

$$-a_i u_i = \sum_{j=1, j\neq i}^{n} a_j u_j \tag{3.15}$$

Then, we left multiply by $-a_i^{-1}$, to get

$$u_i = \sum_{j=1, j\neq i}^{n} (-a_i^{-1} \cdot a_j) u_j \tag{3.16}$$

Math 110: Linear Algebra **Fall 2019**

Lecture 4: Linear independence, bases

Lecturer: Edward Frenkel *11 September* *Aditya Sengupta*

4.1 Linear Independence

If $S \subset V$ does not satisfy the property of linear dependence, then S is called linearly independent.

Equivalently, S is linearly independent if for any finite subset of distinct elements $\{u_1, \dots, u_n\}$, the equation

$$a_1 u_1 + \cdots + a_n u_n = \underline{0}_V, a_i \in F \tag{4.1}$$

has only one solution, i.e. all $a_i = 0 \in F$.

Example 4.1. Consider $V = \mathbb{R}^3 = \{\begin{bmatrix} x_1 & x_2 & x_3 \mid x_i \in \mathbb{R}\end{bmatrix}\}$, and $S = \{\begin{bmatrix}1 & 0 & 0\end{bmatrix}^\top, \begin{bmatrix}0 & 1 & 0\end{bmatrix}^\top, \begin{bmatrix}0 & 0 & 1\end{bmatrix}^\top\}$. We want to verify that S is linearly independent, i.e. that $a_1 u_1 + a_2 u_2 + a_3 u_3 = \begin{bmatrix}0 & 0 & 0\end{bmatrix}^\top$ has only one solution.

The sum of the three vectors in S scaled by a_1, a_2, a_3 respectively is

$$\begin{bmatrix} a_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ a_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.2}$$

This can only be satisfied if all three a_is are zero. Therefore S is linearly independent.

□

Example 4.2. Consider the set

$$S_1 = \{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \} \tag{4.3}$$

Checking if this set is linearly independent is equivalent to solving the system of equations

$$
\begin{aligned}
a_1 + a_2 \quad & \quad +2a_3 = 0 && (4.4)\\
+ a_2 \quad & \quad +a_3 = 0 && (4.5)\\
-a_1 + a_2 \quad & \quad = 0 && (4.6)\\
& && (4.7)
\end{aligned}
$$

Adding the first and third equations gives us $2a_2 + 2a_3 = 0$, which from the second and third equations gives us $a_1 = a_2 = -a_3$. Therefore, if we take an arbitrary a_3, that constrains a_1 and a_2. This gives us a family of nontrivial solutions. Therefore, the set is linearly dependent.

□

In the above example, the system of equations was homogeneous, meaning all the right-hand-side constant terms were zero. In general, if we have a homogeneous system, and $(a_1, \ldots, a_n)$ is a solution, then $\forall k \in F$, $(ka_1, \ldots, ka_n)$ is a solution as well.

Proposition 4.1. *Suppose we have a homogeneous system of linear equations on $a_1, \ldots, a_n$. Then the set of solutions is a subspace of F^n.*

The set of solutions of an inhomogeneous system is not a subspace, because its solutions cannot be scaled to obtain new solutions.

Example 4.3. Suppose M is a square upper triangular matrix over $\mathbb{R}$ with nonzero diagonal entries. We want to show that the set of columns of M is linearly independent.

To illustrate the problem, we can pick a particular M and show its columns are linearly independent.

$$
M = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix} \tag{4.8}
$$

This has a corresponding set of equations,

$$
\begin{aligned}
a_1 + a_2 \quad & \quad +2a_3 = 0 && (4.9)\\
+ a_2 \quad & \quad +a_3 = 0 && (4.10)\\
& \quad 2a_3 = 0 && (4.11)
\end{aligned}
$$

Therefore, working from bottom to top to eliminate variables, we can show that $a_3 = 0 \implies a_2 = 0 \implies a_1 = 0$. Therefore the columns are linearly independent.

As long as the coefficients on the a_is are nonzero (which is guaranteed by a nonzero diagonal), this procedure works for any n. Therefore the only solution of $a_1u_1 +$

$\cdots + a_n u_n = \underline{0}_V$ is the trivial solution, so the set is linearly independent. $\square$

4.2 Bases

We've established the notions of a generating subset of a vector space V over a field F, and a linearly independent subset. Now, we can combine them into the idea of a *basis* of V.

Definition 4.1. *A subset S of V is called a basis of V if it generates V, i.e. if* $\operatorname{span}(S) = V$*, and if it is linearly independent.*

A basis is the minimal subset of a vector space that generates it. A basis is not unique, but any basis of a vector space has the same size. This size is called the dimension of the vector space.

Example 4.4. In $\mathbb{R}^n$, the standard basis is

$$\beta = \{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \} \tag{4.12}$$

We previously saw that $\operatorname{span}(\beta) = \mathbb{R}^n$. Now, we will show that β is linearly independent. $\square$

Lemma 4.2. *The set β_M of columns of a square upper-triangular matrix M with nonzero diagonal entries is also a basis of $\mathbb{R}^n$.*

Proof. We showed above that β_M is linearly independent. We can show that $\operatorname{span}(\beta_M) = \mathbb{R}^n$.

Recall that the system of linear equations generated by M is

$$M_{11}a_1 + M_{12}a_2 + \dots \qquad + M_{1n}a_n = x_1 \tag{4.13}$$
$$+ M_{22}a_2 + \dots \qquad + M_{2n}a_n = x_2 \tag{4.14}$$
$$\vdots \tag{4.15}$$
$$+ M_{nn}a_n = x_n \tag{4.16}$$

Working backwards, we can show that $a_n = \frac{x_n}{M_{nn}}$, which gives us a unique value for a_{n-1}, which in turn gives us a unique value for a_{n-2}, and so on. Therefore there exists a unique $\vec{a} = \{a_1, \dots, a_n\}$ such that $M\vec{a} = \vec{x}$ for any $\vec{x} \in \mathbb{R}^n$. $\square$

Proposition 4.3. *Let V be a vector space over F. A subset $\beta \subset V$ is a basis of V if and only if all $v \in V$ can be expressed uniquely as a linear combination of elements of β.*

Proof. In the forward direction, we assume β is a basis of V and we show that every $v \in V$ can be expressed as a linear combination of elements of β. Further, we show that this combination is unique.

Since $\operatorname{span}(\beta) = V$, every $v \in V$ can be expressed as a combination of elements of β. To show this expression is unique, we suppose there are two such expressions,

$$v = a_1 u_1 + \cdots + a_n u_n \tag{4.17}$$
$$v = b_1 u_1 + \cdots + b_n u_n \tag{4.18}$$

where $a_i, b_i \in F$. We subtract these two and use distributivity to collect terms:

$$\underline{0}_V = (a_1 - b_1)u_1 + \cdots + (a_n - b_n)u_n \tag{4.19}$$

Since β is a basis, it is linearly independent, therefore $a_i - b_i = 0$ for all i, so $a_i = b_i$. $\square$

Math 110: Linear Algebra **Fall 2019**

Lecture 5: Dimension, Replacement Theorem

Lecturer: Edward Frenkel *16 September* *Aditya Sengupta*

Suppose a vector space V over F has a finite basis $\beta = \{x_1, \dots, x_n\}$. Any other basis of V has the same number of elements, n. This is referred to as the dimension on V. The crucial step in proving this is the replacement theorem.

Theorem 5.1. *Let V be a vector space over F. Let $G \subset V$ be a subset of V of cardinality n that generates V. Further, let $L \subset V$ be another subset, of cardinality m, which is linearly independent. Then:*

1. $m \leq n$
2. $\exists H \subset G, |H| = n - m$ *such that* $L \cup H$ *generates* V.

The book proof is by induction on m, which is weird because part of what we're trying to prove is that m is bounded above.

As an alternative, we state two theorems and prove them more directly.

Theorem 5.2. *Suppose $V, G, L, |G| = n, |L| = m$ are defined as above, and suppose that $m \leq n$. Then $\exists H$ defined as above.*

Theorem 5.3. *Suppose $V, G, L, |G| = n, |L| = m$ are defined as above. Then $m \leq n$.*

Proof of Theorem **??**. We prove this by induction on $m \geq 0$. For $m = 0$, L has cardinality 0, so it must be the empty set, so $H \subset G$ must be G. Then $L \cup H = \varnothing \cup G = G$. This generates V.

Next, we suppose that Theorem **??** is true for $|L| = k$, and we show it is true for $|L| = k+1$. Let $L = \{v_i\}_{i=1}^{k+1}$. Say $|G| = n$ and G generates V. Suppose $G = \{u_i\}_{i=1}^{n}$. By our inductive hypothesis, Theorem **??** is true for $m = k$, so there exists $H \subset G, |H| = n - k$ such that $L \cup H$ generates V. Call it $H = \{w\}_{i=1}^{n-m}$.

Now, we take $L \cup H = \{v_1, \dots, v_{m+1}, w_1, \dots, w_{n-k}\}$. We claim that this has one degree of linear dependence, i.e. we can choose a w_i that is a linear combination of the others, and remove it to get a set spanning V.

We add v_{k+1} to L. Since $L \cup H$ spans V, v_{m+1} can be written as a linear combination of its elements:

$$v_{k+1} = a_1 v_1 + \cdots + a_k v_k + b_1 w_1 + \cdots + b_{n-k} w_{n-k} \tag{5.1}$$

We claim that at least one of the b_is is nonzero. We prove this by contradiction; if this is not the case, then all b_is are 0, then we get v_{k+1} as a linear combination of the v_is, which contradicts L being linearly independent. So at least one b_i is nonzero. Without loss of generality suppose $b_1 \neq 0$. Then, Equation **??** implies

$$w_1 = (-b_1^{-1} a_1) v_1 + \cdots + (-b_1^{-1} a_k) v_k + b_1^{-1} v_{k+1} + (-b_1^{-1} b_2) w_2 + \dots (-b_1^{-1} b_{n-k}) u_{n-k} \tag{5.2}$$

Therefore, $w_1 \in \operatorname{span}\{v_1, \dots, v_m, w_2, \dots, w_{n-k}\}$ and so we can drop w_1. Therefore, $L \cup H$ has cardinality n and generates V by the inductive hypothesis. $\square$

The proof above is based on knowing that $n - m > 0$, so that there exists at least one b_i that can be shown to be nonzero and so can be inverted to create a linear dependence. We now prove this.

Proof of Theorem **??**. Suppose this is not the case. Then there exists a linearly independent subset $L \subset V, |L| = m > n$. Choose $L' \subset L$ with n elements. It is still linearly independent. Apply Theorem **??**, so there exists an H such that $L' \cup H$ generates V. But $|L'| = n$, so $|H| = 0$, i.e. $H = \varnothing$. Therefore L' generates V. Let v_{n+1} be an element of L such that $v_{n+1} \notin L'$. But since L' generates V, v_{n+1} can be written as a linear combination of elements of L'. This means L is linearly dependent. This is a contradiction, therefore $m \leq n$. □

Corollary 5.4. *Let V be a vector space over F, and β be a basis of V. Suppose β is finite and $|\beta| = n$. Then any other basis of V also has n elements.*

Proof of Corollary **??**. Let γ be another basis of V and let $|\gamma| = m$. Take γ equivalent to L in Theorem **??**, and take β equivalent to G. By Theorem **??**, we get $m \leq n$. But if we take γ equivalent to G and β equivalent to L, we get $n \geq m$. Therefore $m = n$. □

Now, we know that every basis has an equal size, which allows us to introduce the notion of dimension.

Definition 5.1. *Let V be a vector space over F.*

1. *Suppose that V has a finite basis β. Then V is called finite-dimensional, and we say that the dimension is equal to $|\beta|$.*
2. *Otherwise, we say that V is infinite-dimensional.*

Remark 5.5. *The same set V may be a vector space over different fields. Depending on which field we choose, V may have different dimensions.*

Example 5.1. Let $V = \mathbb{C}$ over $F = \mathbb{C}$. The dimension of V is 1. More generally, if $V = \mathbb{C}^n$, its dimension over $\mathbb{C}$ is n. But we can also view $\mathbb{C}$ as a vector space over $\mathbb{R}$, with the same addition but scalar multiplication only by reals. This has dimension 2, with basis $\{1, i\}$. If $V = \mathbb{C}^n$, its dimension over $\mathbb{R}$ is $2n$.

□

Corollary 5.6. *Let V be a vector space over F and let $\dim V = n$. Any generating subset of V has at least n elements, and any linearly independent subset of V has no more than n elements. If a generating subset has n elements, then it is a basis; if a linearly independent subset has n elements, then it is a basis.*

Math 110: Linear Algebra **Fall 2019**

Lecture 6: Linear Transformations

Lecturer: Edward Frenkel *18 September* *Aditya Sengupta*

6.1 Maps Between Vector Spaces

So far, we've discussed an individual vector space, which is a set V with an associated field F such that two operations are defined: addition, $+ : V \times V \to V$, and scalar multiplication, $\cdot : F \times V \to V$, and such that they satisfy certain axioms. We'll now look at interactions between different vector spaces, specifically maps (functions) $V \to W$.

In set theory, we have a notion of a map or function between sets A and B, $f : A \to B$. This is a rule assigning an element of B to every element of A.

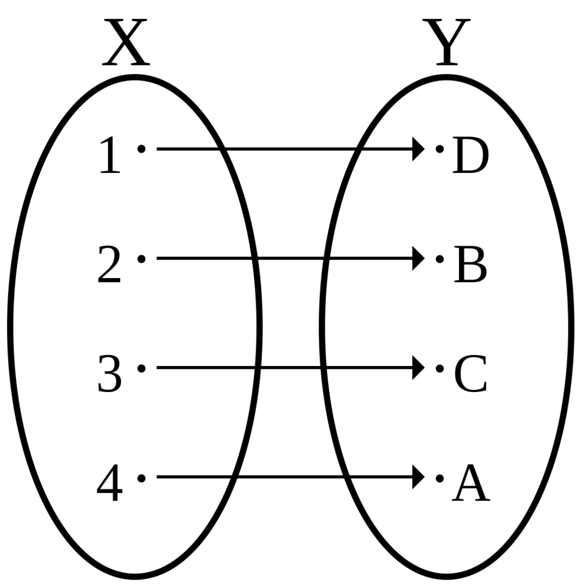


Figure 6.1: A bijection (a special kind of function) between sets X and Y.

Definition 6.1. *A function $f : A \to B$ is well defined if for every $a \in A$, there is a single specific element $b \in B$ such that $f(a) = b$.*

Figure **??** shows a specific kind of mapping called a bijection, in which the mapping is well defined and satisfies two additional properties, being injective (one-to-one) and surjective (onto). However, to be a mapping, it suffices to be well defined according to Definition **??**.

6.2 Linear Transformations

We don't want to define just any function between V and W; instead, we want to deal with maps that respect the structure of V and W. Specifically,

1. V and W should be over the same field F.
2. The map has to be compatible with $+, \cdot$.

Let's call this map $T : V_{+,\cdot} \to W_{+,\cdot}$. Then we can express this notion of respecting structure in more mathematical terms:

$$\forall x, y \in V, T(x + y) = T(x) + T(y) \tag{6.1}$$

We view $T(x + y), T(x), T(y)$ as elements of W and require that they satisfy superposition. This means T is compatible with the addition operation. Similarly,

$$\forall x \in V, c \in F, T(c \cdot x) = c \cdot T(x) \tag{6.2}$$

Similarly, we view $T(c \cdot x)$ and $T(x)$ as elements of W and require that they satisfy scaling in the common field. This means T is compatible with the multiplication operation.

Definition 6.2. *Let V, W be two vector spaces over F. A map $T : V \to W$ is called a linear transformation if it satisfies Equations **??** and **??**.*

Lemma 6.1. *For all $x_1, \ldots, x_n \in V, a_1, \ldots, a_n \in F$,*

$$T\left(\sum_{i=1}^{n} a_i \cdot x_i\right) = \sum_{i=1}^{n} a_i T(x_i) \tag{6.3}$$

This is a generalization of Equations **??** in conjunction with Equation **??**; in particular, associativity shows that Equation **??** implies Lemma **??**:

$$T(x + y + z) = T((x + y) + z) = T(x + y) + T(z) = T(x) + T(y) + T(z) \tag{6.4}$$

Lemma 6.2. *V and W both have zero elements $\underline{0}_V \in V, \underline{0}_W \in W$. For any linear transformation T, the zero element in V maps to that in W:*

$$T(\underline{0}_V) = \underline{0}_W \tag{6.5}$$

*Proof of Lemma **??**.* We get this from the scaling property of a linear transformation.

$$\underline{0}_V = 0 \cdot x \ \forall x \in V \tag{6.6}$$
$$\tag{6.7}$$

Therefore, without loss of generality we choose $x \in V$ and apply scaling:

$$T(\underline{0}_V) = T(0 \cdot x) = 0 \cdot T(x) = \underline{0}_W \tag{6.8}$$

□

6.3 Vector Space Examples

Example 6.1. Consider the simplest possible vector space, $V = \{\underline{0}\}$ over some field F. Let W be any vector space over F. Then there is only one linear transformation $T : V \to W$, namely $T(\underline{0}) = \underline{0}_W$.

□

Example 6.2. We can consider a vector space that is itself the field. Let $V = W = F = \mathbb{R}$. There are many possible functions $f : \mathbb{R} \to \mathbb{R}$, and we've spent at least a year (Math 1A and Math 1B) studying them, but the ones we care about satisfy superposition and scaling:

$$\forall x, y \in \mathbb{R}, f(x+y) = f(x) + f(y) \tag{6.9}$$
$$\forall x, c \in \mathbb{R}, f(cx) = cf(x) \tag{6.10}$$

□

Lemma 6.3. *If $f : \mathbb{R} \to \mathbb{R}$ satisfies Equations* **??** *and* **??**, *then $\exists \alpha \in \mathbb{R}$ such that $f(x) = \alpha x$.*

It's a little sad that we have to restrict our purview to only these functions. This is in some sense what we expected: this is a linear algebra course, so it makes sense that we can only look at linear functions. But it feels like we're throwing out a lot. However, it turns out that we can approximate tons of functions to linear functions. This is what we're doing when we draw a tangent line to a point:

Example 6.3. Consider $f : \mathbb{R}^2 \to \mathbb{R}, \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \to \alpha_1 x_1 + \alpha_2 x_2$ for fixed $\alpha_1, \alpha_2 \in \mathbb{R}$. We can verify that this satisfies superposition and scaling, and so this is a valid linear transformation.

□

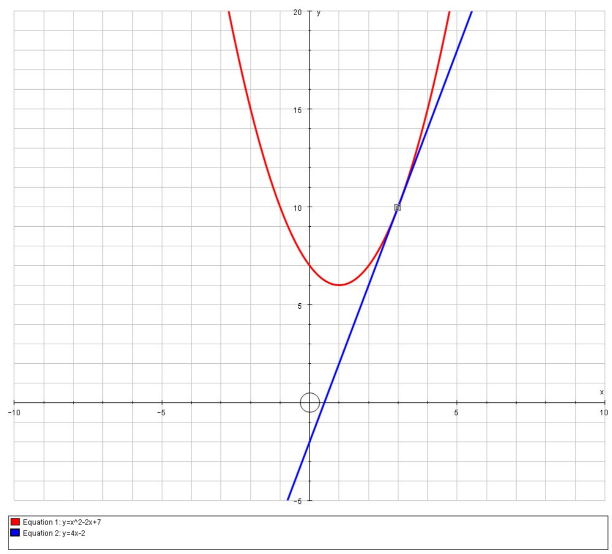


Figure 6.2: A tangent line locally approximating a quadratic function.

6.4 Natural Subspaces

Transformations have to satisfy the same linearity requirements as vector spaces, so we can construct certain vector spaces naturally from valid transformations. There are two natural subspaces of V and W that arise from a choice of the transformation T.

1. The null space of $T, N(T) = \{x \in V \mid T(x) = \underline{0}_W\}$.
2. The range of $T, R(T) = \{y \in W \mid \exists x \in V \,\text{s.t.}\, T(x) = y\}$

Proposition 6.4. *$N(T)$ and $R(T)$ are subspaces of V and W respectively.*

Proof. We previously saw that if a space was closed under addition and under scalar multiplication, and it contained the zero element, then it was a subspace. Here, we only check that both of the proposed subspaces have zero elements (the rest was left as an exercise for the reader, no, really):

1. $\underline{0}_V \in N(T)$ by Lemma **??**.
2. Take $y = \underline{0}_W$. Again, by Lemma **??**, $T(\underline{0}_V) = \underline{0}_W$, so $\underline{0}_W \in R(T)$.

□

Theorem 6.5. *Dimension Theorem: let V be a finite-dimensional vector space over F and let W be a vector space over F. Then $N(T)$ and $R(T)$ are also finite-dimensional, and*

$$\dim N(T) + \dim R(T) = \dim V. \tag{6.11}$$

Proof. $N(T) \subset V$; let $\dim V = n$, and let $\dim N(T) = k$. Then $k \leq n$. Since $N(T)$ is a vector space with dimension k, it has a basis of k elements. Let $T(x_i) = \underline{0}_W$ for all $i = 1, \ldots, k$ form a basis of $N(T)$. Now, we extend it to a basis of V, $\{x_1, \ldots, x_k, x_{k+1}, \ldots, x_n\}$.

Now, we can say that if V is a finite dimensional vector space and $\gamma = \{x_1, \ldots, x_k\}$ is a linearly independent subset of V with cardinality k, then there exists a subset $H = \{y_1, \ldots, y_{n-k}\}$ such that $\gamma \cup H$ is a basis of V. We can see this by the replacement theorem; take as G any basis of V, and take γ as L. Then we see that there exists $H \subset G$ such that $|H| = n - k$ and $\gamma \cup H$ is linearly independent. Since it has n elements and is linearly independent, it is a basis.

Next, we apply linearity to each of the basis elements, $T(x_i) = 0$ for $i = 1, \ldots, k$, and $T(x_i) \in W$ for $i = k+1, \ldots, n$. These are all some element of W, but we don't know anything specifically about them yet. We claim that $T(x_{k+1}), \ldots, T(x_n)$ form a basis in $R(T)$. If this is the case, then we're done because we've shown that $\dim N(T) = k$ and $\dim R(T) = n - k$, so $\dim N(T) + \dim R(T) = n$.

We first show that the T_{k+i}s are linearly independent, i.e. that

$$b_{k+1}T(x_{k+1}) + \cdots + b_n T(x_n) = \underline{0}_W, b_i \in F \implies b_{k+i} = 0 \tag{6.12}$$

"We can put the genie back into the bottle" and apply linearity to the sum to put everything inside of the transformation:

$$T(b_{k+1}x_{k+1} + \cdots + b_n x_n) = \underline{0}_W \tag{6.13}$$

Therefore, by definition, $b_{k+1}x_{k+1} + \cdots + b_n x_n \in N(T)$. But $\{x_1, \ldots, x_k\}$ is a basis of $N(T)$, so there exists some linear combination of them that adds up to the sum:

$$b_{k+1}x_{k+1} + \cdots + b_n x_n = a_1 x_1 + \cdots + a_k x_k \tag{6.14}$$

Since $\{x_i\}_{i=1}^n$ is a basis, all the a_is and b_{k+i}s are zero, therefore $b_{k+1} = \cdots = b_n = 0$, which implies that $T(x_{k+1}), \ldots, T(x_n)$ are linearly independent.

Now, we show that the T_{k+i} generate $R(T)$. Recall the definition of the range,

$$R(T) = \{T(x) \mid x \in V\} \tag{6.15}$$

Since V is a vector space with dimension n, we can write any element of V as a linear combination of n basis elements:

$$x = c_1 x_1 + \cdots + c_n x_n \tag{6.16}$$

Therefore, applying Lemma **??**:

$$T(x) = T(c_1 x_1 + \cdots + c_n x_n) = c_1 T(x_1) + \cdots + c_k T(x_k) + c_{k+1} + \cdots + c_n T(x_n) \tag{6.17}$$

We know that $T(x_1)$ through $T(x_k)$ are zero. Therefore, every element in $R(T)$ is a linear combination of the T_{k+i}s. Therefore

$$R(T) = \text{span}(T(x_{k+1}), \ldots, T(x_n)) \tag{6.18}$$

□

6.5 Numerical Representation

Let V be a finite-dimensional vector space over F and let $\dim V = n$. Fix a basis of $V, \beta = \{x_1, \ldots, x_n\}$. For all $v \in V$, there exists a unique representation as $\sum_{i=1}^{n} a_i x_i$ for some $a_i \in F$. We can think of these a_is as the coordinates of v relative to β. Essentially, this creates a map $V \to F^n$,

$$v \to \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = [v]_\beta \tag{6.19}$$

This places V and F^n in one-to-one correspondence.

Now, consider V, W to be finite-dimensional vector spaces over F. Let $\dim V = n, \dim W = m$. Suppose there is a transformation between them, $T : V \to W$. Choose a basis of V, $\beta = \{x_1, \ldots, x_n\}$, and choose a basis of W, $\gamma = \{y_1, \ldots, y_m\}$. For each basis element, we can take the transformation and find the resulting vector's representation in F^m. We denote this by $[T(x_i)]_\gamma$. If we do this for all basis elements, we get a collection of n columns (for each of the x_is), each of which is an element of F^m. This constructs a matrix in $F^{m\times n}$ that encodes the linear transformation. We denote this by $[T]_\beta^\gamma$.

Math 110: Linear Algebra **Fall 2019**

Lecture 7: The Vector Space of Linear Transformations

Lecturer: Edward Frenkel *23 September* *Aditya Sengupta*

7.1 Meta

Let $\mathcal{L}(V, W)$ be the set of all linear transformations from V to W. Define addition of $T : V \to W, U : V \to W$ by

$$\forall x \in V, (T + U)(x) = T(x) + U(x) \tag{7.1}$$

Define scalar multiplication: given $a \in F$, we say

$$\forall x \in V, (aT)(x) = aT(x) \tag{7.2}$$

Looks like this is a set with two natural operations.

Lemma 7.1. *$\mathcal{L}(V, W)$ with these two operations is a vector space over F.*

Suppose $\dim V = n, \dim W = m$. The dimension of $\mathcal{L}(V, W)$ is mn, and the dimension of $V \oplus W = n + m$.

7.2 Compositions and Numeric Representations

Consider functions $f : S_1 \to S_2, g : S_2 \to S_3$. We can define $g \circ f : S_1 \to S_3$ by $\forall x \in S_1, (g \circ f)(x) = g(f(x))$. $f(x)$ is an element of S_2 so it can be considered an input to g.

For vector spaces, consider the mapping

$$V \xrightarrow{T} W \xrightarrow{U} Z \tag{7.3}$$

between vector spaces V, W, Z. Suppose T and U are linear. Then, the composition can be given the name $U \circ T$.

Lemma 7.2. *$U \circ T$ is a linear function.*

To help us understand this, we employ a numerical representation. Let V be a finite-dimensional vector space over F. Choose a basis β in V; then there is a bijection $v \in V \to [v]_\beta \in F^n$, $v = \sum_{i=1}^n a_i v_i \to \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}^T$.

Next, consider vector spaces V, W over F and a transformation between them $T \in \mathcal{L}(V, W)$. Suppose there is an ordered basis $\alpha = \{x_i\}_{i=1}^n$ of V and an ordered basis $\beta = \{y_i\}_{i=1}^m$ of W. The transformation can be represented by a matrix in $M_{m\times n}(F)$. Row j of the matrix, $1 \leq j \leq n$, is given by $[T(x_j)]_\beta$; each of these columns has dimension m, i.e. they are vectors in F^m. We refer to this whole transformation by $[T]_\alpha^\beta$.

If we know $[T]_\alpha^\beta$, α, and β, then we know T. Further,

1. $[T]_\alpha^\beta$ tells us what the $T(x_i)$s are.
2. Once we know $T(x_i), mi = 1, \dots, n$, we know $T(x)$ for all $x \in V$ because it is sufficient to specify T on a basis and apply linearity to reach anywhere else in the space.

In general, this can be encoded in a matrix multiplication by

$$[T]_\alpha^\beta = [x]_\alpha \cdot [T(x)]_(\beta) \tag{7.4}$$

The transformation is encoded in an $m \times n$ matrix, which by multiplication takes a vector in F^n (isomorphic to V) to one in F^m (isomorphic to W).

Suppose we have ordered bases α, β, γ on V, W, Z which are vector spaces of dimension n, m, p. When we try to encode the transformation $V \to Z$ in terms of the natural ones given by the bases from $V \to W$ and $W \to Z$, we get a composition $U \circ T$ naturally encoded by an $p \times n$ that is the product of a $p \times m$ and $m \times n$ matrix.

$$[UT]_\alpha^\gamma [x]_\alpha = [U]_\beta^\gamma [T]_\alpha^\beta [x]_\alpha \tag{7.5}$$

Example 7.1. Consider the transformation $T_\theta : Vec_2 \to Vec_2$, representing rotation by an angle θ counterclockwise. Modulo 2π, this is linear, i.e. $T_\theta \circ T_\phi = T_{\theta+\phi}$. We get this from geometry. Pick a basis of two-dimensional space, $\alpha = \{x_1, x_2\}$; then

$$[T_\theta]_\alpha^\alpha = \left([T_\theta(x_1)]_\alpha \, [T_\theta(x_2)]_\alpha \right) \tag{7.6}$$

We use triangles and see that x_1 gets mapped to $x_1 \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$, and x_2 gets mapped to $\begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$. We can use linearity to derive the sine and cosine sum laws:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} cos(\theta+\phi) & \sin(\theta+\phi) \\ -\sin(\theta+\phi) & \cos(\theta+\phi) \end{bmatrix} \tag{7.7}$$

I don't want to actually multiply matrices, but we get

$$\cos\theta\cos\phi - \sin\theta\sin\phi = \cos(\theta+\phi) \tag{7.8}$$
$$\sin\theta\cos\phi + \cos\theta\sin\phi = \sin(\theta+\phi) \tag{7.9}$$

The rotations of a circle are a group!

$\square$

Example 7.2. Consider $V = W = \mathbb{C}$ viewed as a vector space over $\mathbb{R}$. $\dim \mathbb{C} = 2$, and choose $\alpha = \{1, i\}$. Every complex number is uniquely written as $a + bi = a \cdot 1 + b \cdot i$.

Let $T_{a+bi} : \mathbb{C} \to \mathbb{C}, x \to (a+bi)x$. This transforms the basis by $1 \to \begin{bmatrix} a \\ b \end{bmatrix}, i \to \begin{bmatrix} -b \\ a \end{bmatrix}$. That looks a lot like rotation. If we constrain $a^2 + b^2 = 1$, we get that $e^{i\pi} = -1$.

$\square$

Math 110: Linear Algebra **Fall 2019**

Lecture 8: Topic

Lecturer: Edward Frenkel *25 September* *Aditya Sengupta*

8.1

We previously saw that a finite-dimensional vector space V such that $\dim V = n$ was isomorphic to F^n for any particular choice of a basis of V. Consider the basis for F^n,

$$\{\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}, \begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix}, \dots, \begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}\} \tag{8.1}$$

This has the corresponding matrix representation $I_n = \delta_{ij}$, i.e. a matrix that is 1 down the diagonal and 0 otherwise.

Suppose the linear transformation ϕ_β goes $V \to F^n, v \to [v]_\beta$. This linear transformation ϕ_β is invertible; for all $\vec{y} \in F^n$, there exists a unique v such that $\phi_\beta(v) = \vec{y}$. This follows from the fact that all v have a unique representation as $\sum_{i=1}^n a_i x_i$.

(Definitions of injective and surjective here)

Consider two vector spaces, V, W over the same field, and a linear transformation $T : V \to W$.

Definition 8.1. *T is called invertible (an isomorphism) if $\exists T^{-1} : W \to V$ such that $T^{-1} \circ T = Id_V, T \circ T^{-1} \circ Id_W$.*

Lemma 8.1. *If it exists, T^{-1} is a linear transformation.*

Proof. $\forall y, z \in W, T^{-1}(y + z) = T^{-1}y + T^{-1}z$ as follows. For $y, \exists! x \in V$ such that $T(x) = y$, and similarly there exists a $u \in V$ for z. Then

$$T^{-1}(y + z) = T^{-1}(T(x) + T(u)) = T^{-1}(T(x + u)) = x + u = T^{-1}(y) + T^{-1}(z) \tag{8.2}$$

The proof of scalar multiplication is similar. □

Suppose V, W are finite-dimensional vector spaces over F.

Proposition 8.2. *There is an isomorphism $V \to W$ if and only if $\dim V = \dim W$.*

Proof. In the forward direction, suppose there exists $T : V \to W$ invertible. Then T is one-to-one onto, meaning that $R(T) = W$ and $N(T) = \{\underline{0}_V\}$. By the dimension theoremn, $\dim V = \dim N(T) + \dim R(T) = \dim W$.

In the backward direction, suppose $\dim V = \dim W = n$. Chooses bases for both, β and γ. Then there exists a unique linear transformation $T : V \to W, x_i \to y_i$. For all $v \in V$ there exists a unique $\{a_i\}$ such that $v = \sum a_i x_i$. This is an invertible linear transformation. I'll just copy this from someone later, right now I have priorities. $\square$

Math 110: Linear Algebra **Fall 2019**

Lecture 9: Change of Coordinates, Linear Bijections

Lecturer: Edward Frenkel *30 September* *Aditya Sengupta*

9.1 Linear Bijections

Example 9.1. Suppose $A, B \in M_{n\times n}(F)$ such that $AB = I_n$. Show that $BA = I_n$.

A encodes a linear transformation $L_A : F^n \to F^n, \vec{v} \to A\vec{v}$. This correspondence holds both ways, i.e. for a choice of basis of F^n, A is the unique matrix encoding $L_A : F^n \to F^n$. Therefore A corresponds uniquely to L_A under an isomorphism $M_{n\times n} \overset{\mathcal{L}}{\leftrightarrow} (F^n, F^n)$. Similarly B corresponds uniquely to some L_B. Since functions that compose to give the identity element must be the identity, $L_A \circ L_B = L_B \circ L_A = I_{F^n}$. Therefore we can apply the bijection and show that $AB = BA = I_n$. $\square$

This example relied on the following result,

Lemma 9.1. *Suppose $f : A \to B$ and f is invertible, meaning there exists $f^{-1} : B \to A$. Then if $g : B \to A$ such that $g \circ f = I_A$ and g is invertible, then $f \circ g = I_A$ and in fact $g = f^{-1}$.*

Theorem 9.2. *Let $T : V \to W$ where both are finite-dimensional vector spaces. If we know either that T is onto or one-to-one, it must be invertible.*

Proof. If T is onto, then $R(T) = W$. By the dimension theorem, $\dim N(T) = 0$, so $N(T) = \{\underline{0}\}$, meaning T is one-to-one and therefore bijective, so it is invertible.

If T is one-to-one, then $N(T) = \{\underline{0}\}$, so $\dim N(T) = 0$, so $\dim R(T) = \dim W$, so $R(T) = W$ meaning T is onto $\square$

9.2 Change of Coordinates

Suppose we have two bases β, β' of V/F. Let $\beta = \{x_i\}$ and $\beta' = \{x_i'\}$. Suppose there is some linear transformation L_Q between β and β'. Then, any linear transformation using β can be written as follows:

$$\phi_\beta(v) = L_Q \circ \phi_{\beta'}(v) \tag{9.1}$$

$$[v]_\beta = Q[v]_{\beta'} \tag{9.2}$$

$$[x_i']_\beta = Q[x_i']_{\beta'} \tag{9.3}$$

$[x_i']$ in β' coordinates is just a coordinate vector with one 1 and 0s everywhere else. This allows us to construct Q:

$$Q = [[x'_i]_\beta \mid 1 \leq i \leq n] \tag{9.4}$$

Math 110: Linear Algebra **Fall 2019**

Lecture 10: Matrices are Encoded Linear Transforms

Lecturer: Edward Frenkel *2 October* *Aditya Sengupta*

10.1 Cascading Transforms

(lowkey I'm super behind on my SLC reading so I'm gonna do that in parallel, sorry if I miss things)

We previously constructed a change-of-basis matrix $\beta' \to \beta$,

$$Q = \begin{bmatrix} [x_1']_\beta & [x_2']_\beta & \dots & [x_n']_\beta \end{bmatrix} \tag{10.1}$$

Here, we can get the basis elements by $[v]_\beta = Q[v]_{\beta'}$ for all $v \in V$. We can compactly write this by $Q = [I_v]_{\beta'}^{\beta}$. In general, a linear transformation $T : V \to W$ between bases α of V (with dimension n) and γ of W (with dimension m) is specified by the matrix

$$[T]_\alpha^\gamma = \begin{bmatrix} [T(y_1)]_\gamma & [T(y_2)]_\gamma & \dots & [T(y_n)]_\gamma \end{bmatrix} \tag{10.2}$$

and by applying the transform to each element of the basis α in turn, we get the relationship

$$[T(v)]_\gamma = [T]_\alpha^\gamma [v]_\alpha \tag{10.3}$$

Remark 10.1. *Often, if $T : V \to V$, we take $\alpha = \gamma$ and so we take the matrix $[T]_\alpha^\alpha$ or more simply just $[T]_\alpha$. In this case, if $T = I_V$, we get the identity matrix,*

$$[I_V]_\alpha^\alpha = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} = I_n \tag{10.4}$$

Suppose we have $T : V \to V$ with matrix A. I really don't feel like doing tikzpicture (some day!) so here's a picture of the diagram he drew on the board.

(note: the arrow going up to ϕ_β^{-1} on the right should be going down.)

Given β', we get $A' = [T]_{\beta'}$.

If we're translating from one basis of V to another, we have a lot of things going on: translation to F^n, change of basis in F^n, the actual transformation, and the change back to V. We can do this in any order. This gives us the Crazy Transformation Cube. That's the formal math term, trust me I'm an EMS major.

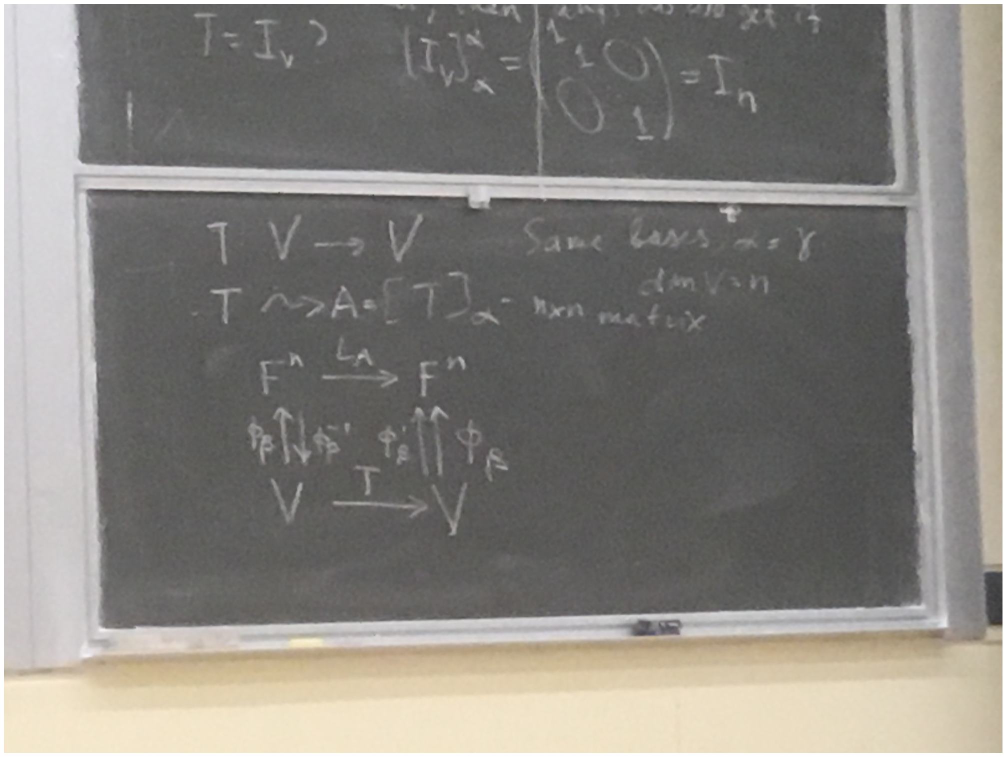


Figure 10.3: A visualization of the transformations between V and F^n

By carrying out the transformations, we can get the relationship $L'_A = L_Q^{-1} L_A L_Q$ (where L_A is the transformation associated to A in F^n-space and L_Q is the F^n change of basis) which gives us that $A' = Q^{-1}AQ$; this is how you transform a change-of-basis formula between bases. A wild concept that arises from this is that linear change-of-basis transforms are a conceptual normal subgroup of linear transforms. Although technically any matrix could be associated to a change of basis so this isn't that surprising. If you specify something like the matrix has to have unit determinant, you can specify a subgroup and say it has to be normal. Math is amazing. `https://yutsumura.com/special-linear-group-is-a-normal-subgroup-of-general-linear-group/`

10.2 Space of Linear Transformations

Consider finite-dimensional vector spaces V and W, and the space $\mathcal{L}(V, W)$ of linear transformations $T : V \to W$. Let $\dim V = n$ and $\dim W = m$, and say there exist bases β and γ of V and W, corresponding to which there exist isomorphisms $\phi_\beta : V \to F^n, \phi_\gamma : W \to F^m$. Then, we get an isomorphism $\mathcal{L}(V, W) \to M_{m\times n}(F), T \to [T]_\beta^\gamma$. Hence $\dim \mathcal{L}(V, W) = \dim M_{m\times n}(F) = mn$. A special case of this is where $W = F$.

Definition 10.1. *Given a vector space V over F, the dual vector space V^* to V is $\mathcal{L}(V, F)$.*

(this is the first thing we've done that wasn't in 54 get hyped)

$\dim V^* = 1 \cdot n = n = \dim V$. There exist isomorphisms between V and V^* but there are many and there is no preferred isomorphism in general. However, we will prove that $V^{**} = (V^*)^*$ is canonically isomorphic to V without having to choose an isomorphism.

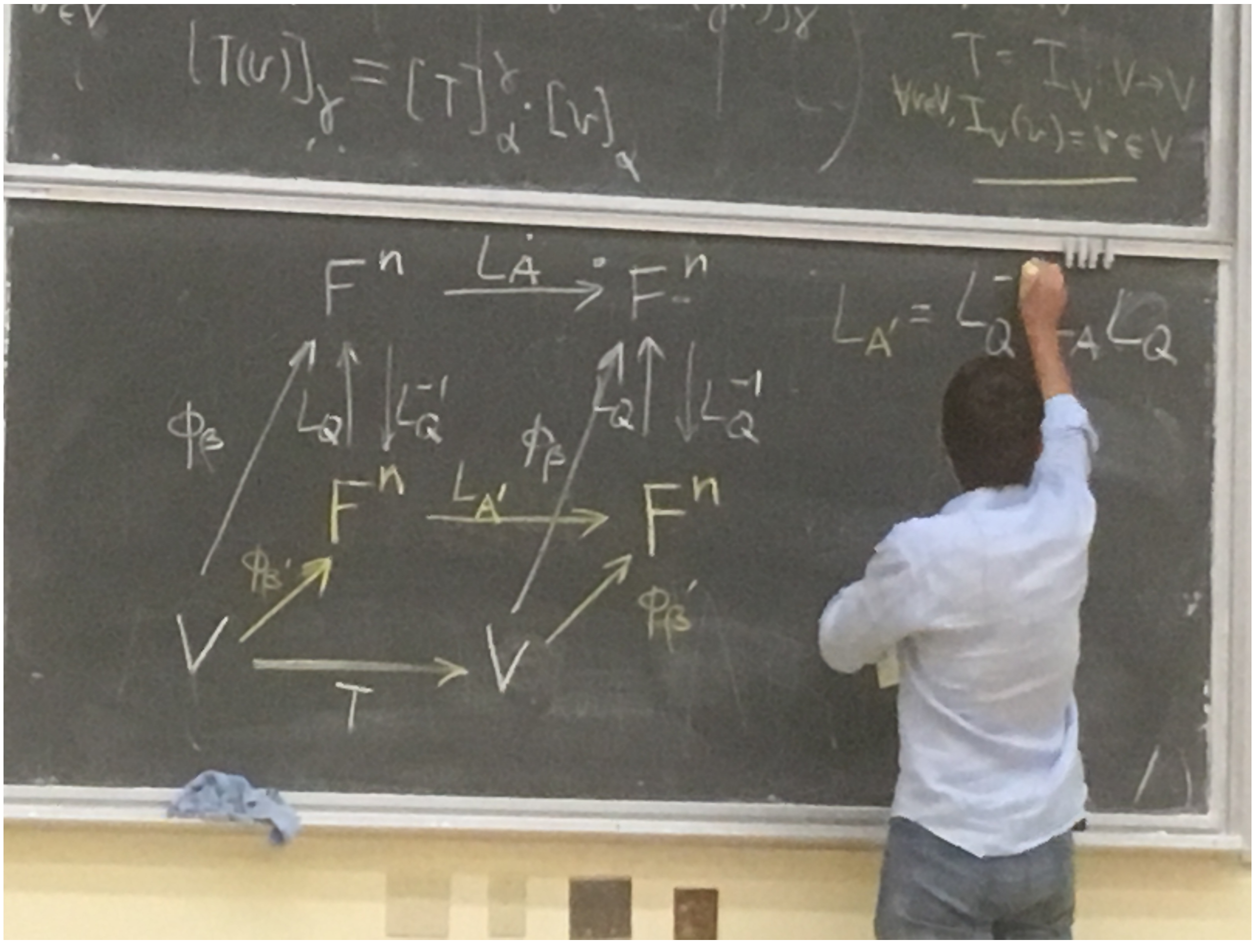


Figure 10.4: The Crazy Transformation Cube

A transformation $V \to W$ gets changed to $T^t : W^* \to V^*$ under the duals of both vector spaces. Its associated matrix is just the transpose of the matrix associated to T.

Elements of V^* are linear transformations $V \to F$. We call these <u>linear functionals</u>, $f : V \to F$. These are essentially multivariate functions. For example, let's say $F = \mathbb{R}$; then V^* consists of elements $f : V \to \mathbb{R}$. Further, as V is isomorphic to $\mathbb{R}^n$ for some n, these are functions $f : \mathbb{R}^n \to \mathbb{R}$. For example, take $n = 3$ and you get a 53-type vector valued function. More specifically, a linear functional $\mathbb{R}^3 \to \mathbb{R}$ can be written as

$$f\left(\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}\right) = k_1 a_1 + k_2 a_2 + k_3 a_3 \tag{10.5}$$

for some $k_1, k_2, k_3 \in \mathbb{R}$. Note that this can be equivalently rewritten as

$$f\left(\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}\right) = \begin{bmatrix}k_1 & k_2 & k_3\end{bmatrix}\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix} \tag{10.6}$$

For another example, let's take $V = P_n(\mathbb{R})$. Each eleement of V is a function $\mathbb{R}$ to $\mathbb{R}$. Then, V^* is the space of all functions on these functions (wild) which take $p(t) \to f(p(t))$. Let $x_0 \in \mathbb{R}$, and define $f_{x_0}(p(t)) = p(a)$. For example, if $n = 2$, $P_2(\mathbb{R}) = \{a_0 + a_1t + a_2t^2 \mid a_i \in \mathbb{R}\}$. Then $f_{x_0}(p(t)) = p(x_0), a_0 + a_1t + a_2t^2 \to a_0 + a_1x_0 + a_2x_0^2$, which is just a number. For $x_0 = 2$, we get $f_2(p(t)) = a_0 + 2a_1 + 4a_2$ which we can see is linear!

Linear functionals usually correspond to natural operations; for example, integration is linear:

$$f_{a,b}(p(t)) = \int_a^b p(t)dt. \tag{10.7}$$

Suppose $\dim V = n$ and $\beta = \{x_i\}$ is a basis. Can we construct a basis in V^*? Yes! This is called the dual basis to β.

Theorem 10.2. *Define $f_i : V \to F$ by*

$$f_i\left(\sum_{j=1}^{n} a_jx_j\right) = a_i \in F \tag{10.8}$$

These are linear transformations $V \to F$ and $\{f_i\}$ is a basis of V^*.

Math 110: Linear Algebra **Fall 2019**

Lecture 11: Dual Spaces

Lecturer: Edward Frenkel *7 October* *Aditya Sengupta*

11.1 Basis of a Dual Space

Recall that the dual space of a vector space V over F is $V^* = \mathcal{L}(V, F) = \{f : V \to F \mid f \text{ linear}\}$. We want to construct a basis β^* of V^* based on a basis β of V.

Let $\beta = \{x_1, \dots, x_n\}$. Define linear functionals $f_i \in V^*$ by the formula

$$f_i(x_j) = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{11.1}$$

Then, we can use linearity to write

$$f_i(v) = f_i\left(\sum_{j=1}^{n} a_j x_j\right) = \sum_{i=1}^{n} a_j f_i(x_j) = a_i \in F \tag{11.2}$$

We suppose that $\beta^* = \{f_1, f_2, \dots, f_n\}$. To show this is a basis, we essentially want to show that we can write any $f\ inV^*$ as a linear combination of the f_is.

$$f = \sum_{i=1}^{n} c_i f_i \tag{11.3}$$

$$f(x_j) = \left(\sum_{i=1}^{n} c_i f_i\right)(x_j) = \sum_{i=1}^{n} c_i f_i(x_j) = c_j \tag{11.4}$$

Based on this, we make the guess

$$f = \sum_{i=1}^{n} \color{red}{f(x_i)} f_i \tag{11.5}$$

where $\color{red}{f(x_i)}$ is in red to denote that it is an element of F, instead of a functional. From this, we will get that $\operatorname{span}(\beta^*) = V^*$. Since $|\beta^*| = n = \dim V^*$, we get that β^* is a basis of V^*.

Example 11.1. Let $V = \mathbb{P}_1(\mathbb{R}) = \{a + bt \mid a, b \in \mathbb{R}\}$. Recall that $\phi_c : \mathbb{P}_n(\mathbb{R}) \to \mathbb{R}, p(t) \to p(c)$, is a linear functional (the evaluation homomorphism). Then, define $\phi_{a,b} : \mathbb{P}_n(\mathbb{R}) \to \mathbb{R}, p(t) \to \int_a^b p(t)dt$. This is also a linear functional.

Let $f_1(p) = \int_0^1 p(t)dt, f_2(p) = \int_0^2 p(t)dt$. To show that $\{f_1, f_2\}$ is a basis of $(\mathbb{P}_1(\mathbb{R}))^*$, it is sufficient to find its dual basis, i.e. to find $p_1^{(t)}, p_2^{(t)} \in \mathbb{P}_1(\mathbb{R})$ such that $f_i(p_j) = \delta_{ij}$.

We let both of them have free parameters:

$$p_1(t) = a_1 + b_1 t \implies f_1(p_1) = \int_0^1 (a_1 + b_1 t)dt = a_1 + \frac{1}{2}b_1 = 1 \tag{11.6}$$

$$f_2(p_1) = \int_0^2 (a_2 + b_2 t)dt = 2a_1 + 2b_1 = 0 \tag{11.7}$$

From this, we get $a_1 = -b_1$ and $a_1 - \frac{1}{2}a_1 = 1$ so $a_1 = 2$ and $b_1 = -2$.

Likewise, we require that $f_1(p_2) = 0$ and $f_2(p_2) = 1$, which gives us $p_1(t) = 2 - 2t, p_2(t) = -\frac{1}{2} + t$. These are clearly linearly independent, so they form a basis. Therefore $\{f_1, f_2\}$ form a basis of the dual space.

□

Remark 11.1. *Note that we can generalize the above idea to $\mathcal{F}(S, \mathbb{R}) = \{p : S \to \mathbb{R}\}$ where S is a set. This set has to be such that linearity is satisfied, i.e. $(p + q)(s) = p(s) + q(s)$ and $(kp)(s) = kp(s)$ for all $s \in S, k \in F$.*

For example, fix $c \in S$, and let

$$\phi_c : \mathcal{F}(S, \mathbb{R}) \to \mathbb{R}, p \to p(c) \tag{11.8}$$

i.e. the evaluation of p at c.

The basis $\beta = \{p_1, p_2\}$ we constructed before can be used to specify the evaluation homomorphism:

$$\phi_c = \phi_c(p_1)f_1 + \phi_c(f_2) = (2 - 2c)f_1 + \left(-\frac{1}{2} + c\right) f_2 \tag{11.9}$$

For example, for any polynomial in $\mathbb{P}_1(\mathbb{R})$,

$$p(10) = -18 \int_0^1 p(t)dt + \frac{19}{2} \int_0^2 p(t)dt \tag{11.10}$$

11.2 Linear Transformations on Dual Spaces

Recall that under a dual-space transformation, $T : V \to W$ goes to $T^\intercal : W^* \to V^*$. In the original vector spaces, suppose we have a linear functional $g : W \to F$, $g \in W^*$. We can compose the original transformation $T : V \to W$ with this to get a linear functional $V \to F$, which is an element of V^*. For a fixed T, this specifies a transformation from W^* to V^*:

$$T^\intercal : g \to g \circ T \tag{11.11}$$

Suppose that V has a basis $\beta = \{x_1, \ldots, x_n\}$ and W has a basis $\gamma = \{y_1, \ldots, y_n\}$. Then, to each transformation $T : V \to W$, we can associate a matrix $[T]_\beta^\gamma : V \to W, \vec{v} \to A_{m\times n}\vec{v}$. There is also a dual transformation $T^\intercal : W^* \to V^*$. This has an associated matrix $[T^t]_{\gamma^*}^{\beta^*} : W^* \to V^*, w(t) \to A_{n\times n}^\intercal w(t)$ where $w(t) \in W^*$.

The transpose operation is an isomorphism $A \in M_{m\times n}(F) \to A^\intercal \in M_{n\times n}(F)$. We don't yet know for sure that the matrix transpose says anything about the transformation, but we can show that.

Theorem 11.2.

$$[T^\intercal]_{\gamma^*}^{\beta^*} = \left([T]_\beta^\gamma\right)^\intercal \tag{11.12}$$

The matrix operation $A \to A^\intercal$ is trivial, but it's not obvious that changing T to $T^\intercal$ would give us anything meaningful. To make this clearer, we have to start thinking of V and V^* as equal dual pairs. Just like how the time domain is clearly superior to the frequency domain, right up until you come across a convolution integral.

Consider an operation $V^* \times V \to F; f, v \to f(v) = \langle f|v\rangle$. To make this more concrete, let $V = \mathbb{R}^n$ with the canonical basis $\{e_1, \ldots, e_n\}$. The dual basis is $\{f_1, \ldots, f_n\}$ where $f_i(e_j) = \delta_{ij}$. We can specify these elements by row vectors; for example, multiplying $\begin{bmatrix}1 & 0 & \ldots & 0\end{bmatrix}$ by an element of V is the same as applying f_1 on it. We can specify the dual space by $V^* = \{(b_1, \ldots, b_n) \mid b_i \in \mathbb{R}\}$. Then we can evaluate the generic $f(v)$:

$$f(v) = \begin{bmatrix}b_1 & b_2 & \ldots & b_n\end{bmatrix} \cdot \begin{bmatrix}a_1 & a_2 & \ldots & a_n\end{bmatrix}^\intercal = a_1 b_1 + \cdots + a_n b_n \tag{11.13}$$

Like we'd expected, this is just a number. More explicitly, $\sum_{i=1}^n a_i b_i$ is the value of $f = \begin{bmatrix}b_1 & b_2 & \ldots & b_n\end{bmatrix}$ on $\begin{bmatrix}a_1 & a_2 & \ldots & a_n\end{bmatrix}^\intercal \in \mathbb{R}^n$.

Now that we have this structure, we can allow both f and v to vary. This gives us $|v\rangle = \begin{bmatrix}a_1 & a_2 & \ldots & a_n\end{bmatrix}^\intercal$ and $\langle f| = \begin{bmatrix}b_1 & b_2 & \ldots & b_n\end{bmatrix}$ as quantities we can look at separately, as well as the product $\langle f|\,|v\rangle$.

The whole notion of dual spaces assumes that V and $(V^*)^*$ are isomorphic. We can construct a linear transformation $V \to (V^*)^*$ and show it's an isomorphism to prove that.

Lecture 12: Review

Lecturer: Edward Frenkel *14 October* *Aditya Sengupta*

31 is a Mersenne prime and I'll thank you to remember it and that's why Halloween is better than Christmas.

Math 110: Linear Algebra **Fall 2019**

Lecture 13: Diagonalization

Lecturer: Edward Frenkel *21 October* *Aditya Sengupta*

13.1 Motivation

Diagonalization gives us a new way of looking at a linear transformation, that can tell us more about what it does. For example, in general, if we don't specify anything about two transformations then they won't commute, i.e. $AB \neq BA$. However, diagonal matrices do commute.

The structure of a diagonal matrix is $a_{ij} = \lambda_i \delta_{ij}$, i.e. values are nonzero only on the diagonal.

$$\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Theorem 13.1. *If A and B are diagonal, then $AB = BA$.*

13.2 Eigenstuff

Recall that $A_{n\times n}$ encodes a linear transformation $L_A : F^n \to F^n$, with respect to some basis $\{x_1, \dots, x_n\}$. For now let's say it's the standard basis. A is diagonal if and only if $L_A(x_i) = \lambda_i e_i$. This is called the eigenvector equation.

If x_i satisfies this equation, then x_i is an eigenvector of A (or equivalently of L_A). λ_i is the eigenvalue of A on x_i.

Definition 13.1. *An eigenbasis is a basis of F^n such that every element is an eigenvector.*

Theorem 13.2. *A is diagonal if and only if the standard basis is an eigenbasis.*

Even if A is not diagonal, i.e. the standard basis β is not an eigenbasis, there may be another basis $\beta' = \{x'_1, \dots, x'_n\}$ such that $x'_i \in F^n$ which is an eigenbasis of L_A. Then we have the coordinate change matrix Q:

$$Q = \begin{bmatrix} [x'_1]_\beta & \dots & [x'_n]_\beta \end{bmatrix} \tag{13.1}$$

which is defined so that $[v]_\beta = Q[v]_{\beta'}$. Then, we've previously seen that a transformation can be done in another basis according to

$$[T]_{\beta'} = Q^{-1}[T]_\beta Q \tag{13.2}$$

We first translate $\beta' \to \beta$, do the transformation in β, and finally translate back $\beta' \to \beta$. Therefore, if we assume that β' is our desired eigenbasis, some algebra gives us $A = QDQ^{-1}$. We say that A is similar to D, i.e. there exists some change of basis (an invertible $n \times n$ matrix) Q such that $A = QDQ^{-1}$.

13.3 Properties of Matrices Similar to Diagonal Matrices

If we have two matrices that are not diagonal, but are both similar to diagonal matrices with the same Q, we can see they commute by nice matrix multiplication cancellations:

$$
\begin{aligned}
A = QDQ^{-1}, B = QCQ^{-1} \\
AB &= QDQ^{-1}QCQ^{-1} = QDICQ^{-1} = QDCQ^{-1} \\
&= QCDQ^{-1} = QCIDQ^{-1} = QCQ^{-1}QDQ^{-1} \\
&= BA
\end{aligned}
$$

Further, diagonalization makes it easy to raise matrices to arbitrary powers:

$$
\begin{aligned}
A^n &= (QDQ^{-1})^n = (QDQ^{-1})(QDQ^{-1})\dots(QDQ^{-1}) && (13.3)\\
&= QD^nQ^{-1} && (13.4)
\end{aligned}
$$

where D^n is just $\lambda_i^n \delta_{ij}$.

13.4 Definitions and Setup

Definition 13.2. *Let V be a finite-dimensional vector space over F and let $T : V \to V$ be a linear operator. An element $v \in V$ is called an eigenvector of T if*

1. $v \neq 0$
2. $T(v) = \lambda v$ *for some* $\lambda \in F$.

We can translate this into a statement on matrices. If β is some basis of V, let $A = [T]_\beta$. Let $\vec{y} = [v]_\beta$. Then if v is an eigenvector of T with eigenvalue λ, then $\vec{y}$ is an eigenvector of A with eigenvalue λ. At the outset, we are given A, an $ntimesn$ matrix. We want to find $y_1, \dots, y_n$ and λ such that

$$
A \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \lambda \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{13.5}
$$

It seems difficult to look for its eigenvalues and eigenvectors, because there are n equations and $n+1$ variables. However, it'll turn out that we can find all of these values withuot solving any systems of equations. As a first step, say we find $\{\lambda_1, \dots, \lambda_k\}_{k \leq n}$ of eigenvalues. Then, we can solve the fully-constrained system $A \cdot \vec{x} = \lambda_i \vec{x}$ to find the eigenvectors.

13.5 Determinants

To find the eigenvalues, we introduce the determinant. This is a function $M_{n\times n}(F) \to F, A \to \det(A)$. This is not a linear functional.

We're used to the really ugly formula for the determinant, through cofactor expansion. There is a much nicer formulation of the determinant that involves dual spaces. However, for now let's just accept the ugly formula and trust that it will get the job done.

The key properties of the determinant are

(i) A is invertible if and only if $\det(A) \neq 0$.

(ii) The product AB has a determinant $\det(A)\det(B)$.

To start with 2×2 matrices, we attempt to come up with a formula that will tell us when the matrix is invertible. A 2×2 matrix is non-invertible if one of its columns is a scalar multiple of the other: if you view the columns of $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ as fractions, then for invertibility we want $\frac{a_{11}}{a_{21}} \neq \frac{a_{12}}{a_{22}}$. Cross-multiplying, we get a natural formula for the determinant: $\det(A) = a_{11}a_{22} - a_{12}a_{21}$.

For higher dimensions, we define it recursively and decide we hate intuition.

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+1} a_{1i} \det(\hat{A}_{1i}) \tag{13.6}$$

where $\hat{A}_{1i}$ is the matrix we get by removing the first row and ith column of A.

13.6 Finding eigenvalues

Suppose $\vec{x}$ is an eigenvector of A with eigenvalue λ. Then

$$\begin{aligned} A\vec{x} &= \lambda\vec{x} \\ A\vec{x} - \lambda I_n \vec{x} &= \vec{0} \\ (A - \lambda I_n) \cdot \vec{x} &= \vec{0} \end{aligned}$$

Recall that $\vec{x} \neq \vec{0}$; we get that $\vec{x} \in N(A - \lambda I_n)$ and $\dim N(A - \lambda I_n) > 0$. That is, $A - \lambda I_n$ is not invertible, which means its determinant must be zero.

To find all eigenvalues of A, we introduce an unknown t and compute

$$f_A(t) = \det(A - tI_n) \tag{13.7}$$

This is called the characteristic polynomial of A, and we have shown that the eigenvalues of A are the roots of this polynomial, because we want to make its determinant zero.

Example 13.1. Let

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix}$$

We subtract tI_n and take a determinant:

$$\det(A - tI_2) = \begin{bmatrix} 1-t & -1 \\ 2 & 4-t \end{bmatrix} = (1-t)(4-t) + 2 = t^2 - 5t + 6$$

This has roots $t = 2, t = 3$. First, we take the eigenvalue $\lambda = 2$, and we get

$$(A - 2I_2) \cdot \vec{x} = 0$$
$$\begin{bmatrix} -1 & -1 \\ 2 & 2 \end{bmatrix} \cdot \vec{x} = 0$$

A vector spanning the null space of this matrix is $\vec{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. We see that there's only one unique constraint being applied, but the vector $\vec{x}$ is in F^2, so the dimension of the null space is 1. Therefore, the general eigenvector is $\begin{bmatrix} a \\ -a \end{bmatrix}$, but we can pick a representative. We'll later have the idea of normalizing the eigenvectors, but we don't know what an inner product is yet, so we just pick the simplest representative.

Next, we take the eigenvalue $\lambda = 3$, and we get

$$(A - 3I_2) \cdot \vec{x} = 0$$
$$\begin{bmatrix} -2 & -1 \\ 1 & 2 \end{bmatrix} \cdot \vec{x} = 0$$

A vector spanning the null space of this matrix is $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$.

Therefore, the eigenbasis we wanted is $\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \end{bmatrix}\}$. We can construct Q and Q^{-1}:

$$Q = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}, Q^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

Frenkel just barely remembered how to do an inverse.

Therefore, we can say that

$$A = Q \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} Q^{-1} \tag{13.8}$$

□

Math 110: Linear Algebra **Fall 2019**

Lecture 14: Eigenstuff

Lecturer: Edward Frenkel *23 October* *Aditya Sengupta*

Theorem 14.1. *The eigenvalues of A are the roots of the characteristic polynomial $f_A(t) = \det(A - tI_n)$ and vice versa.*

If λ is a root of $f(t)$, then it can be written as $(t-\lambda)g(t)$ where $g(t)$ is another polynomial. Continuing this process, if we have k roots $\lambda_1, \dots, \lambda_k$, then $f(t) = \prod_{i=1}^{k}(t-\lambda_i)g(t)$.

We say that $f_A(t)$ is split if we can write it entirely like this, i.e. there exist $\lambda_1, \dots, \lambda_n$ such that

$$f_A(t) = (-1)^n \prod_{i=1}^{n}(t-\lambda_i) \tag{14.1}$$

We can write out the determinant $A - tI_n$ as follows to identify the λ_is and also recognize why we have the $(-1)^n$ above:

$$\begin{vmatrix} a_{11}-t & \dots & \dots & \dots \\ \dots & a_{22} & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & \dots & a_{nn} \end{vmatrix} = (-t)^n + (-t)^{n-1}\operatorname{tr} A + \cdots + t^0 \det A$$

There's no easy expression for the $\dots$ in the middle of that polynomial in t. (Not sure of the purpose of doing this.)

Let $\lambda_1, \dots, \lambda_k$ be the distinct roots; $\lambda_i \neq \lambda_j$ if $i \neq j$. Let m_i be the algebraic multiplicity of i, i.e. the exponent if we write out the polynomial in the form

$$f_A(t) = (-1)^n \prod_{i=1}^{k}(t-\lambda_i)^{m_i} \tag{14.2}$$

Each m_i is at least 1 and is some integer. For example, if all the m_is are 1, then $k = n$ so $f_A(t)$ has n distinct roots.

Suppose we have A such that its characteristic polynomial is split. Then A has eigenvalues $\lambda_1, \dots, \lambda_k$ with multiplicities $m_1, \dots, m_k$. First, consider the generic case $k = n$. Then there exists $y_i \in F^n$ such that $Ay_i = \lambda_i y_i$ for all $i = 1, \dots, n$.

Theorem 14.2. *The set $\{y_1, \dots, y_n\}$ is an eigenbasis.*

Proof. We know these are all eigenvectors, so it remains to prove they are a basis. Since $\dim F^n = n = |\{y_1, \dots, y_n\}|$, it will suffice to prove it is linearly independent, i.e. that if $\sum_i a_i y_i = \underline{0}$ and $a_i \in F$, then all the a_is must be 0.

We will prove this by induction. If $n = 1$ then $\{y_1\}$ is trivially linearly independent. Suppose true for some $k < n$, that is, $\{y_1, \dots, y_k\}$ is linearly independent, then we add in y_{k+1}.

$$a_1 y_1 + \cdots + a_k y_k + a_{k+1} y_{k+1} = \underline{0} \tag{14.3}$$

Apply $(A - \lambda_{k+1} I_n)$ to both sides. Then

$$(A - \lambda_{k+1} I_n) y_i = A y_i - \lambda_{k+1} y_i = \lambda_i y_i - \lambda_{k+1} y_i = (\lambda_i - \lambda_{k+1}) y_i \tag{14.4}$$

Therefore

$$a_1(\lambda_1 - \lambda_{k+1}) y_1 + a_2(\lambda_2 - \lambda_{k+1}) y_2 + \cdots + a_k(\lambda_k - \lambda_{k+1}) y_k + a_{k+1}(\lambda_{k+1} - \lambda_{k+1}) y_{k+1} = \underline{0} \tag{14.5}$$

The last term drops out and we're left with a linear combination of $\{y_1, \dots, y_k\}$. We know this is linearly independent, so $a_i(\lambda_i - \lambda_{k+1}) = 0$. But recall that we took all the λ_is distinct, so $\lambda_i - \lambda_{k+1}$ is always nonzero. Therefore the a_is up to k are all zero by linear independence. This means a_{k+1} must also be 0. So we have shown $\{y_1, \dots, y_{k+1}\}$ is linearly independent. $\square$

Next, let's consider the case where $k < n$, so there exists at least one $m_i > 1$. Look at 2×2 diagonal matrices. For example, the characteristic polynomial of $\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ is $(t - \lambda_1)(t - \lambda_2)$. If we have a multiplicity greater than 1, this reduces to $\lambda_1 = \lambda_2 = \lambda$ and the polynomial reduces to $(t - \lambda)^2$. We'll find that this corresponds to two linearly independent eigenvectors with the same eigenvalue.

Definition 14.1. *An eigenspace of A corresponding to a given eigenvalue λ is the set of all solutions to $Av = \lambda v$ where $v \in F^n$.*

We denote this by $E_\lambda \subset F^n$.

Lemma 14.3.

1. *E_λ is a subspace of F^n.*
2. *In the generic case where no two eigenvalues are the same,* $\dim E_{\lambda_i} = 1$ *for all i.*
3. *$E_\lambda \neq \{\underline{0}\} \iff \lambda$ is an eigenvalue of A.*

Lemma 14.4. *Suppose A has eigenvalues $\lambda_1, \dots, \lambda_k$ with multiplicities $m_1, \dots, m_k$. Then* $\dim E_{\lambda_i} \leq m_i$.

Proof. Choose a basis of E_{λ_i} and extend it to a basis of F^n. Relative to this basis, our matrix is diagonal in the first k rows and columns, with all the diagonal elements equal to λ_1 and is zero below this diagonal block. (See q5 on the practice midterm.) That means the characteristic polynomial has the form $(\lambda_1 - t)^{\dim E_{\lambda_1}} g(t)$. Therefore $\dim E_{\lambda_1} \leq m_1$. $\square$

Theorem 14.5. *A is diagonalizable if and only if* $\dim E_{\lambda_i} = m_i$ *for all $i = 1, \dots, k$. Suppose β_i is a basis of E_{λ_i}; then an eigenbasis is a disjunctive union $\beta = \beta_1 \sqcup \beta_2 \sqcup \cdots \sqcup \beta_k$.*

Proof. In the backward direction, β_i has m_i elements, so the cardinality of β is $\sum_i m_i = \deg f_A(t) = n$. Therefore β is a set with n elements. It remains to be shown that β is linearly independent. This is done like we did it last time.

In the forward direction, we know that A is diagonalizable. So there is an eigenbasis $\gamma_1, \ldots, \gamma_k$ where γ_i contains all eigenvectors with eigenvalue λ_i. $|\gamma_i| \leq \dim E_{\lambda_i} \leq m_i$ by the lemma. Therefore

$$n = \sum_{i=1}^{k} |\gamma_i| \leq \sum_{i=1}^{k} m_i = n \tag{14.6}$$

Therefore $|\gamma_i| = m_i$ so $\dim E_{\lambda_i} = m_i$. $\square$

Math 110: Linear Algebra **Fall 2019**

Lecture 15: Jordan canonical form

Lecturer: Edward Frenkel *30 October* *Aditya Sengupta*

Recall that A is diagonalizable if there exists some matrix Q such that $A = QDQ^{-1}$ where D is diagonal. The existence of Q is equivalent to the existence of an eigenbasis $\{x_1, \dots, x_n\}$; if we have this basis, we can construct Q just by treating each of the elements of the eigenbasis as a column.

$$Q = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$$

Last time, we saw that A is diagonalizable if two conditions are met:

1. $f_A(t)$ is split, i.e. there exists an expansion of the characteristic polynomial,

$$f_A(t) = \det(A - tI_n) = (-1)^n \prod_{i=1}^{k} (t - \lambda_i)^{m_i}. \tag{15.1}$$

   $\lambda_1, \dots, \lambda_k$ are distinct eigenvalues, and m_i is referred to as the multiplicity of λ_i.

2. $\dim E_{\lambda_i} = m_i$; recall that $E_{\lambda_i} = \{v \in F^n \mid (A - \lambda_i I_n)v = \underline{0}\}$ is the eigenspace corresponding to λ_i. In general, $E_{\lambda_i} \neq \{\underline{0}\}$, and $1 \leq \dim E_{\lambda_i} \leq m_i$. This condition is satisfied almost all of the time.

A field F is called algebraically closed if any polynomial over F splits, i.e. the first condition is always satisfied. For example, $\mathbb{C}$ is algebraically closed. Note that $\mathbb{R}$ is not algebraically closed, because $t^2 + 1$ has no solutions in $\mathbb{R}$. Suppose we have the matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, which has the characteristic polynomial $t^2 + 1$. This does not split over $\mathbb{R}$ but it does split over $\mathbb{C}$. We get eigenvectors defined over $\mathbb{C}^2$ instead of $\mathbb{R}^2$.

Consider the following matrices over $\mathbb{R}$: $A = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$ and $B = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$. Both of these have the characteristic polynomial $(t - \lambda)^2$, i.e. they have eigenvalue λ with multiplicity 2. To find the eigenvectors of B, we want to find the vectors that satisfy

$$\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \lambda \begin{bmatrix} a \\ b \end{bmatrix} \tag{15.2}$$

i.e.

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{15.3}$$

The solutions to this are of the form $\begin{bmatrix} a \\ 0 \end{bmatrix}$, so $E_\lambda = \text{span}\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\}$ and $\dim E_\lambda = 1 < 2$. The special thing about B is that we could read the eigenvalues off the diagonal: even though B isn't diagonalizable, it's as close to it as we can get, and this allowed us to read off the eigenvalues.

We say that B is in Jordan canonical form, which is something I don't feel like TeXing today. Check out Figure **??**.

The Jordan canonical form of a matrix consists of blocks of the same eigenvalue down the diagonal, with 1s to their right. You can go through some algebra and show that the characteristic polynomial of a Jordan canonical matrix that only has one block is what we expect:

$$\det(J_p(\lambda) - tI_p) = (-1)^p(t-\lambda)^p \tag{15.4}$$

which is the same as that of a diagonal matrix with λs all down the diagonal.

If A has $f_A = (-1)^n \prod_{i=1}^{k}(t-\lambda_i)^{m_i}$ and A can be brought to Jordan form, then m_i is the sum of sizes of the Jordan blocks $J_p(\lambda_i)$.

Definition 15.1. *We say that A can be brought to Jordan form if there exists an invertible matrix $Q_{n\times n}$ such that $A = QJQ^{-1}$ where J is in Jordan form.*

Note that this definition includes diagonalizable matrices, because a diagonal matrix is said to just consist of n 1×1 Jordan blocks.

Consider the matrix $J_p(\lambda) - \lambda I_p$, whose null space we want to find to get the eigenvectors of $J_p(\lambda)$.

$$J_p(\lambda) - \lambda I_p \begin{bmatrix} 0 & 1 & 0 & \dots & \dots \\ \dots & 0 & 1 & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & \dots & 0 & 1 \\ \dots & \dots & \dots & \dots & 0 \end{bmatrix} \tag{15.5}$$

Conceptually, this annihilates standard basis element x_1, sends x_2 to x_1, sends x_3 to x_2, and so on. This creates a cycle whose length corresponds to the multiplicity of λ. This is equivalent to saying there is a basis of F^n which is a union of cycles like this.

The x_is are not all eigenvectors: only the one that is annihilated, x_1, is an eigenvector. The rest are so-called generalized eigenvectors: in general, x_i is annihilated by $(A-\lambda I_n)^i$, because $A - \lambda I_n$ sends x_i to x_{i-1} so if we apply it i times it gets annihilated.

Definition 15.2. *$v \in F^n$ is called a generalized eigenvector of A corresponding to λ if $(A-\lambda I_n)^p v = 0$ for some $p > 0$.*

The rest of the lecture is an example with the matrix $A = \begin{bmatrix} 3 & 1 & -2 \\ -1 & 0 & 5 \\ -1 & -1 & 4 \end{bmatrix}$ which feels like a lot to actually write out, so I just won't.

Math 110: Linear Algebra **Fall 2019**

Lecture 16: Jordan form contd.

Lecturer: Edward Frenkel *4 November* *Aditya Sengupta*

Recall that a generalized eigenvector v has the property that for some p, $(A - \lambda I)^p v = \underline{0}$. This gives us the generalized eigenspace K_λ:

$$K_\lambda = \{v \in V \mid (A - \lambda I)^p v = 0 \text{ for some } p \in \{1, 2, \dots\}\}$$

When we talked about diagonalization, we introduced the notion of an eigenspace, so this is just a generalization of that.

Remark 16.1. *We can talk about linear transformations either in terms of the abstraction in which V over F is a vector space of dimension n, and $T : V \to V$ is a linear operator, or in terms of a matrix $A \in M_{n\times n}(F)$.*

We previously proved that if $T : V \to V$ is diagonalized, then $V = \oplus_{i=1}^k E_{\lambda_i}$. Now, we will prove that in general we have a similar decomposition in terms of K_{λ_i}s for any transformation whose matrix can be brought to Jordan form.

Here, we assume that the characteristic polynomial $f_T(t)$ splits over F by

$$f_T(t) = (-1)^n \prod_{i=1}^{k} (t - \lambda_i)^{m_i}$$

where $\lambda_i \neq \lambda_j$ for all $i \neq j$. We get a list $\lambda_1, \dots, \lambda_k$ of eigenvalues with multiplicities $m_1, \dots, m_k$. This is going to be more difficult to generalize, because previously we had that every vector in some E_{λ_i} satisfied $(T - \lambda_i I)v = 0$. But now, the equation that we have to satisfy won't always be the same.

If T is diagonalizable, we just construct the eigenbasis from that process that we already saw. If it is not, there are two levels of subdivision:

1. $V = \oplus_{i=1}^k K_{\lambda_i}$.
2. K_{λ_i} will have a basis that is a union of cycles.

Then, we can take $\beta_i = \gamma_1^i \sqcup \gamma_2^i \sqcup \cdots \sqcup \gamma_{q_i}^i$.

A generalized eigenvector has associated to it actual eigenvectors. If v satisfies $(A - \lambda I)^p v = 0$, then $(A - \lambda I)^{p-1} v$ is an actual eigenvector. We'll call the generalized eigenvector that starts the chain the <u>final vector</u> of the cycle, and the generalized eigenvector that's actually an eigenvector the <u>initial vector</u>.

The full proof of the Jordan canonical form decomposition being valid requires the Cayley-Hamilton theorem. Basically, this says that if you've got a matrix A that has a characteristic polynomial $\sum_i a_i t^i$, then $\sum_i a_i A^i = 0$: any matrix satisfies its own characteristic polynomial.

It's sort of a cornerstone of bad math pedagogy because high school linalg classes (wherever they exist) tend to just state it as fact and not justify it with anything about why it's true from a linear transforms perspective. So I was kind of excited to do it here. PG&E has deprived me of math.

Anyway, we've now built up enough machinery that we should just commit to actually doing something. First, here's the theorem we want to prove overall:

Theorem 16.2. *1. There exists a basis of V that is a union of cycles of T.*

2. $V = \oplus_{i=1}^{k} K_{\lambda_i}$.

3. $\dim K_{\lambda_i} = m_i$.

The proof of this relies on the following lemmas.

Lemma 16.3 (Linear independence of cycles)**.** *Let $\gamma_1, \gamma_2, \ldots, \gamma_q$ be cycles of generalized eigenvectors of T corresponding to λ, whose initial vectors are linearly independent. Then $\gamma_1 \sqcup \cdots \sqcup \gamma_q$ is linearly independent.*

Lemma 16.4 (Linear independence of vectors from different cycles)**.** *Suppose $v_i \in K_{\lambda_i}$ for $i = 1, \ldots, k$ such that $v_1 + v_2 + \cdots + v_k = \underline{0}$. Then $v_i = \underline{0}$ for all i.*

We proved Lemma **??** for the E_{λ_i}s before.

We can prove part 1 of Theorem **??** from these two lemmas, i.e. that there exists a basis of V that is a union of cycles of T.

Proof. We induct on $\dim V = n$.

Base case: $n = 1$. Then V is a one-dimensional vector space over F, and $T : V \to V$ is linear, meaning there exists some $a \in F$ such that $T(v) = av$.

Inductive step: let the inductive hypothesis be true for $\dim V = n - 1$. Recall that $\lambda_1, \ldots, \lambda_k$ are roots of $f_T(t)$. Let $W = R(T - \lambda_1 I)$; this is a subspace of V. We know that $\dim N(T - \lambda_1 I) > 0$, so by the dimension theorem, $\dim W < n$. We claim that W is $T-$invariant, i.e. that for all $w \in W, T(w) \in W$.

Since W is the range of $T - \lambda_1 I$, there exists some y such that $w = (T - \lambda_1 I)(y)$. Therefore

$$T(w) = T((T - \lambda_1 I)(y)) = (T - \lambda_1 I)(T(y)) \in W$$

Therefore $W \subset V$ and $\dim W < n = \dim V$.

By our inductive assumption, W has a basis of cycles of T. There will be cycles corresponding to eigenvalues $\lambda_1, \ldots, \lambda_k$ of T. So let's look at the cycles for λ_1 in W. Call them $\gamma_1', \ldots, \gamma_q'$. Since they are in $W = R(T - \lambda_1 I)$, we can extend each γ_i to a cycle in V by adjoining a vector $y_i \in V$ such that $(T - \lambda_1 I)y_i$ is equal to the final vector in the cycle γ_i. Therefore each $\gamma_i = \gamma_i' \cup \{y_i\}$, and the γ_is and γ_i's share initial vectors. Since the disjunctive union $\sqcup \gamma_i'$ is a subset of a basis, it follows that this set is linearly independent, therefore the initial vectors are linearly independent. By Lemma **??**, the disjunctive union of the γ_is is linearly independent, like we wanted. ∎

Math 110: Linear Algebra **Fall 2019**

Lecture 17: Generalized eigenspace basis, Markov chains

Lecturer: Edward Frenkel *6 November* *Aditya Sengupta*

I walked in ten minutes late, rip.

17.1 Generalized eigenspace basis proof

Previously, we constructed a disjunctive union of cycles γ_i for λ_1, and showed it was a basis. This was based on two lemmas, the second of which we'll prove here.

Lemma 17.1. *(Previously, this was Lemma 16.4, or in his document Theorem 2). Let K_{λ_i} be the generalized eigenspace on V corresponding to λ_i. Suppose we have $v_i \in K_{\lambda_i}$ such that $v_1 + v_2 + \cdots + v_k = \underline{0}$. Then each $v_i = 0$.*

Proof. We induct on k. This is trivially true for $k = 1$. Suppose we proved that if $v_1 + \cdots + v_m = \underline{0}$ then $v_1, v_2, \ldots, v_m = \underline{0}$. Then, we consider some combination such that $v_1 + \cdots + v_m + v_{m+1} = \underline{0}$, and we prove that $v_1 = \cdots = v_{m+1} = \underline{0}$.

First, let's consider the case where $v_i \in E_{\lambda_i}$. Apply $(T - \lambda_{m+1} I)$ to both sides of the initial equation:

$$(\lambda_1 + \lambda_{m+1})v_1 + \cdots + (\lambda_m - \lambda_{m+1})v_m = \underline{0} \tag{17.1}$$

The $m + 1$ term got murdered. Then, by the inductive hypothesis, all of $(\lambda_i - \lambda_{m+1})v_i = \underline{0}$. Since all the λ_is must be distinct, we get that $v_i = \underline{0}$ for all $i = 1, \ldots, m$.

Now, let's try to generalize this argument to $v_i \in K_{\lambda_i}$. Then, the annihilating operator is no longer $(T - \lambda_i I)$: it's now $(T - \lambda_i I)^{p_i}$ for some p_i. So, we have to apply $(T - \lambda_{m+1} I)^{p_{m+1}}$. This gives us

$$(T - \lambda_{m+1} I)^{p_{m+1}} v_1 + \cdots + (T - \lambda_{m+1} I)^{p_{m+1}} v_m \tag{17.2}$$

The $m + 1$ term is killed again, but now we don't know that all of these are eigenvectors. However, this turns out not to be necessary. All we need to know is that $(T - \lambda_{m+1} I)^{p_{m+1}}$ sends $v_i \neq \underline{0}$ to $v_i' \in K_{\lambda_i} \neq \underline{0}$.

To do this, we want to show that K_{λ_i} is invariant under $(T - \lambda_{m+1} I)^{p_{m+1}}$, and that it is an isomorphism for K_{λ_i}.

We know that for all $v \in K_{\lambda_i}$, $\exists p$ such that $(T - \lambda_i I)^p v = \underline{0}$. We claim that

$$(T - \lambda_i I)^p \left((T - \lambda_{m+1} I)v\right) = \underline{0} \tag{17.3}$$

that is, if v is killed by the power p of the operator, then $(T - \lambda_{m+1} I)v$ is as well. Because of this, we look at $(T - \lambda_{m+1} I)v$ as an element of K_{λ_i}, since it is annihilated. This suggests we look at the transformation restricted to K_{λ_i}:

$$(T + \lambda_{m+1} I)|_{K_{\lambda_i}} : K_{\lambda_i} \to K_{\lambda_i} \tag{17.4}$$

We want to prove that its null space is just $\{\underline{0}\}$: this comes out from $(T - \lambda_{m+1} I)w = 0$ which implies $Tw = \lambda_{m+1} w$. This can only hold along with the condition for being an element of K_{λ_i}, i.e. that $(T - \lambda_i I)^p v = 0$ for some p, if $w = 0$.

"It won't do much good for me to keep talking about it, so I'm moving on." □

17.2 Markov chains

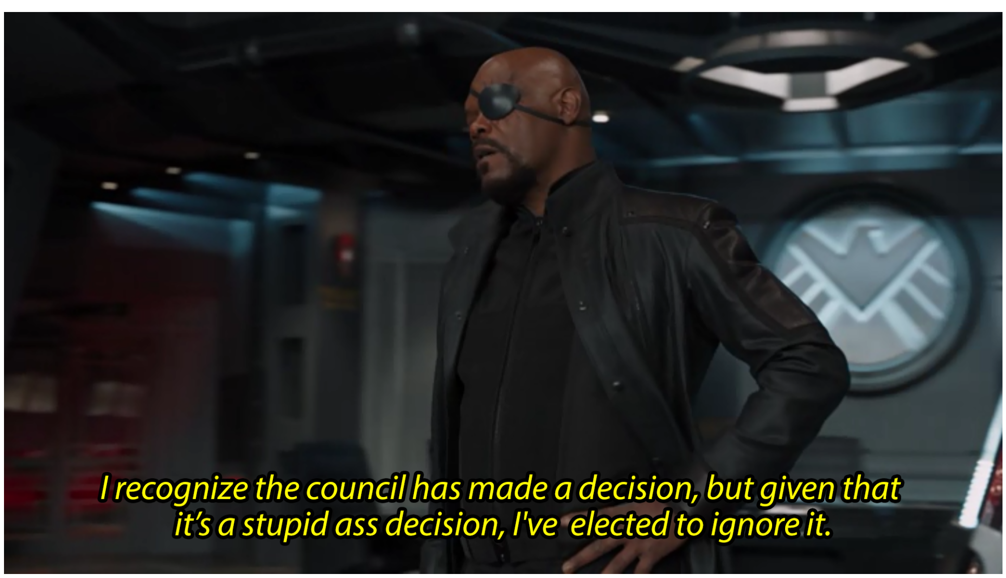
A Markov chain is an example of a stochastic process, i.e. a process that evolves over time according to certain probability laws.

A discrete-time Markov chain describes a process whose state changes in discrete steps and cycles between a finite set of possible states, with certain probabilities that are stationary and do not evolve over time. It also satisfies the Markov property, that the state only depends on the previous state:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1) = \mathbb{P}(X_{n+1} = j \mid X_n = i) = p_{ij} \tag{17.5}$$

In words, "the probability that the chain is in state j at time $n+1$ given that it was in state i at time n, state x_{n-1} at time $n-1$, ..., and state x_1 at time 1 is the same as the probability that the chain is in state j at time $n+1$ given that it was in state i at time n". More simply, the state-transition rules for the next step only depend on your current state, not on your history. We denote the quantity $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ by p_{ij}, the probability of transitioning from i to j.

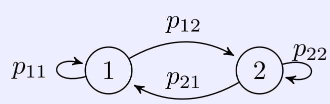
Note that he's using the opposite convention to the convention that's used basically everywhere. Everyone else says p_{ij} is the probability of going i to j, but he uses it as the probability of going j to i.



Because you have to jump to something (self-loops p_{ii} are allowed), we can say that $\sum_{j=1}^{N} p_{ij} = 1$ and that $0 \leq p_{ij} \leq 1$. We can encode these in a transition matrix A:

$$A = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \tag{17.6}$$

Example 17.1. Suppose we want to model the weather over multiple days. Let's encode "it is sunny" in state 1, and "it is rainy" in state 2.



Note again that I'm flipping his convention so that it matches the convention that's used in every other online resource about Markov chains. This also means that I'm transposing every matrix he writes.

Suppose the transition matrix is $A = \begin{bmatrix} 0.8 & 0.2 \\ 0.7 & 0.3 \end{bmatrix}$. Then, if today, at $t = 0$, we're in state 1, then we can find the probabilities it'll be sunny or rainy on the next day by left multiplying a row vector of the current distribution. This is the purpose of his formulation: if you use the definition $j \to i$ has p_{ij}, you get to do the usual right multiplication, but with the canonical definition, you have to think about distributions as row vectors.

$$\begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}0.8 & 0.2\\ 0.7 & 0.3\end{bmatrix} = \begin{bmatrix}0.8 & 0.2\end{bmatrix} \tag{17.7}$$

Then, at $t = 2$, the probabilities are

$$\begin{bmatrix}0.8 & 0.2\end{bmatrix}\begin{bmatrix}0.8 & 0.2\\ 0.7 & 0.3\end{bmatrix} = \begin{bmatrix}0.78 & 0.22\end{bmatrix} \tag{17.8}$$

□

Is it possible to predict the long-term behaviour of a Markov chain?

$$\vec{p_i} = \vec{p_{i-1}}A = \vec{p_{i-2}}A^2 = \cdots = \vec{p_0}A^i \tag{17.9}$$

We can predict long-term behaviour if we write A in Jordan canonical form:

$$A = QJQ^{-1} \implies A^i = QJ^iQ^{-1} \tag{17.10}$$

If J is diagonal, i.e. it's $\lambda_1, \ldots, \lambda_n$ down the diagonal and 0 everywhere else, then D^i is just $\lambda_1^i, \ldots, \lambda_n^i$ down the diagonal and 0 everywhere else, which is easy to compute.

Example 17.2. In our previous example,

$$\det\begin{bmatrix}0.8-t & 0.2\\ 0.7 & 0.3-t\end{bmatrix} = (t-1)(t-0.1) \tag{17.11}$$

If we take the null space in both case, we get that the eigenvector for $\lambda = 1$ is $\begin{bmatrix}\frac{7}{9} & \frac{2}{9}\end{bmatrix}$ and the eigenvector for $\lambda = 0.1$ is $\begin{bmatrix}1 & -1\end{bmatrix}$. (This bears checking, because I did it with my convention and might have messed up the transpose.) We construct Q^{-1} by stacking up these row vectors - I think this is Q^{-1} instead of Q because of some weirdness with transposing. Therefore, we get

$$A = \begin{bmatrix}1 & \frac{2}{9}\\ 1 & -\frac{7}{9}\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 0.1\end{bmatrix}\begin{bmatrix}\frac{7}{9}\frac{2}{9} & \\ 1 & -1\end{bmatrix} \tag{17.12}$$

(I verified this on WolframAlpha, so we're good)

If we raise both sides to the ith power, we get

$$A^i = \begin{bmatrix} 1 & \frac{2}{9} \\ 1 & -\frac{7}{9} \end{bmatrix} \begin{bmatrix} 1^i & 0 \\ 0 & 0.1^i \end{bmatrix} \begin{bmatrix} \frac{7}{9}\,\frac{2}{9} & \\ 1 & -1 \end{bmatrix} \tag{17.13}$$

As $i \to \infty$, J goes to $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, therefore our overall i is

$$\begin{bmatrix} 1 & \frac{2}{9} \\ 1 & -\frac{7}{9} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{7}{9}\,\frac{2}{9} & \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} \frac{7}{9} & \frac{2}{9} \\ \frac{7}{9} & \frac{2}{9} \end{bmatrix} \tag{17.14}$$

□

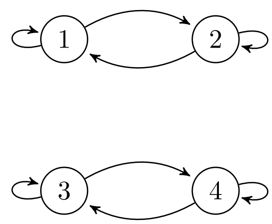
This is true for a general transition matrix satisfying regularity, which means that there exists $m \in \{1, 2, \dots\}$ such that $(A^m)_{ij} \neq 0$ for all i, j: every state is accessible from every other one. If this isthe case, there is a unique probability vector such that $\vec{p} = \vec{p}A$: this is the stationary distribution. In this case, every row of the Markov chain is this vector.

Math 110: Linear Algebra **Fall 2019**

Lecture 18: Markov Chains contd.

Lecturer: Edward Frenkel *13 November* *Aditya Sengupta*

To analyze (something about) a Markov chain, we impose regularity, i.e. we state that there exists some $m \in \mathbb{N}/\{0\}$ such that all entries of A^m are strictly positive. This prevents Markov chains like this one:



Here, the class $\{1, 2\}$ does not communicate with the class $\{3, 4\}$. A regular Markov chain has all of its states connected, i.e. for every pair of states i, j, there exists some n such that $(A^n)_{i,j} > 0$: the probability of transitioning i to j in n steps is nonzero.

Theorem 18.1. *Let A be a regular transition matrix. Let λ represent an eigenvalue of A. Then:*

1. $|\lambda| < 1$ *or* $\lambda = 1$.
2. *The eigenspace corresponding to* $\lambda = 1$ *is 1-dimensional.*
3. E_1 *contains a unique probability vector* $\vec{p}$.
4. *The unique Jordan block for* $\lambda = 1$ *is 1x1.*
5. *The limit* $\lim_{m \to \infty} A^m$ *exists and is equal to* $\begin{bmatrix} \vec{p} \\ \vec{p} \\ \vdots \end{bmatrix}$.

Proof. Let $B = A^T$. We know that A and B have the same Jordan form. (Note that I'm still flipping his convention, so B for me is what he considers the normal state-transition matrix A.) We can see this by

$$\det(A - tI) = \det((A - tI)^T) = det(A^T - tI) \tag{18.1}$$

The characteristic polynomials therefore both admit the same representation $(-1)^n \prod_{i=1}^{k} (t - \lambda_i)^{m_i}$. Since the eigenvalues and their multiplicities are the same, we know that

$$\dim N((A - \lambda_i I)^p) = n - \operatorname{rank}(A - \lambda_i I)^p = n - \operatorname{rank}(A^T - \lambda_i I)^p = \dim N((A^T - \lambda_i I)^p) \tag{18.2}$$

i.e. the Jordan forms are the same.

Alternatively, we can say that

$$A = QJQ^{-1} \implies A^T = (Q^T)^{-1} J^T Q^T \tag{18.3}$$

If J^T were similar to J, i.e. there existed some R such that $J^T = RJR^{-1}$, we could say A and A^T shared a Jordan form (A^T similar to J^T similar to J similar to A.)

We can see why this is true by looking at the typical Jordan form: on the diagonal, J and J^T are the same, and off the diagonal, a Jordan matrix has 1s below/to the left of each eigenvalue. If we moved those to above/to the right of each eigenvalue, we'd get the transposed Jordan form, which we can achieve just by a flipped version of the identity matrix, with 1s down the right diagonal.

All this to say he's now using the standard convention for Markov chains, and that either formalism works.

Note that the vector $\vec{u} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}$ is a left eigenvector of A with eigenvalue 1: $\vec{u}A = \vec{u}$. Suppose there is another eigenvector of A with eigenvalue 1, linearly independent from $\vec{u}$. Denote this by $\vec{v} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$. Take $\vec{v} - v_k\vec{u} = \begin{bmatrix} v_1 - v_k & v_2 - v_k \dots & 0 & \dots & v_n - v_k \end{bmatrix}$. Then

$$(\vec{v} - v_k\vec{u})A^m = (\vec{v} - v_k\vec{u}) \tag{18.4}$$

For some m, all entries of A^m are greateer than 0. But if we multiply the 0 element thus generated by A^m, we'll get something greater than 0 unless all the other entries are 0, which is a contradiction. Therefore $\vec{u}$ spans the $1-$eigenspace.

It remains to be shown that any eigenvalue that is not 1 must have a magnitude less than 1. We know that $\sum_{j=1}^n A_{ij} = 1$ for all k. Suppose $\vec{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$ such that $\vec{x}A = \lambda\vec{x}$. We want to prove that $|\lambda| < 1$.

$$\sum_{j=1}^n A_{ij} x_j = \lambda x_i \tag{18.5}$$

Let $k = \arg\max |x_i|$, i.e. the index of the maximum element of $\vec{x}$. Then

$$\begin{aligned}
|\lambda - A_{kk}||x_k| &= |\lambda x_k - A_{kk} x_k| && (18.6)\\
&= \left| \sum_{[} j = 1]^n A_{kj} x_j A_{kk} x_k \right| && (18.7)\\
&= \left| \sum_{j=1}^n A_{kj} x_j \right| && (18.8)\\
&\le \sum_{j \ne k} |A_{kj}||x_j| \le \sum_{j \ne k} A_{kj} |x_k| = (1 - A_{kk})|x_k| && (18.9)
\end{aligned}$$

Therefore, we get that $|\lambda - A_{kk}| \le |1 - A_{kk}|$. Therefore, we can use the triangle inequality to show what we want:

$$|\lambda| = |(\lambda - A_{kk}) + A_{kk}| \leq |\lambda - A_{kk}| + |A_{kk}| \tag{18.10}$$
$$\leq |1 - A_{kk}| + |A_{kk}| = 1 - A_{kk} + A_{kk} = 1 \tag{18.11}$$

(we can drop the absolute values because A is nonnegative everywhere).

It remains to show that $|\lambda| = 1 \implies \lambda = 1$. This just comes out of A_{kk} being positive: $|\lambda - A_{kk}| = |-1 - A_{kk} = 1 + A_{kk}$ which can't be less than $1 - A_{kk}$ like we want. (we assume WLOG that $A_{kk} > 0$.)

We've shown the first two statements. The third one comes out of the $1-$eigenspace being one-dimensional, meaning there's a unique way to normalize it. To show the fourth one, we note that if A is a transition matrix, then so is A^m for any $m \in \mathbb{N}$. Therefore, A^m is bounded, meaning that J^m can't grow to infinity as it's similar to A^m. If J contains a Jordan block with $\lambda = 1$ of size greater than 1, we get a contradiction. To see this, we note that

$$\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}^m = \begin{bmatrix} \lambda^m & m\lambda^{m-1} \\ 0 & \lambda^m \end{bmatrix} \xrightarrow[m \to \infty]{\infty} \tag{18.12}$$

Similarly, any Jordan block with $\lambda = 1$ of size $R > 1$ blows up to infinity as $m \to \infty$.

Finally, we note that A^m is a transition matrix, and so is the limiting matrix $L = \lim\limits_{m \to \infty} A^m$. We want to enforce that $LA = L$, i.e. $LA = \lim\limits_{m \to \infty} A^{m+1} = L$. We get this just by stacking up all the row vectors in L, denoting them y_i:

$$LA = \begin{bmatrix} \vec{y_1} A \\ \vec{y_2} A \\ \vdots \\ \vec{y_n} A \end{bmatrix} = L \tag{18.13}$$

Therefore $\vec{y_i} A = \vec{y_i}$, so it's an eigenvector with eigenvalue 1. That means that each $\vec{y_i}$ must be the same as $\vec{p}$. $\square$

Math 110: Linear Algebra **Fall 2019**

Lecture 19: Inner Product Spaces

Lecturer: Edward Frenkel *18 November* *Aditya Sengupta*

19.1 Inner Products over $\mathbb{R}$ and $\mathbb{C}$

In $\mathbb{R}^n$, we have the notion of a magnitude, and of orthogonality, that are connected to the idea of an inner product. We'd like to generalize these!

The squared-magnitude of $\vec{x} \in \mathbb{R}^n$ is $\|\vec{x}\| = \vec{x} \cdot \vec{x}$, and two vectors in $\mathbb{R}^n$ are orthogonal if they dot to 0: $\vec{x} \perp \vec{y} \iff \vec{x} \cdot \vec{y} = 0$. To generalize these, we introduce the following definition of an inner product:

Definition 19.1. *Let V be a vector space over $\mathbb{R}$ or $\mathbb{C}$. An inner product on V is an operation $V \times V \to F, (\vec{x}, \vec{y}) \to \langle x, y \rangle$ satisfying the axioms*

(1) $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$.

(2) $\langle c \cdot x, y \rangle = c\langle x, y \rangle$.

(3) $\langle x, y \rangle = \langle y, x \rangle$, *or for the complex case,* $\langle x, y \rangle = \langle y, x \rangle^*$.

(4) $\langle x, x \rangle \geq 0$ *for all* $x \in V$.

For $V = \mathbb{R}^n$ over $F = \mathbb{R}$, the standard inner product $\langle \vec{x}, \vec{y} \rangle = \sum_i x_i y_i$ satisfies all of these, but for $V = \mathbb{C}^n$ over $\mathbb{C}$, this doesn't necessarily work. It satisfies all the axioms, but it doesn't satisfy the property that we want, that the inner product of a vector with itself gives us the magnitude of the vector. Therefore, over $\mathbb{C}$, we take $\langle \vec{x}, \vec{y} \rangle = \sum_i x_i y_i^*$.

Lemma 19.1. *If $F = \mathbb{R}$, then the first two properties of an inner product also hold on the second argument:*

$$\langle y, x + z \rangle = \langle y, x \rangle + \langle y, z \rangle \tag{19.1}$$
$$\langle y, cx \rangle = c\langle y, x \rangle \tag{19.2}$$

That is, the inner product over $\mathbb{R}$ is bilinear: it's linear in both arguments.

If $F = \mathbb{C}$, we have semilinearity: $\langle y, x + z \rangle = \langle y, x \rangle + \langle y, z \rangle$ and $\langle y, cx \rangle = c^* \langle y, x \rangle$. The inner product over $\mathbb{C}$ is therefore sesquilinear: it's almost bilinear, but has to be "twisted" a little bit to get it there.

19.2 Inner Products over Other Spaces

Let $V = C([0, 1])^\circ$ be the space of real-valued continuous functions on $[0, 1]$. Suppose $f(t), g(t) \in V$; then, set

$$\langle f, g \rangle = \int_0^1 f(t)g(t)dt \tag{19.3}$$

This satisfies the fourth axiom, because $\langle f, f \rangle = \int_0^1 f(t)^2 dt$ which is always nonnegative, and is positive only if $f(t) = 0$.

I've tuned out because I just can't do this today.

Math 110: Linear Algebra **Fall 2019**

Lecture 20: Topic

Lecturer: Edward Frenkel *November 20* *Aditya Sengupta*

I have no idea what happened here.

Math 110: Linear Algebra **Fall 2019**

Lecture 21: Inner Product Spaces as Dual Spaces

Lecturer: Edward Frenkel *November 25* *Aditya Sengupta*

Recall that for any finite-dimensional vector space V over F, we have defined its dual, V^*, which gives some meaning to the transpose of a linear transformation: $T : V \to V$ goes to $T^t : V^* \to V^*$. To make sense of this equation, we need an isomorphism $V \simeq V^*$. If F is $\mathbb{R}$ or $\mathbb{C}$, we can identify what this isomorphism is through an inner product on V through some relationship $T^t \simeq T^*$. Operators that satisfy this are called Hermitian.

Suppose V is a finite-dimensional vector space over $\mathbb{R}$ with an inner product $\langle \cdot, \cdot \rangle$. Construct a linear transformation

$$P_{\langle \cdot,\cdot \rangle} : V \to V^*, y \to f_y \tag{21.1}$$

where $f_y(x) = \langle x, y \rangle$. We can verify that f_y is a linear functional by showing that $f_y(x+z) = f_y(x) + f_y(z)$ and $f_y(cx) = cf_y(x)$. Therefore, $P_{\langle \cdot,\cdot \rangle} : V \to V^*$ is a linear transformation. We cllaim this is an isomorphism. Since $\dim(V) = \dim(V^*)$, it is sufficient to show that $P_{\langle \cdot,\cdot \rangle}$ is onto. Pick $g : V \to \mathbb{R}$. Then, we want to $y \in V$ such that $g = f_y$.

We know that every finite-dimensional inner product space has an orthonormal basis. Therefore, we choose such a basis, $\{x_1, \dots, x_n\}$. We need to choose y such that for all $x \in V$, $g(x) = f_y(x)$. We express every $z \in V$ in the orthonormal basis:

$$z = \sum_{i=1}^{n} \langle z, x_i \rangle x_i \tag{21.2}$$

So for the y we want, we can decompose it to a linear combination with weights of the form $\langle y, x_i \rangle$ at which it must match g. Therefore, we take

$$y = \sum_{i=1}^{n} g(x_i) x_i \tag{21.3}$$

Since this choice satisfies the requirement on a basis, it satisfies it everywhere, so $g = f_y$. Therefore $P_{\langle \cdot,\cdot \rangle}$ is an isomorphism.

Now, given $T : V \to V$, we have $T^t : V^* \to V^*$ and we transport T^t to an operator $T^* : V \to V$, such that $T^t g = g \circ T$. With inner products, this is given by a simple formula that we see in the next definition.

Definition 21.1. *The adjoint operator T^* to an operator T is defined by the formula*

$$\langle T(x), y \rangle = \langle x, T^*(y) \rangle \;\forall x, y \in V \tag{21.4}$$

We can show that such a T^* exists:

$$T^*(y) = \sum_{j=1}^{n} \langle y, T(x_j) \rangle x_j \tag{21.5}$$

To check the above, it is sufficient to check it over a basis.

$$\langle T(x_i), y \rangle = \langle x_i, T^*(y) \rangle = \langle x_i, \sum_{j=1}^{n} \langle y, T(x_j) \rangle x_j \rangle = \langle y, T(x_i) \rangle \tag{21.6}$$

Further, we can show it is unique. Suppose we had another operator $U : V \to V$ such that $\langle T(x), y \rangle = \langle x, U(y) \rangle$. Then $\langle x, T^*(y) \rangle = \langle x, U(y) \rangle$ for all $x, y \in V$. But then for all $y \in V$, $T^*(y)$ and $U(y)$ have thee same inner products with alll $x \in V$.

Lemma 21.1. *Suppose $u, v \in V$ are such that $\langle x, u \rangle = \langle x, v \rangle$ for all $x \in V$. Then $u = v$.*

Proof. We can just use linearity:

$$\langle x, u - v \rangle = 0 \ \forall x \tag{21.7}$$

Pick $x = u - v$: then $\|u - v\| = 0$, so $u - v = \vec{0}$. □

Therefore, since T^* and U match everywhere, they must be the same operator. A couple examples here.

Math 110: Linear Algebra		**Fall 2019**

Lecture 22: Topic

Lecturer: Edward Frenkel *November 27* *Aditya Sengupta*

Skipped.

Math 110: Linear Algebra **Fall 2019**

Lecture 23: Topic

Lecturer: Edward Frenkel *4 December* *Aditya Sengupta*

Let V be a finite-dimensional inner product space over $\mathbb{R}$ or $\mathbb{C}$, and let $T : V \to V$ be a self-adjoint operator.

Theorem 23.1. *1. T is diagonalizable.*

2. There is an orthonormal basis in V that is an eigenbasis of T.

3. All eigenvalues of T are real.

Proof. We induct on $\dim V$. This is trivially true if $\dim V = 1$, so suppose inductively there is an orthonormal basis of V that is an eigenbasis of T in the case $\dim V = k$. Now, consider the case $\dim V = k + 1$ We know that T has at least one eigenvector x_1, such that $T(x_1) = \lambda_1 x_1$. Therefore, let $W = \text{span}\{x_1\} \subset V$. $\dim W = 1$. Then,

$$W^\perp = \{y \in V \mid \langle x, y\rangle = 0 \ \forall x \in W\} \tag{23.1}$$

We claim that $W^\perp$ is $T-$invariant: for all $y \in W^\perp, T(y) \in W^\perp$. We can show this as follows:

$$\langle x_1, T(y)\rangle = \langle T(x_1), y\rangle = \langle \lambda_1 x_1, y\rangle = \lambda_1 \cdot 0 = 0 \tag{23.2}$$

Recall that $V = W \oplus W^\perp$. Therefore, the union of bases of W and $W^\perp$ is a basis of V. The dimension of the basis of W is 1, and that of $W^\perp$ is k; by the inductive hypothesis, we can choose an orthonormal basis of $W^\perp$ that is an eigenbasis of T restricted to $W^\perp$. So by induction, we have an orthonormal basis of V that is an eigenbasis of T.

It remains to show that all eigenvalues of T are real. Let λ be an eigenvalue of T. Then $T(x) = \lambda x$ for some $x \neq 0$. Then

$$\lambda\langle x, x\rangle = \langle \lambda x, x\rangle = \langle T(x), x\rangle = \langle x, T(x)\rangle = \langle x, \lambda x\rangle = \bar{\lambda}\langle x, x\rangle \tag{23.3}$$

Therefore $\lambda = \bar{\lambda}$, so it must be real. □

Given a self-adjoint $T : V \to V$, we can construct a corresponding orthonormal eigenbasis. Consider the characteristic polynomial,

$$f_T(t) = (-1)^n \prod_{j=1}^{k} (t - \lambda_j)^{m_j} \tag{23.4}$$

Lemma 23.2. *Let $T : V \to V$ be self-adjoint on an inner product space V. Suppose x_1, x_2 are eigenvectors of T with distinct eigenvalues. Then $\langle x_1, x_2\rangle = 0$.*

Proof.

$$\langle T(x_1), x_2 \rangle = \langle x_1, T(x_2) \rangle \tag{23.5}$$
$$\lambda_1 \langle x_1, x_2 \rangle = \bar{\lambda_2} \langle x_1, x_2 \rangle \tag{23.6}$$

Since both eigenvalues are real, we get that

$$(\lambda_1 - \lambda_2)\langle x_1, x_2 \rangle = 0 \tag{23.7}$$

Since we chose $\lambda_1 \neq \lambda_2$, we get $\langle x_1, x_2 \rangle = 0$. ☐

Corollary 23.3. *$E_{\lambda_j} \perp E_{\lambda_p}$ for all $j \neq p$.*

Now recall that $V = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_k}$ (these are pairwise orthogonal). Therefore, to construct an orthonormal eigenbasis of T in V, we choose any orthonormal basis in each $E_{\lambda_j}, j = 1, \ldots, k$, and call it β_j. Then we take $\beta = \cup_{j=1}^{k} \beta_j$.

An operator T over an inner product space V is called normal if $TT^* = T^*T$.

For any linear operator, let

$$T_1 = \frac{1}{2}(T + T^*), T_2 = \frac{1}{2i}(T - T^*) \tag{23.8}$$

Then T_1 and T_2 are both self-adjoint, and $T = T_1 + iT_2$. If we could find a joint eigenbasis β for T_1 and T_2, this would be an eigenbasis for T.

Further, T is normal if and only if T_1 and T_2 commute:

$$T_1 T_2 = T_2 T_1 \tag{23.9}$$
$$(T + T^*)(T - T^*) = (T - T^*)(T + T^*) \tag{23.10}$$
$$2T^*T = 2TT^* \tag{23.11}$$

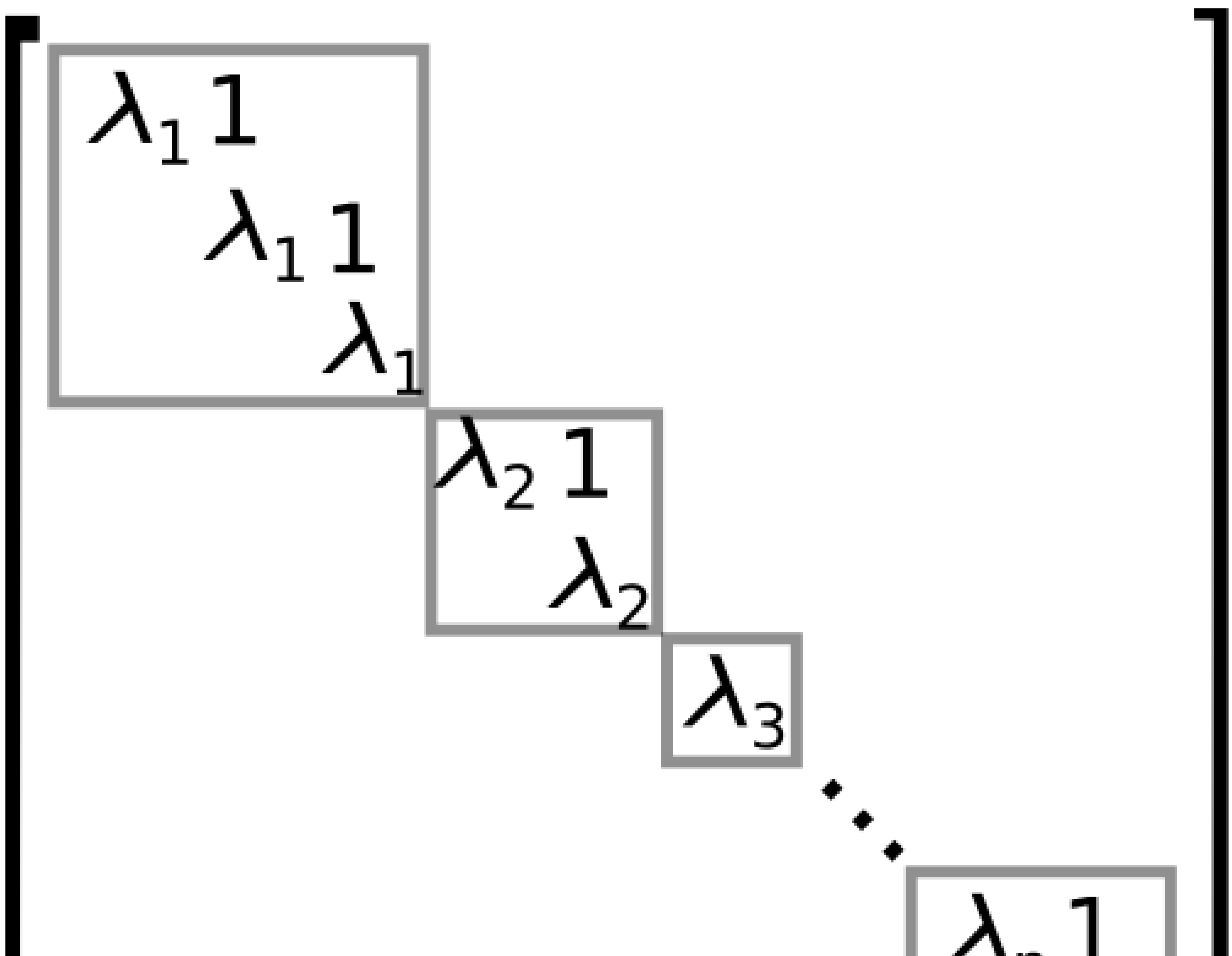


Figure 15.5: An entirely too large picture of the Jordan canonical form

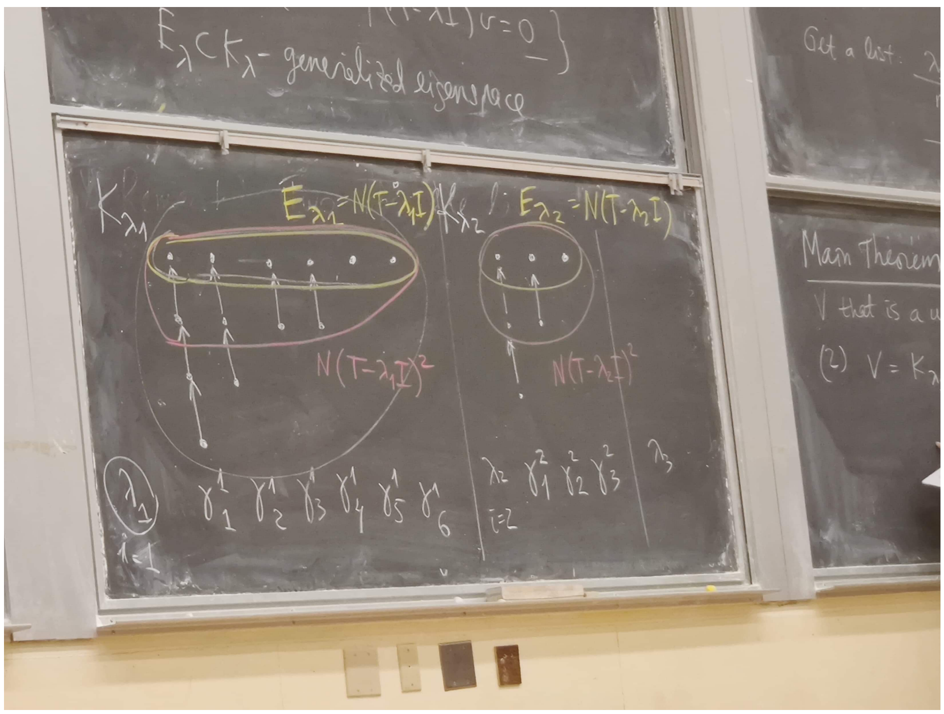


Figure 16.6: The “dot diagram” for arbitrary cycles