Math 110 Notes

I scribed these notes during the Summer 2021 iteration of Math 110, which was taught by Dr. Mira Peterka. We followed *Linear Algebra Done Right* by Sheldon Axler pretty closely and as such, you may notice several references to that textbook throughout these notes. In the interest of time, we decided to skip several sections from LADR, like 3E (product and quotient spaces), 3F (dual spaces) and 4 (polynomials). Moreover, we only skimmed chapters 9 (Operators on Real Vector Spaces) and 10 (Trace and Determinant) and did not cover them in as much detail as the first 8 chapters.

Contents

1 Lecture 1

1.1 Vector Spaces

Definition 1.1: Addition
A (vector) addition on a set V is a function that assigns an element $u + v \in V$ for each pair of elements $u, v \in V$.

Definition 1.2: Multiplication
A (scalar) multiplication on a set V is a function that assigns an element $\lambda v \in V$ for each $\lambda \in \mathbb{F}$ (here $\mathbb{F}$ is a field) and $v \in V$.

The fields that we will commonly work with in this course are $\mathbb{R}$ (real numbers) and $\mathbb{C}$ (complex numbers). However, the properties that we will prove should hold for any abstract field $\mathbb{F}$ that may be finite (ex. a Galois Field) or infinite.

Definition 1.3: Vector Space
A vector space over $\mathbb{F}$ is a set V along with an addition and a scalar multiplication on V such that the following properties hold:

- Commutativity: $u + v = v + u$ for all $u, v \in V$
- Associativity: $(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and all $a, b \in \mathbb{F}$
- Additive Identity: there exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$
- Additive inverse: there exists a $w \in V$ such that $v + w = 0$ for all $v \in V$
- Multiplicative identity: there exists an element $1 \in \mathbb{F}$ such that $1v = v$ for all $v \in V$
- Distributive properties: $a(u + v) = au + av$ and $(a + b)v = av + bv$ for all $u, v \in V$ and for all $a, b \in \mathbb{F}$

Example 1.1
Here are some common vector spaces:

1. $\mathbb{R}^n = \{(u_1, \dots, u_n) \mid u_j \in \mathbb{R}\}$ is a vector space over $\mathbb{R}$ for the operations:

$$(u_1, \dots, u_n) + (v_1, \dots, v_n) = (u_1 + v_1, \dots, u_n + v_n)$$
$$\lambda(u_1, \dots, u_n) = (\lambda u_1, \dots, \lambda u) \text{ for } \lambda \in \mathbb{R}$$

   The additive identity is $0 = (0, \dots, 0) \in \mathbb{R}^n$.

2. $\mathbb{F}^n = \{(u_1, \dots, u_n) \mid u_j \in \mathbb{F}\}$ is a vector space over $\mathbb{F}$ for the same operations defined above.
3. $\mathbb{F}^\infty = \{(u_1, u_2, \dots) \mid u_j \in \mathbb{F}, j \leq \infty\}$ is a vector space over $\mathbb{F}$ for the operations:

$$(u_1, u_2, \dots) + (v_1, v_2, \dots) = (u_1 + v_1, u_2 + v_2, \dots)$$
$$\lambda(u_1, u_2, \dots) = (\lambda u_1, \lambda u_2, \dots) \text{ for } \lambda \in \mathbb{F}$$

   The additive identity is $0 = (0, 0, \dots) \in \mathbb{F}^\infty$.

Example 1.2
Let S be a set and $\mathbb{F}$ be a field. Then, $\mathbb{F}^S = \{f : S \mapsto \mathbb{F}\}$ denotes the set of all functions from S to $\mathbb{F}$ and is a vector space over the following operations:

$$(f + g)(x) = f(x) + g(x) \quad \forall x \in S$$

$$(\lambda f)(x) = \lambda \cdot f(x) \qquad \forall x \in S, \forall \lambda \in \mathbb{F}$$

The additive identity is the zero function: $0(x) = 0$ for all $x \in S$.

Example 1.3

Let $V = \mathbb{R}_+ = \{x \in \mathbb{R} \mid x > 0\}$. Define "$x + y$" $= xy$ for $x, y \in V$ as an addition that is generally considered ordinary multiplication. Similarly, define "cx" $= x^c$ for $c \in \mathbb{R}$ and $x \in V$ as a multiplication that is generally considered ordinary exponentiation. With these operations, V is a vector space over $\mathbb{R}$ since

$$\begin{aligned} \text{“}c(x+y)\text{”} &= (xy)^c \\ &= x^c y^c \\ &= \text{“}cx + cy\text{”} \end{aligned}$$

Moreover, 1 is the additive identity (the 0 of this vector space) since "$1 + x$" $= 1 \cdot x =$ "x" for all $x \in V$.

1.2 Subspaces

Definition 1.4: Subspace

Suppose V is a vector space over $\mathbb{F}$. A subset $U \subseteq V$ is a subspace of V iff it

- has the additive identity: $0 \in U$
- satisfies vector addition: $u, v \in U \implies u + v \in U$
- satisfies scalar multiplication: $\lambda \in \mathbb{F}, u \in U \implies \lambda u \in U$

Note 1.1

A subspace U of V is itself a vector space over the field $\mathbb{F}$ for the same addition and scalar multiplication operations defined for V.

Example 1.4

Let $V = \mathbb{R}^2$. Then, V has exactly one "2-dimensional" subspace, namely V itself. Moreover, each line passing through the origin is a "1-dimensional" subspace of V. Finally, V has exactly one "0-dimensional" subspace, which is $\{0\} = \{(0,0)\} \in V$. We will formalize the notion of dimension later in the course.

Example 1.5

Here are some cases that are *not* examples of subspaces. Let $V = \mathbb{R}^2$ for both cases.

- Let $U = \{(x, y) \mid y \geq 0\}$ be the closed upper half of $\mathbb{R}^2$. Then,
  1. U contains the additive identity $0 = (0, 0)$ since $0 \geq 0$
  2. U is closed under addition, i.e., $u + v \in U$ for $u, v \in U$
  3. U is *not* closed under scalar multiplication: note that $(1, 1) \in U$ and $-1 \in \mathbb{R}$ but $-1 \cdot (1, 1) = (-1, -1) \notin U$
- Let $W = \{(x, y) \mid xy \geq 0\}$ be the first and third quadrant of $\mathbb{R}^2$. Then,
  1. W contains the additive identity $0 = (0, 0)$ since $0 \cdot 0 \geq 0$
  2. W is closed under scalar multiplication
  3. W is *not* closed under addition in V: note that $u = (1, 2) \in W$ and $w = (-2, -1) \in W$ but $u + w = (-1, 1) \notin W$

Definition 1.5: Sum of Subsets
Let V be a vector space over $\mathbb{F}$ and $U_1, \ldots, U_k$ be subsets of V. Then, we define

$$U_1 + \cdots + U_k = \left\{u_1 + \cdots + u_k \mid u_j \in U_j \text{ for } j \leq k\right\}$$

Note 1.2
If $U_1, \ldots, U_k$ are subspaces of V, then $U_1 + \cdots + U_k$ is a subspace of V too.

Definition 1.6: Direct Sum
Suppose $U_1, \ldots, U_k$ are subspaces of V. The sum $U_1 + \cdots + U_k$ is called a direct sum if each element of $U_1 + \cdots + U_k$ can be written in one and one way only as a sum $u_1 + \cdots + u_k$ where each $u_j \in U_j$ for $j \leq k$. If $U_1 + \cdots + U_k$ is a direct sum, we can denote it by $U_1 \oplus U_2 \oplus \cdots \oplus U_k$.

Theorem 1.1
Let U, W be subspaces of V. Then $U + W$ is a direct sum iff $U \cap W = \{0\}$.

Proof: We will prove both parts of the biconditional:

- Suppose $x \in U \cap W$ and $U \cap W$ is a direct sum. Then, we write

$$\begin{aligned} x &= u + 0 \quad u \in U,\ 0 \in W \\ x &= 0 + w \quad 0 \in U, w \in W \end{aligned}$$

  Since $U + W$ is a direct sum, x can only be written in one way which makes $u = 0$ and $w = 0$. So $x = u + w = 0 + 0 = 0$ and $U \cap W = \{0\}$.

- Now suppose $u_1 + w_1 = x = u_2 + w_2$ where $u_1, u_2 \in U, w_1, w_2 \in W$. Then, $u_1 - u_2 = w_2 - w_1$ for $u_1 - u_2 \in U$ and $w_2 - w_1 \in W$. However, $U \cap W = \{0\}$ so $u_1 - u_2 = w_2 - w_1 = 0$ which implies $u_1 = u_2$ and $w_1 = w_2$. Thus, there is only one way to write $x = u_1 + w_1$, making $U + W$ a direct sum.

Note 1.3
This theorem works well for 2 subspaces but it does not generalize for 3 or more subspaces.

Example 1.6: 1C Exercise 20
Let $U = \left\{(x, x, y, y) \in \mathbb{F}^4 \mid x, y \in \mathbb{F}\right\}$. Find a subspace $W \subseteq \mathbb{R}^4$ so that $U \oplus W = \mathbb{F}^4$. There are many possible correct answers. Two of them are:

- $W = \left\{(x, -x, y, -y) \in \mathbb{F}^4 \mid x, y \in \mathbb{F}\right\}$
- $W = \left\{(0, x, 0, y) \in \mathbb{F}^4 \mid x, y, \in \mathbb{F}\right\}$

Example 1.7: 1C Exercise 24
Let $V = \mathbb{R}^{\mathbb{R}}$. Define U_e to be the set of even functions from $\mathbb{R}$ to $\mathbb{R}$ and U_o to be the set of odd functions. Show that $\mathbb{R}^{\mathbb{R}} = U_e \oplus U_o$.

Proof: Showing that U_e and U_o are subspaces of $\mathbb{R}^{\mathbb{R}}$ is left as an exercise to the reader.

- We will appeal to the theorem above and show that $U_e \cap U_o = \{0(x)\}$, the zero function. If $f \in U_e \cap U_o$, then $f(-x) = f(x)$ and $f(-x) = -f(x)$ for all $x \in \mathbb{R}$. Since $f(x) = -f(x)$ for all $x \in \mathbb{R}$, this implies that $f(x) = 0(x)$ so $U_e \cap U_o = \{0(x)\}$ as expected.
- What does the direct sum look like? Suppose $f \in \mathbb{R}^{\mathbb{R}}$.

Define $f_{\text{even}}(x) = \frac{f(x)+f(-x)}{2}$. Since

$$f_{\text{even}}(-x) = \frac{f(-x)+f(-(-x))}{2} = \frac{f(-x)+f(x)}{2} = f_{\text{even}}(x)$$

we get that $f_{\text{even}} \in U_e$.

Now consider $f - f_{\text{even}}$, which we need to show is in U_o. Note that

$$(f - f_{\text{even}})(x) = f(x) - \frac{f(x)+f(-x)}{2} = \frac{f(x)-f(-x)}{2}$$

Then,

$$(f - f_{\text{even}})(-x) = \frac{f(-x)-f(-(-x))}{2} = -\left(\frac{f(x)-f(-x)}{2}\right) = -(f - f_{\text{even}})(x)$$

Thus, $f - f_{\text{even}} \in U_o$ as desired.

Now, $f(x) = f_{\text{even}}(x) + (f - f_{\text{even}})(x) \in U_e + U_o$ so $\mathbb{R}^{\mathbb{R}} = U_e \oplus U_o$.

2 Lecture 2

2.1 Span

Definition 2.1: Linear Combination
Let V be a vector space over a field $\mathbb{F}$ and let $v_1, \ldots, v_k \in V$. A linear combination of $v_1, \ldots, v_k$ is a vector of the form $c_1 v_1 + \cdots + c_k v_k$ where $c_1, \ldots, c_k \in \mathbb{F}$.

Definition 2.2: Span
The span of $v_1, \ldots, v_k$ is the set of all linear combinations of $v_1, \ldots, v_k$, i.e.

$$\mathrm{span}(v_1, \ldots, v_k) = \{a_1 v_1 + \cdots + a_k v_k \mid a_1, \ldots, a_k \in \mathbb{F}\}$$

Note 2.1
Very common but not universal notation: let $\emptyset$ be the empty set/list. We declare that $\mathrm{span}(\emptyset) = \{0\}$.

Example 2.1
Is the vector $(7, 8, 9)$ a linear combination of the vectors $(1, 2, 3)$ and $(4, 5, 6)$, i.e., is $(7, 8, 9) \in \mathrm{span}((1, 2, 3), (4, 5, 6))$? Note that $\mathrm{span}((1, 2, 3), (4, 5, 6))$ is a plane in $\mathbb{R}^2$ so we are essentially asking if $(7, 8, 9)$ lies on that plane. If it does, it will satisfy

$$c_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + c_2 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$$

for some choice of c_1 and c_2. We will attempt to solve for those values in the following system

$$\begin{aligned} c_1 + 4c_2 &= 7 \\ 2c_1 + 5c_2 &= 8 \\ 3c_1 + 6c_2 &= 9 \end{aligned}$$

We can row reduce the following matrix

$$\left[\begin{array}{cc|c} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{array}\right]$$

to get $c_1 = -1$ and $c_2 = 2$. Since $\begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 2 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, it must be a linear combination of the two given vectors.

Theorem 2.1
$v_1, \ldots, v_k \in V \implies \mathrm{span}(v_1, \ldots, v_k)$ is a subspace of V.

Proof: We will check all three conditions for subspaces:

- Identity: $0 = 0v_1 + \cdots + 0v_k$ so $0 \in \mathrm{span}(v_1, \ldots, v_k)$
- Addition: suppose $u, w \in \mathrm{span}(v_1, \ldots, v_k)$. Then, there are $a_1, \ldots, a_k, b_1, \ldots, b_k \in \mathbb{F}$ such that

$$\begin{aligned} u &= a_1 v_1 + \cdots + a_k v_k \\ w &= b_1 v_2 + \cdots + b_k v_k \end{aligned}$$

so,

$$\begin{aligned} u + w &= a_1v_1 + \cdots + a_kv_k + b_1v_1 + \cdots + b_kv_k \\ &= (a_1 + b_1)v_1 + \cdots + (a_k + b_k)v_k \\ &\in \operatorname{span}(v_1, \ldots, v_k) \end{aligned}$$

- Scalar Multiplication: if $u = a_1v_1 + \cdots + a_kv_k \in \operatorname{span}(v_1, \ldots, v_k)$ and $\lambda \in \mathbb{F}$, then

$$\begin{aligned} \lambda u &= \lambda(a_1v_1 + \cdots + a_kv_k) \\ &= (\lambda a_1)v_1 + \cdots + (\lambda a_k)v_k \\ &\in \operatorname{span}(v_1, \ldots, v_k) \end{aligned}$$

Theorem 2.2
If $v_1, \ldots, v_k \in V$, then $\operatorname{span}(v_1, \ldots, v_k)$ is the smallest subspace of V containing $v_1, \ldots, v_k$, i.e., $\operatorname{span}(v_1, \ldots, v_k)$ is the subspace of V generated by $v_1, \ldots, v_k$.

Proof: Suppose $W \subseteq V$ is a subspace of V and that $v_1, \ldots, v_k \in W$. Then any linear combination of $v_1, \ldots, v_k$ must be in W since W is a subspace of V. So, $\operatorname{span}(v_1, \ldots, v_k) \subseteq W$. In other words, every subspace W of V that contains a list of vectors will also contain their span.

Definition 2.3: Finite-dimensional
A vector space V is finite-dimensional if V is spanned by some finite list of vectors, i.e., $V = \operatorname{span}(v_1, \ldots, v_m)$ for some list $v_1, \ldots, v_m$.

Example 2.2
We know that $\mathbb{R}^3$ is finite-dimensional as it is spanned by

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Similarly, $\mathbb{R}^2$ is also finite-dimensional since it is spanned by

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

or some other set of vectors like

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, v_3 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

Theorem 2.3
Let p be a polynomial with real coefficients. If

$$\begin{aligned} p(x) &= a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \\ &= b_mx^m + b_{m-1}x^{m-1} + \cdots + b_1x + b_0 \end{aligned}$$

where $a_n \neq 0$ and $b_m \neq 0$, then $n = m$. Therefore, p has a well-defined degree, namely $\deg(p(x)) = n = m$.

Proof: WLOG, suppose $n > m$. Then,

$$\frac{\mathrm{d}^{n+1}}{\mathrm{d}x^{n+1}} p(x) = \frac{\mathrm{d}^{n+1}}{\mathrm{d}x^{n+1}} (a_nx^n + \cdots + a_1x + a_0)$$

$$
\begin{aligned}
&= n!a_n \\
&\neq 0
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\frac{\mathrm{d}^{n+1}}{\mathrm{d}x^{n+1}} p(x) &= \frac{\mathrm{d}^{n+1}}{\mathrm{d}x^{n+1}} (b_m x^m + \cdots + b_1 x + b_0) \\
&= 0
\end{aligned}
$$

This is contradiction, which also holds for $m > n$, so $n = m$.

Example 2.3
Let $\mathcal{P}$ be the set of *all* polynomials. Then $\mathcal{P}$ is infinite-dimensional, i.e., not finite-dimensional.

Proof: Suppose $\mathcal{P} = \operatorname{span}(p_1, \ldots, p_k)$. Let j be the max degree of $p_1, \ldots, p_k$. Consider x^{j+1}. Note that

$$x^{j+1} \notin \operatorname{span}(v_1, \ldots, v_k)$$

since $\deg(x^{j+1}) = j + 1$ where any linear combination of $v_1, \ldots, v_k$ has degree at most j. So, we can always keep appending terms like x^{j+1} to the list of vectors that span $\mathcal{P}$, which implies that $\mathcal{P}$ is not finite-dimensional.

2.2 Linear Independence

Definition 2.4: Linearly Independent
The vectors $v_1, \ldots, v_k$ are linearly independent iff they satisfy the trivial relation, i.e., $c_1 v_1 + \cdots + c_k v_k = 0 \implies c_1 = c_2 = \cdots = c_k = 0$.

Definition 2.5: Linearly Dependent
If $v_1, \ldots, v_k$ are not linearly independent, we say that they are linearly dependent. In other words, $v_1, \ldots, v_k$ are linearly dependent if there are $c_1, \ldots, c_k \in \mathbb{F}$ such that there is at least one $c_j \neq 0$ and $c_1 v_1 + \cdots + c_k v_k = 0$.

Note 2.2
We declare that the empty list is linearly independent.

Note 2.3
Any list with the 0 vector is linearly dependent.

Example 2.4
Consider $w_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $w_2 = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$ and $w_3 = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$ in $\mathbb{R}^3$. Then,

$$
\begin{aligned}
2w_1 - 1w_2 + 0w_3 = 0 \\
0w_1 + 3w_2 - 2w_3 = 0
\end{aligned}
$$

so w_1, w_2, w_3 are clearly linearly dependent. Moreover, $\operatorname{span}(w_1, w_2, w_3) = \operatorname{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)$ is a line in $\mathbb{R}^3$.

Theorem 2.4: Linear Dependence Lemma

If $v_1, \ldots, v_m$ are linearly dependent, then the Linear Dependence Lemma (LDL) states that there is some $j \leq m$ such that

- $v_j \in \text{span}(v_1, \ldots, v_{j-1})$
- $\text{span}(v_1, \ldots, v_m) = \text{span}(v_1, \ldots, \hat{v}_j, \ldots, v_m)$ where the RHS denotes the span with v_j removed from the list

Proof: We will prove both parts:

- If $v_1, \ldots, v_m$ are linearly dependent, then we can find coefficients $a_1, \ldots, a_m$, not all 0, so that $a_1v_1 + \cdots + a_mv_m = 0$. Let j be the largest $k \leq m$ such that $a_k \neq 0$. Then,

$$\begin{aligned} a_1v_1 + \cdots + a_jv_j &= 0 \\ a_jv_j &= -a_1v_1 - \cdots - a_{j-1}v_{j-1} \\ v_j &= -\frac{a_1}{a_j}v_1 - \cdots - \frac{a_{j-1}}{a_j}v_{j-1} \\ &\in \text{span}(v_1, \ldots, v_{j-1}) \end{aligned}$$

- If $u = c_1v_1 + \cdots + c_mv_m$, substitute the expression for v_j above to get

$$\begin{aligned} u &= c_1v_1 + \cdots + c_{j-1}v_{j-1} + c_jv_j + c_{j+1}v_{j+1} + \cdots + c_mv_m \\ &= c_1v_1 + \cdots + c_{j-1}v_{j-1} + c_j\left(-\frac{a_1}{a_j}v_1 - \cdots - \frac{a_{j-1}}{a_j}v_{j-1}\right) + c_{j+1}v_{j+1} + \cdots + c_mv_m \\ &= \left(c_1 - \frac{a_1}{a_j}c_j\right)v_1 + \cdots + \left(c_{j-1} + \frac{a_{j-1}}{a_j}c_j\right)v_{j-1} + c_{j+1}v_{j+1} + \cdots + c_mv_m \end{aligned}$$

  Thus, $\text{span}(v_1, \ldots, v_m) = \text{span}(v_1, \ldots, \hat{v}_j, \ldots, v_m)$ since any vector u in the first set can be rewritten as a vector in the second set.

Note 2.4

If $j = 1$, then $v_1 = 0$ and $u = c_2v_2 + \cdots + c_mv_m$. This is precisely because $\text{span}(\emptyset)$ can be defined as $\{0\}$.

Theorem 2.5

Suppose $u_1, \ldots, u_m$ are linearly independent in V and $w_1, \ldots, w_n$ span V. Then, $m \leq n$.

Proof: Look at $u_1, w_1, \ldots, w_n$. Since $u_1 \in \text{span}(w_1, \ldots, w_n)$, the list $u_1, w_1, \ldots, w_n$ is linearly dependent. Then, the Linear Dependence Lemma applies here and some w_j can be removed from $u_1, w_1, \ldots, w_n$ such that $\text{span}(u_1, w_1, \ldots, \hat{w}_j, \ldots w_m) = \text{span}(u_1, w_1, \ldots, w_n) = V$.

How do we know that u_1 is not the vector that gets removed? If it was, it would imply that $u_1 = 0$ (see note 1.4). However, this would be a contradiction to the linear independence of $u_1, \ldots, u_m$ (see note 1.3). Thus, it really had to have been a w_j vector that was removed.

Next, insert u_2 into the second slot of this list to get $u_1, u_2, w_1, \ldots, \hat{w}_j, \ldots, w_m$. This list is once again linearly dependent since $u_2 \in V \implies u_2 \in \text{span}(w_1, \ldots, w_n) = \text{span}(u_1, w_1, \ldots, \hat{w}_j, \ldots, w_m)$, and the LDL can be applied here. Due to the same reasoning as above, u_1 is not removed from the spanning list. However, neither is u_2 since $u_2 \notin \text{span}(u_1)$ (recall the first part of LDL) and u_1, u_2 are linearly independent. Therefore, one of the w_i vectors must be removed again, which will yield a new list $u_1, u_2, w_1, \ldots, \hat{w}_i, \ldots, \hat{w}_j, \ldots, w_n$ (it is entirely possible that w_1 was removed instead — WLOG, we just assumed that $1 < i < j < n$ here) for $i \neq j$ that spans V.

$\vdots$

Next, insert the vector u_j after u_{j-1} into the list above, which will make it linearly dependent once again. However, since $u_1, \ldots, u_j$ themselves are linearly independent, by the LDL, some vector to the right of u_j, i.e. a w vector, can be removed from the spanning list, which will continue to span V. Notice that the exact same logic that was applied to the $j = 2$ case above generalizes here as well.

$$\vdots$$

We can repeat this process until all of $u_1, \ldots, u_m$ has been inserted into the spanning list. The list still contains n vectors since every vector appended is also followed by a single removal. The first m vectors are $u_1, \ldots, u_m$ so we are left with $n - m$ total w vectors in the end. Since this new list also spans V and the number of vectors can never physically be negative, it would follow that $n - m \geq 0 \implies n \geq m$.

Theorem 2.6
If V is a finite-dimensional vector space and U is a subspace of V, then U is also finite-dimensional.

Proof: If $U = \{0\}$, then it is finite-dimensional. Otherwise, choose some $u_1 \neq 0$ in U. Then, either $U = \operatorname{span}(u_1)$ and we are done, or there is some $u_2 \in U$ not in $\operatorname{span}(u_1)$, i.e., u_1, u_2 are linearly independent. Similarly, either $\operatorname{span}(u_1, u_2) = U$ or there is a third vector $u_3 \in U$ such that $u_3 \notin \operatorname{span}(u_1, u_2)$, i.e., u_1, u_2, u_3 are linearly independent.

We can continue this process, which must eventually terminate. Why? Suppose $w_1, \ldots, w_m$ spans V. Then, by the theorem above, no list in V and, therefore, no spanning list in U can be of length $m + 1$ or more.

2.3 Discussion Problems

Problem 2.1
Prove that if $v_n \notin \operatorname{span}(v_1, \ldots, v_{n-1})$ and $v_1, \ldots, v_{n-1}$ are linearly independent, then $v_1, \ldots, v_n$ are also linearly independent.

Answer: Suppose $c_1v_1 + \ldots c_nv_n = 0$. Thus, $c_nv_n = -c_1v_1 - \cdots - c_{n-1}v_{n-1}$. However, $v_n \notin \operatorname{span}(v_1, \ldots, v_{n-1})$ which implies that $c_n = 0$. So, $c_1v_1 + \cdots + c_{n-1}v_{n-1} + 0v_n = 0$. However, $v_1, \ldots, v_{n-1}$ are already linearly independent as given, so $c_1 = \cdots = c_{n-1} = 0$. Since all c_i are 0, the list $v_1, \ldots, v_n$ is indeed linearly independent.

Problem 2.2
Let $w_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $w_2 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ and $w_3 = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$. Are $\operatorname{span}(w_1, w_2)$ and $\operatorname{span}(w_1, w_2)$ the same subspace in $\mathbb{R}^3$?

Answer: Note that w_1 is in both sets. Also observe that $w_3 = w_2 - w_1$. Thus, $\operatorname{span}(w_1, w_3) \subseteq \operatorname{span}(w_1, w_2)$. Similarly, $w_2 = w_3 + w_1$. Thus, $\operatorname{span}(w_1, w_2) \subseteq \operatorname{span}(w_1, w_3)$. Therefore, they are the same sets.

Problem 2.3
If $v_1, \ldots, v_k \in V$ are linearly independent, is it possible that $u_1, \ldots, u_{k-1}$ spans V?

Answer: No! The length of any spanning list is less than or equal to the length of any list of linearly independent vectors.

Problem 2.4
Prove true or give a counterexample: If v_1, v_2, v_3 are linearly dependent, then any v_i for $i = 1, 2, 3$ can be written as a linear combination of the other two.

Answer: Consider $v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $v_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$. Note that v_2 is not a linear combination of v_1 and v_3.

Problem 2.5

Express $V = \mathbb{R}^2 = U \oplus W = U \oplus W'$ where $U = \{(x, 0) \mid X \in \mathbb{R}\}$ and $W \neq W'$.

Answer: Consider $W = \{(0, y) \mid y \in \mathbb{R}\}$ and $W' = \{(y, y) \mid y \in \mathbb{R}\}$.

Problem 2.6

Let $\mathcal{P}(\mathbb{F})$ be the set of polynomials with coefficients in $\mathbb{F}$. Show that x^n and x^m are linearly independent if $n \neq m$.

Answer: Let $a, b \in \mathbb{F}$ such that $ax^n + bx^n = 0$. For $x = 1$, we get $a + b = 0 \implies b = -a$. Then, $ax^n - ax^m = 0 \implies a(x^n - x^m) = 0$. Assume that $a \neq 0$. Then, $a(x^n - x^m) = 0 \implies x^n - x^m = 0 \implies x^n = x^m$. This is a contradiction so $a = 0$ and consequently, $b = 0$. Thus, x^n and x^m are linearly independent.

Note that this statement only holds for fields like $\mathbb{R}$ and $\mathbb{C}$ but does not necessarily work in finite fields like $\mathbb{Z}_3$.

3 Lecture 3

3.1 Basis and Dimension

Definition 3.1: Basis
A list of vectors in V is a basis for V if the list is linearly independent *and* it spans V.

Note 3.1
The list $v_1, \ldots, v_n$ is a basis for V iff every $v \in V$ can be written uniquely in the form $v = a_1 v_1 + \cdots + a_n v_n$ for some $a_1, \ldots, a_n \in \mathbb{F}$.

Theorem 3.1
If $\operatorname{span}(v_1, \ldots, v_k) = V$, then some sub-list (possibly the entire list) of $v_1, \ldots, v_k$ is a basis for V.

Proof: If $v_1, \ldots, v_k$ are linearly independent, then it is a basis of V and we are done. Otherwise, by the LDL, we can throw away some v_j such that $\operatorname{span}(v_1, \ldots, \hat{v}_j, \ldots, v_k) = V$. If $v_1, \ldots, \hat{v}_j, \ldots, v_k$ is linearly independent, it is a basis for V and we are done. Otherwise, apply the LDL once again and remove some v_l to obtain the list $v_1, \ldots, \hat{v}_j, \ldots, \hat{v}_l, \ldots, v_k$ for $j \neq l$ (WLOG, let $j < l$) that still spans V. If it is linearly independent, then we are done or we can continue repeating this same process.

The process must terminate within $k < \infty$ steps, or we will be left with $V = \operatorname{span}(\emptyset) = \{0\}$ after everything in $v_1, \ldots, v_k$ is discarded. Moreover, one this process terminates, we can only be left with a linearly independent list of vectors, which will form the basis of V.

Note 3.2
An immediate conclusion of the theorem above is that every finite-dimensional vector space has a basis.

Theorem 3.2
If $v_1, \ldots, v_n$ and $u_1, \ldots, u_m$ are two bases of V, then $n = m$.

Proof: Since $u_1, \ldots, u_m$ is linearly independent and $v_1, \ldots, v_n$ spans V, we know that $m \leq n$. However, since $v_1, \ldots, v_n$ is linearly independent and $u_1, \ldots, u_m$ spans V, we also know that $n \leq m$. These two inequalities imply that $m = n$.

Definition 3.2: Dimension
If V is finite-dimensional, then $\dim V$ is the length of *any* basis of V.

Example 3.1
Let $V = \mathbb{R}^n$ and $e_1, \ldots, e_n$ be the standard basis for $\mathbb{R}^n$. Any other basis of $\mathbb{R}^n$ can be related to $e_1, \ldots, e_n$ by a change of basis matrix: applying an $n \times n$ invertible matrix A to $e_1, \ldots, e_n$ will yield another valid basis $Ae_1, \ldots, Ae_n$. In fact, one can rewrite A as

$$A = \begin{bmatrix} Ae_1 & \ldots & Ae_n \end{bmatrix}$$

where each Ae_i forms the ith column of A.

Example 3.2
Let $\mathcal{P}_m(\mathbb{F})$ be the set of all polynomials with degree at most m. Then, $\mathcal{P}_m$ is an $m+1$ dimensional vector space over $\mathbb{F}$ with basis $1, x, x^2, \ldots, x^m$.

Example 3.3
The same vector space can have different dimensions over different fields. The vector space $\mathbb{C}$ defined over the field $\mathbb{C}$ has dimension 1, but defined over the field $\mathbb{R}$ has dimension 2.

Theorem 3.3
Any linearly independent list $v_1, \dots, v_k$ in a finite-dimensional vector space V can be extended to form a basis for V.

Proof: Does $v_1, \dots, v_k$ initially span V? If so, then it is already a basis and we are done.

If not, then there is a $v_{k+1} \in V$ such that $v_{k+1} \notin \operatorname{span}(v_1, \dots, v_k)$. Then, $v_1, \dots, v_k, v_{k+1}$ must be linearly independent (recall problem 1.1 from last discussion). Check again: if this new list spans V, then it is a basis for V and we are done.

If not, then there is a $v_{k+2} \in V$ such that $v_{k+2} \notin \operatorname{span}(v_1, \dots, v_{k+1})$. Then, again, $v_1, \dots, v_k, v_{k+1}, v_{k+2}$ must be linearly independent. If it spans V, then it is a basis, and if not, continue this whole process.

It must eventually terminate, yielding a basis for V, because the length of the list of linearly independent vectors must be not be greater than the dimension of V.

Theorem 3.4
If U is a subspace of V and $\dim V < \infty$, then $\dim U \leq \dim V$.

Proof: Since V is finite dimensional, then so is U with, say, $\dim U = k$. Thus, U has a basis $u_1, \dots, u_k$. In particular, $u_1, \dots, u_k$ is a linearly independent list in U, which following the definition of a subspace, must also be a linearly independent list in V. So, by the theorem above, the list $u_1, \dots, u_k$ can be extended to form a basis $u_1, \dots, u_k, v_1, \dots, v_{\dim V - k}$ of V. Since $\dim V - k \geq 0$, we get that $k \leq \dim V \implies \dim U \leq \dim V$.

Theorem 3.5
If V is a finite-dimensional vector space and $U \subseteq V$ a subspace, then there exists a subspace W of V such that $V = U \oplus W$.

Proof: Let $\dim U = k$ and $\dim V = n$. Choose a basis of U such as $u_1, \dots, u_k$. Extend this basis to V, by appending any additional vectors as appropriate to get $u_1, \dots, u_k, w_1, \dots, w_{n-k}$. Let $W = \operatorname{span}(w_1, \dots, w_{n-k})$. Since $\operatorname{span}(u_1, \dots, u_k, w_1, \dots, w_{n-k}) = V$, for any $v \in V$, there are scalars $c_1, \dots, c_k, d_1, \dots, d_{n-k} \in \mathbb{F}$ such that

$$v = c_1 u_1 + \cdots + c_k u_k + d_1 w_1 + \cdots + d_{n-k} w_{n-k}$$

so $V = U + W$.

To show $V = U \oplus W$, we now need to show that $U \cap W = \{0\}$. If $x \in U \cap W$, then $x \in U$ and $x \in W$. Since $x \in U$, it can be written as $x = c_1 u_1 + \cdots + c_k u_k + 0 w_1 + \cdots + 0 w_{n-k}$ for some $c_1, \dots, c_k \in \mathbb{F}$. Similarly, since $x \in W$, it can be written as $x = 0 u_1 + \cdots + 0 u_k + d_1 w_1 + \cdots + d_{n-k} w_{n-k}$ for some $d_1, \dots, d_{n-k} \in \mathbb{F}$. Writing these next to each other, notice that

$$\begin{aligned} x &= c_1 u_1 + \cdots + c_k u_k + 0 w_1 + \cdots + 0 w_{n-k} \\ &= 0 u_1 + \cdots + 0 u_k + d_1 w_1 + \cdots + d_{n-k} w_{n-k} \end{aligned}$$

Comparing term by term will give us $c_1 u_1 + \cdots + c_k u_k = 0$ and $d_1 w_1 + \cdots + d_{n-k} w_{n-k} = 0$. However, both of these lists are linearly independent, which implies that $c_1 = \cdots = c_k = 0$ and $d_1 = \cdots = d_{n-k} = 0$. Thus, $x = 0 u_1 + \cdots + 0 u_k + 0 w_1 + \cdots + 0 w_{n-k} = 0$. Since an arbitrary $x \in U \cap W$ was considered, this implies that $U \cap W = \{0\}$, as desired.

Theorem 3.6
If U_1 and U_2 are finite-dimensional subspaces of a vector space V, then

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$$

3.2 Discussion Problems

Problem 3.1
Suppose $\dim V = 5$. If α is a list of length 4 of vectors in V, which of the following are possible?

1. α is linearly independent
2. α is linearly dependent
3. $\operatorname{span}(\alpha) = V$
4. $\operatorname{span}(\alpha) \neq V$

Answer the same questions for β, a list of length 6 in V.

Answer: For α, items 1 and 2 may be true, item 3 is never true (need a list of length 5 to span V) and as such, item 4 is always true. For β, item 1 is impossible and item 2 is always true (there are 5 linearly independent basis vectors so a 6th one must always be in their span), and items 3 and 4 are possible.

Problem 3.2
Suppose $\dim V = n$ and α is a list of length n of vectors in V. Explain why, if α is linearly independent, it must be a basis for V.

Answer: We can extend α to a basis for V. However, any basis of V has length n, so any extension of α that is a basis of V must be α itself.

Problem 3.3
Suppose $\dim V = n$ and α is a list of length n of vectors in V. Explain why, if $\operatorname{span}(\alpha) = V$, it must be a basis for V.

Answer: We can reduce α to a basis for V. However, any basis of V has the same length as α, so any reduction of α must be to α itself. Thus, α is a basis for V.

Problem 3.4
Prove that if U is a subspace of V with $\dim U = \dim V = n$, then $U = V$.

Answer: Suppose $u_1, \ldots, u_n$ is a basis of U. It is therefore a linearly independent list of length n in V too, which implies it is a basis for V. Then, $U = \operatorname{span}(u_1, \ldots, u_n) = V$.

Problem 3.5
Show that the space $C(\mathbb{R}) = \{\text{continous functions } f : \mathbb{R} \mapsto \mathbb{R}\}$ is infinite-dimensional.

Answer: Note that $\mathcal{P}(\mathbb{R})$ is a subspace of $C(\mathbb{R})$, so $\dim \mathcal{P}(\mathbb{R}) \leq \dim C(\mathbb{R})$. However, $\mathcal{P}(\mathbb{R})$ is infinite-dimensional, which implies that $C(\mathbb{R})$ is infinite-dimensional as well.

This provides an example of $U \subset V$ such that $\dim U = \dim V = \infty$. Consider $\sin(x) \in C(\mathbb{R})$. Since $\sin(x)$ has infinitely many roots, it is not in $\mathcal{P}(\mathbb{R})$.

Problem 3.6
If $v_1, \ldots, v_4$ is a basis of V and U is a subspace of V such that $v_1, v_2 \in U$ but $v_3, v_4 \notin U$, must v_1, v_2 be a basis of U?

Answer: No — U could be $\operatorname{span}(v_1, v_2, v_3 + v_4)$

Problem 3.7
Let $U = \left\{p \in \mathcal{P}_4(\mathbb{R}) \mid \int_{-1}^{1} p(x)\,\mathrm{d}x = 0\right\}$. Find a basis of U.

Answer: Since $p \in \mathcal{P}_4(\mathbb{R})$, every such polynomial is of the form $p(x) = ax^4 + bx^3 + cx^2 + dx + e$. Then,

$$\begin{aligned} \int_{-1}^{1} p(x)\,\mathrm{d}x &= 0 \\ \int_{-1}^{1} (ax^4 + bx^3 + cx^2 + dx + e)\,\mathrm{d}x &= 0 \\ \frac{ax^5}{5} + \frac{bx^4}{4} + \frac{cx^3}{3} + \frac{dx^2}{2} + ex\Big|_{-1}^{1} &= 0 \\ \frac{2a}{5} + \frac{2c}{3} + 2e &= 0 \end{aligned}$$

Thus, b and d are arbitrary so you can take x^3 and x as two basis elements. Now, observe that

$$\begin{aligned} \int_{-1}^{1} \left(x^4 - \frac{1}{5}\right)\mathrm{d}x &= \frac{x^5}{5} - \frac{x}{5}\Big|_{-1}^{1} \\ &= \frac{2}{5} - \frac{2}{5} \\ &= 0 \\ \int_{-1}^{1} \left(x^2 - \frac{1}{3}\right)\mathrm{d}x &= \frac{x^3}{3} - \frac{x}{3}\Big|_{-1}^{1} \\ &= \frac{2}{3} - \frac{2}{3} \\ &= 0 \end{aligned}$$

Thus, these are possible basis elements of degree 4 and 2. Since the only constant function that belongs in U is the zero function $0(x) = 0$, there is no basis element of degree 0. Therefore, one possible basis of U is $x^4 - \frac{1}{5}, x^3, x^2 - \frac{1}{3}, x$.

4 Lecture 4

4.1 Linear Maps

Definition 4.1: Linear Map/Transformation

Let V and W be two vector spaces over the same field $\mathbb{F}$. A mapping $T : V \mapsto W$ is a linear map/linear transformation if it satisfies:

$$\begin{aligned} T(u+v) &= T(u) + T(v) \quad &\text{(additivity)} \\ T(cv) &= cT(v) \quad &\text{(homogeneity)} \end{aligned}$$

for $u, v \in V$ and $c \in \mathbb{F}$.

A linear map $T : V \mapsto W$ is a map that "respects" the vector space structures on V and W, or is "compatible" with the vector space structures on V and W.

Example 4.1

Here is a non-example of a linear map. Consider $T : \mathbb{R} \mapsto \mathbb{R}$ where $T(x) = x^2$. Lets look at the two properties of a linear transformation:

- $T(x+y) = (x+y)^2 \neq x^2 + y^2 = T(x) + T(y)$
- $T(cx) = (cx)^2 = c^2x^2 \neq cx^2 = cT(x)$

Thus, T is neither additive nor homogenous.

Example 4.2

Suppose $T : \mathbb{R} \mapsto \mathbb{R}$ is linear. Then, there exists a unique m such that $T = T_m$ where $T_m(x) = mx$ for all $x \in \mathbb{R}$.

Proof: Let $e_1 = 1 \in V = \mathbb{R}$ be a basis for V. Then, $T(x) = T(x \cdot e_1) = T(x \cdot 1) = xT(1) = T(1)x$. This holds because $\mathbb{R}$ is both V and $\mathbb{F}$ in this case. Conclusion: $T(x) = T_m(x) = mx$ where $m = T(1)$.

Example 4.3

Similar technique shows that if $T : \mathbb{R}^2 \mapsto \mathbb{R}$ is linear, then $T(x, y) = ax + by$ for all $x, y \in \mathbb{R}$ for some $a, b \in \mathbb{R}$.

Note 4.1

Suppose $v_1, \ldots, v_n$ is a basis of V and that $T : V \mapsto W$ is linear. Then, T is completely determined by the vectors $T(v_1), \ldots, T(v_n)$. Why? If $v = c_1v_1 + \cdots + c_nv_n$, then

$$T(v) = T(c_1v_1 + \cdots + c_nv_n) = c_1T(v_1) + \cdots + c_nT(v_n)$$

Theorem 4.1

Let V and W be two n-dimensional vector spaces. Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_n \in W$. There exists a unique linear map $T : V \mapsto W$ such that $T(v_i) = w_i$ for each $i \leq n$.

Proof: Breaking the proof down into multiple parts to make it more digestible:

- Define $T : V \mapsto W$ by $T(c_1v_1 + \cdots + c_nv_n) = c_1w_1 + \cdots + c_nw_n$.
- We first need to check that this is well-defined, i.e.,

$$c_1v_1 + \cdots + c_nv_n = d_1v_1 + \cdots + d_nv_n \implies T(c_1v_1 + \cdots + c_nv_n) = T(d_1v_1 + \cdots + d_nv_n)$$

Look at $T(c_1v_1 + \cdots + c_nv_n) = c_1w_1 + \cdots + c_nw_n$ and $T(d_1v_1 + \cdots + d_nv_n) = d_1w_1 + \cdots + d_nw_n$. Are they equal? Yes — $v_1, \ldots, v_n$ being linearly independent and $c_1v_1 + \cdots + c_nv_n = d_1v_1 + \cdots + d_nv_n$ implies that $c_j = d_j$ for all $j \leq n$. Thus,

$$\begin{aligned} T(c_1v_1 + \cdots + c_nv_n) &= c_1w_1 + \cdots + c_nw_n \\ &= d_1w_1 + \cdots + d_nw_n \\ &= T(d_1v_1 + \cdots + d_nv_n) \end{aligned}$$

as desired.

- Since any v can be expressed as $c_1v_1 + \cdots + c_nv_n$ for some $c_1, \ldots, c_n \in \mathbb{F}$, the map T is defined on all $v \in V$. By taking $c_j = 1$ and $c_l = 0$ for $l \neq j$, we get $T(v_j) = w_j$.
- Next, we verify that T is linear. This should be pretty straightforward. If $u = a_1v_1 + \cdots + a_nv_n$ and $v = b_1v_1 + \cdots + b_nv_n$, then $u + v = (a_1 + b_1)v_1 + \cdots + (a_n + b_b)v_n$. So,

$$\begin{aligned} T(u + w) &= T((a_1 + b_1)v_1 + \cdots + (a_n + b_n)v_n) \\ &= (a_1 + b_1)T(v_1) + \cdots + (a_n + b_n)T(v_n) \\ &= (a_1 + b_1)w_1 + \cdots + (a_n + b_n)w_n \\ &= a_1w_1 + \cdots + a_nw_n + b_1w_1 + \cdots + b_nw_n \\ &= T(u) + T(w) \end{aligned}$$

  Proof for $T(\lambda v) = \lambda T(v)$ is very similar.
- Now, we prove uniqueness. Suppose S is a linear map that takes each v_j to w_j for all $j \leq n$. Then,

$$\begin{aligned} S(c_1v_1 + \cdots + c_nv_n) &= c_1S(v_1) + \cdots + c_nS(v_n) \\ &= c_1w_1 + \cdots + c_nw_n \end{aligned}$$

  for all $c_1, \ldots, c_n \in \mathbb{F}$. Then, $S = T$ as desired.

Theorem 4.2
If $T : V \mapsto W$ is a linear map, then $T(0) = 0$.

Proof: $T(0) = T(0 + 0) = T(0) + T(0) \implies T(0) + T(0) - T(0) = T(0) - T(0) \implies T(0) = 0$

Example 4.4
In the theorem above, we used the linear independence of $v_1, \ldots, v_n$ to argue that T is well-defined. If T is linearly dependent, however, then T won't necessarily be well-defined. Consider $V = W = \mathbb{R}^2$ and

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, v_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, w_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, w_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Let

$$\begin{aligned} T(v_1) &= w_1 \\ T(v_2) &= w_2 \end{aligned}$$

Then,

$$T(2v_1) = 2T(v_1) = 2w_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 0 \end{bmatrix} = w_2 = T(v_2)$$

In other words, $2v_1 = v_2$ does not imply that $T(2v_1) = T(v_2)$. Thus, T does not represent a well-defined map.

Definition 4.2: Vector Space of Linear Maps
Define $\mathcal{L}(V, W) = \{T : V \mapsto W \mid T \text{ is linear}\}$. Then, $\mathcal{L}(V, W)$ is a vector space over $\mathbb{F}$.

The vector space $\mathcal{L}(V, W)$ is valid over the following operations:

- Define an addition on $\mathcal{L}(V, W)$ as $(S + T)(v) = S(v) + T(v)$ for all $v \in V$.
- Define a scalar multiplication on $\mathcal{L}(V, W)$ as $(\lambda T)(v) = \lambda T(v)$ for all $v \in V, \lambda \in \mathbb{F}$.
- Define the zero map $T(v) = 0$ for all $v \in V$ as the additive identity.

The proof of $\mathcal{L}(V, W)$ being a valid vector space over the operations above is left as an exercise for the reader.

Note 4.2
$\dim \mathcal{L}(V, W) = \dim V \dim W$

Let $v_1, \ldots, v_{\dim V}$ and $w_1, \ldots, w_{\dim W}$ be some basis of V and W respectively. Consider T_{ij} such that $T_{ij}(v_i) = w_j$ and $T_{ij}(v_k) = 0$ when $k \neq i$. It is left to the reader to show that T_{ij} for $1 \leq i \leq \dim V$ and $1 \leq j \leq \dim W$ is a basis for $\mathcal{L}(V, W)$.

4.2 Null Space and Range

Definition 4.3: Kernal/Null Space
If $T : V \mapsto W$ is linear, then $\ker(T) = \text{null}(T) = \{v \in V \mid T(v) = 0\}$.

Note 4.3
$\ker(T)$ is a subspace of V.

Definition 4.4: Image/Range
If $T : V \mapsto W$ is linear, then $\text{im}(T) = \text{range}(T) = \{T(v) \mid v \in V\}$.

Note 4.4
$\text{range}(T)$ is a subspace of W.

Example 4.5
Let $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be the reflection about the line $y = x$. Then, $T(e_1) = e_2$ and $T(e_2) = e_1$. In other words, $T(c_1e_1 + c_2e_2) = c_2e_1 + c_1e_2$. Define

$$[T]_e^e = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

to be the matrix, in the standard basis, that represents this transformation. Finally, $\text{range}(T) = \mathbb{R}^2$ and $\ker(T) = \{0\}$

Example 4.6
Let $\text{proj}_W : \mathbb{R}^3 \mapsto \mathbb{R}^3$ be the orthogonal projection onto a plane $W \subseteq \mathbb{R}^2$. Then, $\text{range}(\text{proj}_W) = W \subseteq \mathbb{R}^3$ and $\ker(\text{proj}_W)$ is the line in $\mathbb{R}^3$ that is normal to W.

Definition 4.5: Surjection
The map $T : V \mapsto W$ is onto or surjective if $\text{range}(T) = W$.

Definition 4.6: Injection
The map $T : V \mapsto W$ is injective if it is one-to-one, i.e., $T(v) = T(u) \implies v = u$ for all $u, v \in V$.

Theorem 4.3
$T : V \mapsto W$ is injective iff $T(v) = 0 \implies v = 0$, i.e., iff $\ker(T) = \{0\}$.

Proof: Biconditional proof:

- If T is injective, then it is one to one and $T(v) = 0 = T(0) \implies v = 0$.
- Suppose $T(v) = T(u)$. Then, $T(v) - T(u) = T(v - u) = 0$ and $\ker(T) = \{0\}$ implies $v - u = 0 \implies v = u$ as desired.

Example 4.7
Let $V = W = \mathcal{P}_5(\mathbb{R})$. Then, define $D : \mathcal{P}_5(\mathbb{R}) \mapsto \mathcal{P}_5(\mathbb{R})$ where $Df = f'$ is the derivative of f. Then, $\text{range}(D) = \mathcal{P}_4(\mathbb{R}) \subseteq \mathcal{P}_5(\mathbb{R})$ and $\ker(D)$ is the set of constant functions $(f(x) = c)$. Note that D is neither injective nor surjective. However, if D was instead defined as $D : \mathcal{P}_5(\mathbb{R}) \mapsto \mathcal{P}_4(\mathbb{R})$, then it would be surjective.

Example 4.8
Let's see what happens if we consider an infinite-dimensional vector space $V = W = C^\infty([a, b])$, i.e. the set of ∞-differentiable functions on $[a, b]$. Let $D : C^\infty([a, b]) \mapsto C^\infty([a, b])$ where D is again the differentiation map defined above. Then, $\ker(D)$ is the still the set of constant functions and $\text{range}(D) = C^\infty([a, b]) = W$. So, in this case, D *is* surjective, but it still isn't injective.

Theorem 4.4: Rank-Nullity Theorem
Suppose $T : V \mapsto W$ is a linear map, with $\dim V < \infty$. Then, $V = U \oplus Z$ where $U = \ker(T)$ and Z is a non-unique subspace of V such that $T(Z) = \{T(z) \mid z \in Z\} = \text{range}(T)$ and T is injective when restricted to Z. As an immediate consequence, $\dim V = \dim \ker(T) + \dim \text{range}(T)$.

Proof: Since $\dim V < \infty$ and $\ker(T)$ is a subspace of V, there is a basis $u_1, \ldots, u_m$ for $U = \ker(T)$ that extends to a basis $u_1, \ldots, u_m, v_1, \ldots, v_n$ of V (so $\dim V = m + n$). Let $Z = \text{span}(v_1, \ldots, v_n)$. We claim that $T(v_1), \ldots, T(v_n)$ is a basis of $\text{range}(T)$. First suppose $x = c_1u_1 + \cdots + c_mu_m + d_1v_1 + \cdots + d_nv_n$ for $c_1, \ldots, c_m, d_1, \ldots, d_n \in \mathbb{F}$. Thus, $V = U + Z$ is a direct sum since every $x \in V$ can be written like that. Then,

$$\begin{aligned} T(x) &= T(c_1u_1 + \cdots + c_mu_m + d_1v_1 + \cdots + d_nv_n) \\ &= \underbrace{T(c_1u_1 + \cdots + c_mu_m)}_{0} + T(d_1v_1 + \cdots + d_nv_n) \\ &= d_1T(v_1) + \cdots + d_nT(v_n) \end{aligned}$$

Therefore, $\text{range}(T) = \text{span}(T(v_1), \ldots, T(v_n)) = T(Z)$ as desired.

We will now show that $T(v_1), \ldots, T(v_n)$ are linearly independent. Suppose $a_1T(v_1) + \cdots + a_nT(v_n) = 0$ for some scalars $a_1, \ldots, a_n \in \mathbb{F}$. Then, $T(a_1v_1 + \cdots + a_nv_n) = 0$ and $a_1v_1 + \cdots + a_nv_n \in \ker(T)$. So, $a_1v_1 + \cdots + a_nv_n = b_1u_1 + \cdots + b_mu_m$ for some scalars $b_1, \ldots, b_m \in F$ since $u_1, \ldots, u_m$ is a basis of $U = \ker(T)$. Thus, $b_1u_1 + \cdots + b_mu_m - a_1v_1 - \cdots - a_nv_n = 0$. However, $u_1, \ldots, u_m, v_1, \ldots, v_n$ are linearly independent (since they form a basis) so $b_1 = \ldots b_m = a_1 = \cdots = a_n = 0$. This calculation proves that

- $T(v_1), \ldots, T(v_n)$ are linearly independent and so they form a basis of $\text{range}(T)$
- $\ker(T) \cap Z = \{0\}$, i.e., $V = U \oplus Z$
- $T|_Z$ (the restriction of T to domain Z) is injective since $\ker(T|_Z) = \ker(T) \cap Z = \{0\}$

Note 4.5

If $V = W$, we are *not* claiming that $V = \ker(T) \oplus \text{range}(T)$, though that can happen. For instance, when $T : V \mapsto V$ is injective and $V = \text{range}(T)$.

Example 4.9

Here is one counterexample to the note above — let $V = W = \mathbb{R}^2$ and $T\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix}$. Then,

$$\ker(T) = \left\{ \begin{bmatrix} x \\ 0 \end{bmatrix} \mid x \in \mathbb{R} \right\}$$
$$\text{range}(T) = \left\{ \begin{bmatrix} x \\ 0 \end{bmatrix} \mid x \in \mathbb{R} \right\}$$

Observe that $\ker(T) = \text{range}(T)$ and $V = \mathbb{R}^2 \neq \ker(T) \oplus \text{range}(T)$.

Example 4.10

Define the left shift transformation $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$, i.e., $T(e_1) = 0$, $T(e_2) = e_1$ and $T(e_3) = e_2$. Then,

$$T\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Note that $\ker(T) = \text{span}(e_1)$ and $\text{range}(T) = \text{span}(e_2, e_2)$. Then, $\dim V = 3 = 1 + 2 = \dim\ker(T) + \dim\text{range}(T)$ as expected.

Theorem 4.5

Let $T \in \mathcal{L}(V, W)$ and $v_1, \ldots, v_n$ be a basis of V. Then, T is injective iff $T(v_1), \ldots, T(v_n)$ are linearly independent.

Proof: We will prove both directions:

- Suppose that

$$c_1T(v_1) + \cdots + c_nT(v_n) = 0$$
$$T(c_1v_1 + \cdots + c_nv_n) = 0$$

  for some scalars $c_1, \ldots, c_n \in \mathbb{F}$. The injectivity of T implies $\ker(T) = \{0\}$, i.e., $c_1v_1 + \cdots + c_nv_n = 0$. However, $c_1 = \cdots = c_n = 0$ since $v_1, \ldots, v_n$ is a basis, which implies that $T(v_1), \ldots, T(v_n)$ are linearly independent as well.

- We will show prove the contrapositive: if T is not injective, then $T(v_1), \ldots, T(v_n)$ are linearly dependent. Since $\ker(T) \neq 0$, there is some non-zero $c_1v_1 + \cdots + c_nv_n \in V$ such that

$$T(c_1v_1 + \cdots + c_nv_n) = 0$$
$$c_1T(v_1) + \cdots + c_nT(v_n) = 0$$

  By the linear independence of $v_1, \ldots, v_n$, there has to be at least one non-zero c_i such that $c_1v_1 + \cdots + c_nv_n \neq 0$. However, this also implies that $T(v_1), \ldots, T(v_n)$ are linearly dependent.

5 Lecture 5

5.1 Matrices

Definition 5.1: Matrix

Define

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

to be an $m \times n$ matrix where a_{ij} is the element in row i and column j.

Note 5.1

The vector $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$ can be regarded as both a column vector and an $n \times 1$ matrix.

Definition 5.2: Matrix-Vector Product

The product of an $m \times n$ matrix A and a vector $x \in \mathbb{R}^n$ is defined as

$$Ax = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n \end{bmatrix} \in \mathbb{R}^m$$

The ith row of Ax can be viewed as the dot product of the ith row of A with x.

Note 5.2

We will assume that you already know how matrix addition, scalar multiplication and vector dot-products work from previous linear algebra courses.

Theorem 5.1

$Ax = x_1a_1 + \dots + x_na_n$ where $a_j = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}$ is the jth column of A.

Proof: Observe that

$$\begin{aligned} Ax &= \begin{bmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n \end{bmatrix} \\ &= \begin{bmatrix} a_{11}x_1 \\ \vdots \\ a_{m1}x_1 \end{bmatrix} + \dots + \begin{bmatrix} a_{1n}x_n \\ \vdots \\ x_{mn}x_n \end{bmatrix} \\ &= x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \dots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} \\ &= x_1a_1 + \dots + x_na_n \end{aligned}$$

Given a system of linear equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

with m equations and n unknowns, we can rewrite it as

$$\begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$
$$Ax = b$$

The augmented matrix $(A \mid b)$ for the system corresponds to the $Ax = b$ expression given above. This is among the many good reasons for defining matrix-vector product the way we did above!

Few observations about the matrix-vector product:

1. $Ae_j = a_j$ where e_j is a part of the standard basis of $\mathbb{R}^n$.

2. An easy calculation shows that Ax is linear:

$$\begin{aligned} A(x + x') &= Ax + Ax' \quad && \forall x, x' \in \mathbb{R}^n \\ A(cx) &= cAx && \forall x \in \mathbb{R}^n, c \in \mathbb{R} \end{aligned}$$

Note 5.3

$\mathbb{F}^{m\times n}$ denotes the vector space of $m \times n$ matrices whose elements are in $\mathbb{F}$. Furthermore, $\dim \mathbb{F}^{m\times n} = mn$.

Theorem 5.2: Matrix-Matrix Product

If $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, we want to extend our definition of matrix multiplication so that

1. AB is defined and $AB \in \mathbb{R}^{m\times p}$

2. If $p = 1$, this just reduces to matrix vector multiplication

3. $(AB)x = A(Bx)$ for all $x \in \mathbb{R}^p$

We claim that AB is uniquely determined by the requirements above.

Proof: According to the associativity condition above,

$$\begin{aligned} (AB)x &= A(Bx) \\ &= A(x_1b_1 + \cdots + x_pb_p) \\ &= x_1Ab_1 + \cdots + x_pAb_p \end{aligned}$$

where $b_1, \ldots, b_p$ are the columns of matrix B. Letting $x_j = e_j$ will yield $(AB)e_j = Ab_j$. In other words, the jth column of AB is the product of matrix A and the jth column of B. Thus, AB is uniquely determined (since regular matrix vector multiplication is also uniquely determined).

Example 5.1

If A is $m \times n$ and B is $n \times p$, then the i,jth entry of AB is $(AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$. Why? $(AB)_{ij}$ is the ith entry of

$Ab_j = b_{1j}a_1 + \cdots + b_{nj}a_n$, which is

$$\begin{aligned} b_{1j}a_{i1} + \cdots + b_{nj}a_{in} &= a_{i1}b_{1j} + \cdots + a_{in}b_{nj} \\ &= \sum_{k=1}^{n} a_{ik}b_{kj} \end{aligned}$$

Theorem 5.3: Associativity of Matrix Multiplication
$(AB)C = A(BC)$ where A is $m \times n$, B is $n \times p$ and C is $p \times q$.

Proof: The jth column of $(AB)C$ is $(AB)c_j = A(Bc_j)$. However, Bc_j is the jth column of BC so $A(Bc_j) = A(BC)_j$, which is the jth column of $A(BC)$. Since this is true for all columns $j \leq q$, we get that $(AB)C = A(BC)$.

Example 5.2
Suppose A is an $m \times n$ matrix. Then, $T(x) = Ax$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. This is the same statement as "A is linear" that was made earlier.

However, there is a converse to this fact.

Theorem 5.4
$T : \mathbb{R}^n \mapsto \mathbb{R}^m$ is a linear map iff there is an $m \times n$ matrix A such that $T(x) = Ax$ for all $x \in \mathbb{R}^n$.

Proof: The first part of this biconditional is trivial and already explained. We will now show the converse:

$$\begin{aligned} T(x) &= T(x_1e_1 + \cdots + x_ne_n) \\ &= x_1T(e_1) + \cdots + x_nT(e_n) \\ &= \underbrace{\begin{bmatrix} T(e_1) & \dots & T(e_n) \end{bmatrix}}_{m\times n \text{ matrix}} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \end{aligned}$$

Taking $A = \begin{bmatrix} T(e_1) & \dots & T(e_n) \end{bmatrix}$ will give us $T(x) = Ax$. This matrix A is called the matrix representation of $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ with respect to the standard basis of both $\mathbb{R}^n$ and $\mathbb{R}^m$, and is sometimes denoted by either $[T]_e$ or $[T]_e^e$.

Example 5.3
Let $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be the reflection across the line $y = -x$. Then,

$$\begin{aligned} T(e_1) &= -e_2 \\ &= \begin{bmatrix} 0 \\ -1 \end{bmatrix} \\ T(e_2) &= -e_1 \\ &= \begin{bmatrix} -1 \\ 0 \end{bmatrix} \end{aligned}$$

Thus, $[T]_e = \begin{bmatrix} T(e_1) & T(e_2) \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$.

Example 5.4

Let $T : \mathbb{R}^3 \mapsto \mathbb{R}^2$ be defined such that $T(x, y, z) = (y - x + z, 2x + 3y)$. Then,

$$\begin{aligned} T(e_1) &= T(1,0,0) = (-1,2) \\ T(e_2) &= T(0,1,0) = (1,3) \\ T(e_3) &= T(0,0,1) = (1,0) \end{aligned}$$

Thus, $[T]_e = \begin{bmatrix} T(e_1) & T(e_2) & T(e_3) \end{bmatrix} = \begin{bmatrix} -1 & 1 & 1 \\ 2 & 3 & 0 \end{bmatrix}$.

Definition 5.3: Composition
Suppose $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with $[T]_e = A$ and $S \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^n)$ with $[S]_e = B$. Then, the composition $T \circ S$ is defined as $(T \circ S)(x) = T(Sx)$ for all $x \in \mathbb{R}^p$ and is well defined.

Theorem 5.5
$[T \circ S]_e = [T]_e[S]_e = AB$

Proof: Note that $[T \circ S]_e = AB$ implies that $(T \circ S)e_j$ is the jth column of AB for all $j \leq p$. However, the jth column of AB is $Ab_j = A(Se_j) = T(Se_j) = (T \circ S)e_j$ by the definition of compositions. Thus, $[T \circ S]_e = AB$ is indeed a true statement.

Note 5.4
The proof above also shows that $(T \circ S)(x) = T(Sx)$ for all $x \in \mathbb{R}^p$ is equivalent to the requirement that $(AB)x = A(Bx)$, which we already established earlier.

5.2 Change of Basis

Definition 5.4: Coordinate Vectors
Suppose that V has a basis $\alpha = v_1, \ldots, v_n$. If $x = c_1v_1 + \cdots + c_nv_n$, then define

$$[x]_\alpha = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n$$

to be the coordinate vector of x with respect to the basis α. Denote the map $x \mapsto [x]_\alpha$ (or equivalently, $V \mapsto \mathbb{R}^n$) by $\tilde{\alpha}$, i.e., define $\tilde{\alpha}(x) = [x]_\alpha$ for all $x \in V$. In other words, $\tilde{\alpha}$ is the map that returns the α-coordinates of some $x \in V$.

Note 5.5
Easy calculations show that $\tilde{\alpha}$ is an injective and surjective map from $V \mapsto \mathbb{R}^n$.

Definition 5.5: Matrix Representation of T
Suppose $T \in \mathcal{L}(V, W)$ and $\alpha = v_1, \ldots, v_n, \beta = w_1, \ldots, w_m$ are a basis of V and W respectively. Then,

$$[T]_\alpha^\beta = \begin{bmatrix} [T(v_1)]_\beta & \ldots & [T(v_n)]_\beta \end{bmatrix}$$

is the matrix representation of T with respect to the particular bases α and β. This is also sometimes denoted by $\mathcal{M}(T, (v_1, \ldots, v_n), (w_1, \ldots, w_m))$ or just $\mathcal{M}(T)$ if the choice of bases is clear from the context.

Let $\dim V = n$, $\dim W = m$ and $T : V \mapsto W$. Then, $[T]_\alpha^\beta$ is an $m \times n$ matrix whose jth column is $[T(v_j)]_\beta$. This can also be represented using the following *commutative diagram*:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\tilde{\alpha}"'] & W \arrow[d, "\tilde{\beta}"] \\ \mathbb{R}^n \arrow[r, "{[T]_\alpha^\beta}"'] & \mathbb{R}^m \end{tikzcd}$$

In other words, $\tilde{\beta}(T(v)) = [T]_\alpha^\beta \tilde{\alpha}(v)$ for all $v \in V$. Since $\tilde{\alpha}$ is invertible, it follows that $[T]_\alpha^\beta = \tilde{\beta} \circ T \circ (\tilde{\alpha})^{-1}$ is a linear map from $\mathbb{R}^n \mapsto \mathbb{R}^m$.

Example 5.5
Let $\dim V = n$ and α, β be two different bases for V. Define $I(v) = v$ to be the identity map from V to V. Consider the following diagram:

$$\begin{tikzcd} V \arrow[r, "I"] \arrow[d, "\tilde{\alpha}"'] & V \arrow[d, "\tilde{\beta}"] \\ \mathbb{R}^n \arrow[r, "{[I]_\alpha^\beta}"'] & \mathbb{R}^n \end{tikzcd}$$

Then,

$$\begin{bmatrix} [v_1]_\beta & \dots & [v_n]_\beta \end{bmatrix} = [I]_\alpha^\beta = \tilde{\beta} \circ I \circ (\tilde{\alpha})^{-1} = \tilde{\beta} \circ (\tilde{\alpha})^{-1}$$

where $\alpha = v_1, \dots, v_n$. Note that $(\tilde{\alpha})^{-1}$ yields a vector back from its α-coordinates, after which $\tilde{\beta}$ retrieves its β-coordinates. In other words, $[I]_\alpha^\beta : \mathbb{R}^n \mapsto \mathbb{R}^n$ is the change of basis matrix from basis α to basis β.

As a consequence of the results above, we get some natural tautologies that follow directly:

- $[T]_\alpha^\beta [v]_\alpha = [T(v)]_\beta$

  Proof: Start with the expansion of $[T]_\alpha^\beta$ as

$$\begin{aligned} [T]_\alpha^\beta [v]_\alpha &= (\tilde{\beta} \circ T \circ \tilde{\alpha}^{-1})[v]_\alpha \\ &= (\tilde{\beta} \circ T)(\tilde{\alpha}^{-1}[v]_\alpha) \\ &= (\tilde{\beta} \circ T)(v) \\ &= \tilde{\beta}(T(v)) \\ &= [T(v)]_\beta \end{aligned}$$

- $[T]_\alpha^\beta [S]_\gamma^\alpha = [T \circ S]_\gamma^\beta$

  Proof: We follow the same strategy as above:

$$\begin{aligned} [T]_\alpha^\beta [S]_\gamma^\alpha [v]_\gamma &= [T]_\alpha^\beta [S(v)]_\alpha \\ &= [T(S(v))]_\beta \\ &= [T \circ S]_\gamma^\beta [v]_\gamma \end{aligned}$$

Example 5.6
Since $[T]_\alpha^\beta [v]_\alpha = [T(v)]_\beta$, the change of basis matrix $[I]_\alpha^\beta$ implies $[I]_\alpha^\beta [v]_\alpha = [v]_\beta$ as expected.

Note 5.6
The calculation $[I]_\alpha^\alpha = [I \circ I]_\alpha^\alpha = [I]_\beta^\alpha [I]_\alpha^\beta$ shows that $[I]_\beta^\alpha = ([I]_\alpha^\beta)^{-1}$.

Example 5.7

Let $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be the reflection about the line $y = 3x$. What is $[T]_e^e$? First, choose an appropriate basis to model this problem in: let

$$\alpha = v_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, v_2 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

Since $T(v_1) = T\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} = 1v_1 + 0v_2$ and $T(v_2) = T\begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \end{bmatrix} = 0v_1 - 1v_2$,

$$[T]_\alpha^\alpha = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Then,

$$\begin{aligned} [T]_e^e &= [I \circ T \circ I]_e^e \\ &= [I]_\alpha^e [T]_\alpha^\alpha [I]_e^\alpha \\ &= [I]_\alpha^e [T]_\alpha^\alpha ([I]_\alpha^e)^{-1} \end{aligned}$$

Thus,

$$\begin{aligned} [T]_e^e &= \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \left(-\frac{1}{10} \begin{bmatrix} -1 & -3 \\ -3 & 1 \end{bmatrix} \right) \\ &= \begin{bmatrix} -4/5 & 3/5 \\ 3/5 & 4/5 \end{bmatrix} \end{aligned}$$

We have been using the notion of an inverse so far that you should already be familiar with from other linear algebra courses. We will give the math 110 definition of it next lecture.

6 Lecture 6

6.1 Matrix Representation of Linear Transformations Recap

If $T : V \mapsto W$ is linear, $\alpha = v_1, \dots, v_n$ is a basis of V, $\beta = w_1, \dots, w_m$ is a basis of W and $T(v_j) = a_{1j}w_1 + \cdots + a_{mj}w_m$ for $1 \le j \le n$, then

$$\begin{aligned} A &= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \\ &= [T]_\alpha^\beta \end{aligned}$$

is the matrix representation of T with respect to both α and β.
In other words, the columns of A are $[T(v_1)]_\beta, [T(v_2)]_\beta, \dots, [T(v_n)]_\beta$, the β-coordinates of each $T(v_j)$. We also observed that the change of basis matrix from α-coordinates to β-coordinates is simply just $[I]_\alpha^\beta$, where $I : V \mapsto V$ is the identity map.

Example 6.1
Let $D \in \mathcal{L}(\mathcal{P}_3(\mathbb{R}), \mathcal{P}_2(\mathbb{R}))$ be defined by $Dp = p'$. The bases of $\mathcal{P}_3(\mathbb{R})$ and $\mathcal{P}_2(\mathbb{R})$ are the polynomial standard bases $1, x, x^2, x^3$ and $1, x, x^2$ respectively. Then,

$$[D]_e^e = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

since

$$\begin{aligned} D(1) &= 0 &&= 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ D(x) &= 1 &&= 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ D(x^2) &= 2x &&= 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 \\ D(x^3) &= 3x^2 &&= 0 \cdot 1 + 0 \cdot x + 3 \cdot x^2 \end{aligned}$$

Example 6.2
Let $D \in \mathcal{L}(\mathcal{P}_3(\mathbb{R}), \mathcal{P}_2(\mathbb{R}))$ and $Dp = p'$. Find a basis α of $\mathcal{P}_3(\mathbb{R})$ and a basis β of $\mathcal{P}_2(\mathbb{R})$ such that

$$[D]_\alpha^\beta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

We seek $\alpha = p_1, p_2, p_3, p_4$ and $\beta = q_1, q_2, q_3$ such that

$$\begin{aligned} p_1' &= q_1 \\ p_2' &= q_2 \\ p_3' &= q_3 \\ p_4' &= 0 \end{aligned}$$

The last equation implies that $p_4 \neq 0$ is the constant polynomial. Now, we can choose p_1, p_2, p_3 to be any polynomials we want in $\mathcal{P}_3(\mathbb{R})$ so long as p_1, p_2, p_3, p_4 are linearly independent. For example, say

$$\begin{aligned} p_1 = x^3 &\implies q_1 = 3x^2 \\ p_2 = x^2 &\implies q_2 = 2x \\ p_3 = x &\implies q_3 = x \end{aligned}$$

Are q_1, q_2, q_3 a basis of $\mathcal{P}(\mathbb{R})$? Lets check: $a(3x^2) + b(2x) + c(x) = 0$ for all $x \in \mathbb{R}$ implies that $3a = 2b = c = 0 \implies a = b = c = 0$. Thus, q_1, q_2, q_3 are also linearly independent and they form a basis of the 3-dimensional space $\mathcal{P}_2(\mathbb{R})$. Note that this solution is not unique.

Example 6.3: 3C Exercise 3

Let $\dim V = n$, $\dim W = m$ and $T \in \mathcal{L}(V, W)$. Prove that there is a basis α of V and β of W such that $\mathcal{M}(T) = [T]_\alpha^\beta$ has 0 for all entries other than the (j, j) entries along the diagonal, which are all 1 for $j \leq \dim \operatorname{range}(T)$. In other words, letting $r = \dim \operatorname{range}(T)$, the matrix $\mathcal{M}(T)$ should look like

$$[T]_\alpha^\beta = \left[\begin{array}{ccc|c} 1 & & & \\ & \ddots & & 0_{r\times(n-r)} \\ & & 1 & \\ \hline & 0_{(m-r)\times r} & & 0_{(m-r)\times(n-r)} \end{array}\right]$$

Proof: Recall the statement and context of the abstract version of the rank-nullity theorem. Choose a basis $u_1, \dots, u_k$ of $\ker(T)$ and extend it to a basis $u_1, \dots, u_k, v_1, \dots, v_{n-k}$ of V. Let $Z = \operatorname{span}(v_1, \dots, v_{n-k})$. Then, $V = Z \oplus \ker(T)$ and $T(v_1), \dots, T(v_{n-k})$ is a basis for $\operatorname{range}(T)$ as in the proof of the rank-nullity theorem from earlier. Extend this to a basis $\beta = T(v_1), \dots, T(v_{n-k}), w_{n-k+1}, \dots, w_m$ of W such that $W = \operatorname{range}(T) \oplus \operatorname{span}(w_{n-k+1}, \dots, w_m)$. Taking these, $\mathcal{M}(T)$ should have the desired form, with $n - k$ 1s along the diagonal starting from the top-left, and 0s everywhere else.

Example 6.4: 3C Exercise 4

Let $\alpha = v_1, \dots, v_n$ be a basis of V and $\dim W = m$. If $T \in \mathcal{L}(V, W)$, prove that there is a basis $\beta = w_1, \dots, w_m$ of W such that all entries of the first column of $\mathcal{M}(T)$ are 0 except for possibly a 1 in the $(1, 1)$ position. In other words, the matrix $\mathcal{M}(T)$ should look like

$$[T]_\alpha^\beta = \left[\begin{array}{c|c} 0 \text{ or } 1 & \\ 0 & \text{other} \\ \vdots & \text{junk} \\ 0 & \end{array}\right]$$

.

Proof: There are two cases:

- Suppose $T(v_1) = 0$. Then, choose any basis for W and $\mathcal{M}(T)$ will have the desired form (with $a_{11} = 0$).
- Suppose $T(v_1) \neq 0$. Then choose $w_1 = T(v_1)$ and extend it to a basis $w_1, \dots, w_m$ of W. Again, $\mathcal{M}(T)$ will have the desired form (with $a_{11} = 1$).

Example 6.5: 3C Exercise 5

Let $\beta = w_1, \dots, w_m$ be a basis of W, with $\dim V = n$ and $T \in \mathcal{L}(V, W)$. Prove that there is a basis $\alpha = v_1, \dots, v_n$ of V such that all entries in the first row of $\mathcal{M}(T)$ are 0 except for possibly a 1 in the $(1, 1)$ position. In other words, the matrix $\mathcal{M}(T)$ should look like

$$[T]_\alpha^\beta = \left[\begin{array}{cccc} 0 \text{ or } 1 & 0 & \dots & 0 \\ \hline c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \vdots & \vdots \end{array}\right]$$

Proof: Choose any basis $\gamma = u_1, \ldots, u_n$ of V. Suppose

$$[T]_\gamma^\beta = \begin{bmatrix} c_{11} & c_{12} & \ldots & c_{1n} \\ c_{21} & \ldots & \ldots & \ldots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

There are again two cases:

1. If $c_{11} = c_{12} = \cdots = c_{1n} = 0$, we can choose $\alpha = \gamma$ and we are done, with the matrix $\mathcal{M}(T)$ having $c_{11} = 0$.
2. Otherwise, choose some c_{1j} that is not 0 and let

$$\begin{aligned} v_1 &= \frac{u_j}{c_{1j}} \\ v_i &= u_{i-1} - c_{1,i-1}v_1 && 2 \le i \le j \\ v_i &= u_i - c_{1,i}v_1 && j+1 \le i \le n \end{aligned}$$

Taking $\alpha = v_1, \ldots, v_n$ results in $\mathcal{M}(T)$ of the desired form with $c_{11} = 1$.

6.2 Isomorphisms

Definition 6.1: Invertible
$T \in \mathcal{L}(V, W)$ is called invertible if there exists an $S \in \mathcal{L}(W, V)$ such that $ST = I_V$ and $TS = I_W$.

Note 6.1
If S exists, then it is unique and we denote it by T^{-1} (this is reminiscent of the inverse notation for scalars).

Theorem 6.1
$T \in \mathcal{L}(V, W)$ is invertible iff T is injective and surjective.

Definition 6.2: Isomorphism
An isomorphism is an invertible linear map between vector spaces. Two vector spaces are isomorphic if there is an isomorphism from one to the other.

Note 6.2
If V and W are isomorphic, we denote that by $V \cong W$.

Theorem 6.2
Two finite-dimensional vector spaces over $\mathbb{F}$ are isomorphic iff they have the same dimension.

Proof: We will prove both directions:

- Suppose $\dim V = \dim W = n$. Choose a basis $v_1, \ldots, v_n$ of V and $w_1, \ldots, w_n$ of W. By theorem 4.1 of this note (not Axler), we can let $T\left(\sum_{k=1}^n c_k v_k\right) = \sum_{k=1}^n c_k w_k$ define a unique linear map from V to W. It is surjective because $\operatorname{span}(w_1, \ldots, w_n) = W$ and injective by the linear independence of $w_1, \ldots, w_n$. Thus, T is an isomorphism so $V \cong W$.
- We will now show that if finite-dimensional V and W are isomorphic, then $\dim V = \dim W$. If $v_1, \ldots, v_n$ is a basis of V and $T : V \mapsto W$ is surjective, then $\operatorname{span}(T(v_1), \ldots, T(v_n)) = W$. Thus, $\dim W \le \dim V = n$. Since T is injective, the vectors $T(v_1), \ldots, T(v_n)$ are linearly independent and $n = \dim V \le \dim W$. Combining the two inequalities yields $\dim V = \dim W = n$.

Example 6.6

Choosing bases $\alpha = v_1, \ldots, v_n$ of V and $\beta = w_1, \ldots, w_m$ of W sets up isomorphisms between $V \cong \mathbb{R}^n$ and $W \cong \mathbb{R}^m$ via the maps $\tilde{\alpha}$ and $\tilde{\beta}$ respectively. Furthermore, these also "induce" an isomorphism from $\mathcal{L}(V, W)$ to $\mathcal{M}_{m,n}(\mathbb{R}) = \mathbb{R}^{m\times n}$ (the set of all $m \times n$ matrices). To visualize these, look at the following commutative diagrams:

$$\begin{tikzcd}
V \arrow[r, "T"] \arrow[d, "\tilde{\alpha}"', "\cong"] & W \arrow[d, "\tilde{\beta}", "\cong"'] \\
\mathbb{R}^n \arrow[r, "{[T]_\alpha^\beta}"] & \mathbb{R}^m
\end{tikzcd}$$

$$\begin{tikzcd}
 & \mathcal{L}(V, W) \arrow[dl, "T\mapsto[T]_\alpha^\beta"', "\cong"] \arrow[dr, "T\mapsto[T]_\gamma^\delta", "\cong"'] & \\
\mathcal{M}_{m,n}(\mathbb{R}) \arrow[rr, "\cong", "A\mapsto[I]_\beta^\delta A[I]_\gamma^\alpha"'] & & \mathcal{M}_{m,n}(\mathbb{R})
\end{tikzcd}$$

$$\begin{tikzcd}
 & V \arrow[dl, "\tilde{\gamma}"', "\cong"] \arrow[dr, "\tilde{\alpha}", "\cong"'] & \\
\mathbb{R}^n \arrow[rr, "\cong", "{[I]_\gamma^\alpha=\tilde{\alpha}\circ(\tilde{\gamma})^{-1}}"'] & & \mathbb{R}^n
\end{tikzcd}
\qquad
\begin{tikzcd}
 & W \arrow[dl, "\tilde{\beta}"', "\cong"] \arrow[dr, "\tilde{\delta}", "\cong"'] & \\
\mathbb{R}^m \arrow[rr, "\cong", "{[I]_\beta^\delta=\tilde{\delta}\circ(\tilde{\beta})^{-1}}"'] & & \mathbb{R}^m
\end{tikzcd}$$

The first commutative diagram shows that $\mathcal{L}(V, W) \cong \mathcal{M}_{m,n}(\mathbb{R})$ via the linear map $T \mapsto [T]_\alpha^\beta$ given by $\tilde{\beta} \circ T \circ (\tilde{\alpha})^{-1}$. The second diagram condenses this and shows the same isomorphism with a different bases γ of V and δ of W. However, $[T]_\gamma^\delta \cong [T]_\alpha^\beta$ (the matrix representations with respect to both sets of bases) are themselves isomorphic, and this isomorphism is induced by the change of basis maps that are presented in the third and fourth diagrams.

Note 6.3

In the case that $V = W$ and $\alpha = \beta$, we get that $\mathcal{L}(V, W) \mapsto \mathcal{M}_{n,n}(\mathbb{F})$ (or equivalently $T \mapsto [T]_\alpha^\alpha$) is actually a ring (called an $\mathbb{F}$-algebra) isomorphism!

6.3 Discussion Problems

Problem 6.1

Show that if $T : V \mapsto W$ is linear, then $\ker(T)$ and $\operatorname{range}(T)$ are subspaces of W.

Answer: Should be pretty straightforward using the conditions of a subspace.

Problem 6.2

Suppose that $\dim V = \dim W = n < \infty$.

1. Show that if $T \in \mathcal{L}(V, W)$ is injective, then it is also surjective.
2. Show that if $T \in \mathcal{L}(V, W)$ is surjective, then it is also injective.

Answer: Use the rank-nullity theorem!

Problem 6.3

Construct explicit counterexamples of V, W and T that will disprove both of the statements above if $\dim V \neq \dim W$.

Problem 6.4
Suppose $T \in \mathcal{L}(\mathbb{R}^5, \mathbb{R}^3)$ and $S \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^5)$. What are the possible dimensions of

1. $\ker(T)$
2. $\operatorname{range}(T)$
3. $\ker(S)$
4. $\operatorname{range}(S)$
5. $\ker(T \circ S)$
6. $\operatorname{range}(T \circ S)$
7. $\ker(S \circ T)$
8. $\operatorname{range}(S \circ T)$

Answer: By the rank-nullity theorem,

$\dim\ker(T)$	$\dim\operatorname{range}(T)$	$\dim\ker(S)$	$\dim\operatorname{range}(S)$
5	0	0	3
4	1	1	2
3	2	2	1
2	3	3	0

$\dim\ker(T \circ S)$	$\dim\operatorname{range}(T \circ S)$	$\dim\ker(S \circ T)$	$\dim\operatorname{range}(S \circ T)$
0	3	5	0
1	2	4	1
2	1	3	2
3	0	2	3

Note that $\dim\ker(T)$ has to be lower bounded by 2 since $\dim\operatorname{range}(T) \leq 3$. Similarly, $\dim\operatorname{range}(S)$ has to be upper bounded by 3 since $\dim\ker(S) \geq 0$.
Since $\ker(S \circ T) \subseteq \ker(T)$, we get that $\dim\ker(S \circ T) \leq \dim\ker(T)$. Again, following the same logic, we also get that $\dim\ker(T \circ S) \leq \dim\ker(S)$.

Problem 6.5
Suppose that $T : V \mapsto W$ is an injective and surjective linear map. Show that the inverse map $T^{-1} : W \mapsto V$ is linear.

Problem 6.6
Consider the map $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ that orthogonally projects $\mathbb{R}^3$ onto the plane $x - y + 2z = 0$. Compute $[T]_e^e$, the matrix representation of T with respect to the standard basis of $\mathbb{R}^3$.

Answer: First, start off by finding a basis for the plane. Observe that we have 2 free variables (say x and y) and one dependent variable (say z). Then,

$$x = 1, y = 1 \implies z = 0$$
$$x = -2, y = 0 \implies z = 1$$

Thus, $\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ will form a basis for the plane. To find a basis for the vectors perpendicular to the plane, take the cross product of the two vectors above to get $\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$. Then, $\alpha = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ will form a basis for $\mathbb{R}^3$ and

$$[T]_\alpha^\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Then,

$$\begin{aligned} [T]_e^e &= [I]_\alpha^e [T]_\alpha^\alpha [I]_e^\alpha \\ &= \begin{bmatrix} -2 & 1 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -2 & 1 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & 2 \end{bmatrix}^{-1} \end{aligned}$$

Problem 6.7

Let $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be the reflection about the line $y = x$ and $S : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be the reflection about the line $y = 2x$. Give $[T]_e^e$, $[S]_e^e$, $[T \circ S]_e^e$ and $[S \circ T]_e^e$.

Answer: This is the same problem as example 5.7, but with the numbers changed. To retrieve $[T \circ S]_e^e$ and $[S \circ T]_e^e$, just multiply $[T]_e^e [S]_e^e$ and $[S]_e^e [T]_e^e$ respectively.

Problem 6.8

Let $\mathcal{P}$ be the set of all polynomials on $\mathbb{R}$ and W be the set of all infinite sequences of real numbers. Define $T \in \mathcal{L}(\mathcal{P}, W)$ by $Tf = (f(0), f'(0), f''(0), f'''(0), \dots)$. Is T injective and/or surjective? Define $\operatorname{range}(T)$ and $\ker(T)$.

Answer: T is not surjective since $(1, 1, 1, \dots, 1, \dots) \in W$ (this is a sequence of all 1s). This sequence cannot be in $\operatorname{range}(T)$ since a polynomial of degree n has $f^{(n+1)}(x) = f^{(n+2)}(x) = \cdots = 0$ and each polynomial must have a finite degree. However, T is injective. If $f = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n$ and $Tf = 0$, then

$$\begin{aligned} f(0) = c_0 &\implies c_0 = 0 \\ f'(0) = c_1 &\implies c_1 = 0 \\ f''(0) = 2c_2 &\implies c_2 = 0 \\ f'''(0) = 3c_2 &\implies c_3 = 0 \end{aligned}$$

and so on. Thus, $\ker(T) = \{0(x)\}$

7 Lecture 7

7.1 Review Problems

We just went over review problems for the upcoming midterm.

8 Lecture 8

8.1 Eigenvectors and Eigenvalues

Definition 8.1: Operator
If $T \in \mathcal{L}(V, V) = \mathcal{L}(V)$, then it is a linear operator.

Theorem 8.1
If V is finite-dimensional and $T \in \mathcal{L}(V)$, then the following are equivalent:

1. T is invertible
2. T is injective
3. T is surjective

Proof: The proof is left as an exercise for the reader. Though, it should follow pretty easily from the rank-nullity theorem!

Definition 8.2: Invariant
A subspace U of V is invariant under T or T-invariant if $T(U) = \mathrm{range}(T) \subseteq U$, i.e., $Tu \in U$ for all $u \in U$.

Example 8.1
Let $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ be the counterclockwise rotation about the z-axis by $\frac{\pi}{4}$ radians, i.e.,

$$[T]_e^e = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then, $\mathbb{R}^3$ has

- one 3-dimensional T-invariant subspace, namely $\mathbb{R}^3$
- one 2-dimensional T-invariant subspace, namely the xy-plane
- one 1-dimensional T-invariant subspace, namely the z-axis
- one 0-dimensional T-invariant subspace, namely $\{0\}$

Example 8.2
Let $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ be the orthogonal projection onto the plane spanned by v_1 and v_2 and let v_3 be their cross-product. This plane, i.e., $\mathrm{span}(v_1, v_2)$, is T-invariant, as is the normal line to the plane, given by $\mathrm{span}(v_3)$

The previous example is a special case of the following simple proposition:

Theorem 8.2
If $T : V \mapsto V$ is linear, then $\mathrm{range}(T)$ and $\ker(T)$ are both T-invariant.

Example 8.3
Let $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be the rotation counterclockwise by θ radians, where θ is a non-integer multiple of π (i.e., $\theta \neq \ldots, -\pi, 0, \pi, 2\pi, 3\pi, \ldots$). Then, T has no 1-dimensional T-invariant subspaces.

Proof: Let $U = \mathrm{span}(v)$ for some $v \neq 0$. Then, U being T-invariant would imply that $T(U) = \{0\}$ or $T(U) = U$. However, T is invertible and T does not rotate any line to itself. So T has no 1-dimensional T-invariant subspaces.

Definition 8.3: Eigenvector and Eigenvalue

Suppose $v \neq 0$ and $Tv = \lambda v$ for some $\lambda \in \mathbb{F}$. We then say that V is an eigenvector of T (with eigenvalue λ) or that v is a λ-eigenvector of T.

Note 8.1

Observe that $\operatorname{span}(v)$ is a 1-dimensional T-invariant subspace of V iff v is an eigenvector of T. Thus, the 1-dimensional T-invariant subspaces of V are precisely the spans of the eigenvectors of T.

Example 8.4

Let $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be a rotation counterclockwise by θ such that $\theta \neq \dots, -\pi, 0, \pi, 2\pi, 3\pi, \dots$. Then, T has no real-valued eigenvalues. More on this later in the course.

Example 8.5

Let $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ be defined by $T(a, b, c) = (b, c, 0)$. What are its eigenvalues and eigenvectors?

$$Tv = \lambda v \implies T\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \lambda \begin{bmatrix} a \\ b \\ c \end{bmatrix} \implies \begin{bmatrix} b \\ c \\ 0 \end{bmatrix} = \begin{bmatrix} \lambda a \\ \lambda b \\ \lambda c \end{bmatrix}$$

If $\lambda \neq 0$, then $c = 0 \implies b = 0 \implies a = 0$. However, $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0$ cannot be an eigenvector. So, $\lambda = 0$ and $\begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}$ is a 0-eigenvector of T.

Note 8.2

In the example above, T is the left or backwards shift in $\mathbb{R}^3$. Note that $\operatorname{span}(e_1, e_2)$ is T-invariant but $\operatorname{span}(e_1, e_3)$ and $\operatorname{span}(e_2, e_3)$ are not.

Example 8.6

Let $T : \mathbb{C}^2 \mapsto \mathbb{C}^2$ the linear map given by $T(w, z) = (-z, w)$. What are its eigenvectors and eigenvalues?

$$\begin{aligned} T(w, z) &= \lambda(w, z) \\ (-z, w) &= (\lambda w, \lambda z) \end{aligned}$$

Note that $z = 0 \implies w = 0$ and vice versa. However, $(0, 0)$ cannot be an eigenvector. Then, combining the two equations,

$$\begin{aligned} -z &= \lambda w \\ &= \lambda(\lambda z) \\ &= \lambda^2 z \end{aligned}$$

Since $z \neq 0$, we get that $\lambda^2 = -1 \implies \lambda = \pm i$. Then, the i-eigenvectors are given by $\{(w, -iw) \mid w \in \mathbb{C}, w \neq 0\}$ and the $-i$-eigenvectors are given by $\{(w, iw) \mid w \in \mathbb{C}, w \neq 0\}$.

Another observation: let V be finite-dimensional, $T \in \mathcal{L}(V)$. Then,

$$\begin{aligned} \lambda \text{ is an eigenvalue of } T &\iff \exists v \neq 0, Tv = \lambda v \\ &\iff \exists v \neq 0, Tv - \lambda v = 0 \\ &\iff \exists v \neq 0, (T - \lambda I)v = 0 \\ &\iff \ker(T - \lambda I) \neq \{0\} \end{aligned}$$

$$\begin{aligned}&\iff T-\lambda I \text{ is not invertible}\\&\iff T-\lambda I \text{ is not injective}\\&\iff T-\lambda I \text{ is not surjective}\end{aligned}$$

In particular, T being invertible is equivalent to saying that 0 is not an eigenvalue of T. If T is not invertible, then $\ker(T) = \operatorname{span}(0-\text{eigenvectors}) \cup \{0\}$.

Example 8.7: 5A Exercise 23

Let V be finite dimensional and $S, T \in \mathcal{L}(V)$. Prove that ST and TS have the same eigenvalues.

Proof: Suppose that $STv = \lambda v$ for $\lambda \neq 0$ and $v \neq 0$. Then, $T(STv) = T(\lambda v) = \lambda Tv$. Since $\lambda v \neq 0$, we know that $STv \neq 0 \implies Tv \neq 0$ (otherwise, $Tv = 0$ would imply $S(0) = 0$, which is not the case here). Thus, Tv is a λ-eigenvector of TS. This proves that non-zero eigenvalues of ST are all eigenvalues of TS too.

Next, note that ST is invertible iff TS is invertible (ST invertible $\iff$ S invertible and T invertible $\iff$ TS invertible). So, 0 is an eigenvalue of ST iff 0 is an eigenvalue of TS.

Example 8.8: 5A Exercise 28

Let V be finite dimensional, with $\dim V \geq 3$, and $T \in \mathcal{L}(V)$ such that every 2-dimensional subspace of V is T-invariant. Prove that $T = cI$ for some $c \in \mathbb{F}$.

Proof: Choose any $v \neq 0$. Extend to a basis $v, v_2, \dots, v_n$ of V. Since $\operatorname{span}(v, v_2)$ is 2-dimensional and T-invariant, $Tv = c_v v + c_2 v_2$ for some $c_v, c_2 \in \mathbb{F}$. Again, $\operatorname{span}(v, v_3)$ is also 2-dimensional and T-invariant so $Tv = d_v v + d_3 v_3$ for some $d_v, d_3 \in \mathbb{F}$. However, since v, v_2, v_3 are linearly independent, $Tv = Tv = c_v v + c_2 v_2 = d_v v + d_3 v_3$ implies that $c_2 = 0 = d_3$ and $c_v = d_v$. Thus, $Tv = c_v v$ for this unique $c_v \in \mathbb{F}$.

Now, we will prove the uniqueness of this c_v. Choose any $w \neq 0$. We need to show that $c_w = c_v$. If $w = av$, then $Tw = c_w w \implies T(av) = aTv = ac_v v = c_v(av) = c_v w$. So, $c_v = c_w$ as required. However, if $w \notin \operatorname{span}(v)$, then v and w are linearly independent and

$$\begin{aligned}T(v+w) &= c_{v+w}(v+w)\\&= c_{v+w}v + c_{v+w}w\\Tv + Tw &= c_v v + c_w w\end{aligned}$$

This implies that $c_v = c_w = c_{v+w}$ and $T = cI$ for $c = c_v$ as desired.

Theorem 8.3

Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T with corresponding eigenvectors $v_1, \dots, v_m$. Then, $v_1, \dots, v_m$ are linearly independent.

Proof: Suppose the list of eigenvectors $v_1, \dots, v_m$ are linearly dependent. Let k be the smallest positive integer such that $v_k \in \operatorname{span}(v_1, \dots, v_{k-1})$. We know that k exists by the LDL so there are scalars $a_i \in \mathbb{F}$ such that $v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1}$. Applying T to both sides,

$$\begin{aligned}Tv_k &= \lambda_k v_k\\&= a_1\lambda_1 v_1 + \cdots + a_{k-1}\lambda_{k-1}v_{k-1}\end{aligned}$$

but

$$\begin{aligned}\lambda_k v_k &= \lambda_k(a_1 v_1 + \cdots + a_{k-1}v_{k-1})\\&= a_1\lambda_k v_1 + \cdots + ak - 1\lambda_k v_{k-1}\end{aligned}$$

However, $v_1, \dots, v_{k-1}$ are linearly independent by the choice of k defined above. Thus,

$$\lambda_k v_k - \lambda_k v_k = 0$$

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

Then, $a_1(\lambda_1 - \lambda_k) = \cdots = a_{k-1}(\lambda_{k-1} - \lambda_k) = 0$. Since all λ_i are distinct, this implies that $a_1 = \cdots = a_{k-1} = 0$ so $v_k = 0$. However, this contradicts that v_k is an eigenvector, so $v_1, \ldots, v_m$ are indeed linearly independent.

Theorem 8.4
If $T \in \mathcal{L}(V)$ and $\dim V = n$, then T has at most n distinct eigenvalues.

Proof: Eigenvectors corresponding to distinct eigenvalues are linearly independent but V can't contain a list of linearly independent vectors greater than n.

Example 8.9: 5A Exercise 13
Let V be finite dimensional, $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Prove that there exists some $\alpha \in \mathbb{F}$ such that $|\alpha - \lambda| < \frac{1}{1000}$ and $T - \alpha I$ is invertible.

Proof: Either $T - \lambda I$ (i.e., λ is not its eigenvalue) is invertible, in which case we can choose $\alpha = \lambda$. Or λ is an eigenvalue of T. Since T has at most $\dim V$ many distinct eigenvalues, there are at most $\dim V$ values of

$$\lambda - \frac{1}{1000} < \alpha < \lambda + \frac{1}{1000}$$

that will make $T - \alpha I$ non-invertible. However, $\mathbb{F}$ is assumed to be some infinite field (like $\mathbb{R}$ or $\mathbb{C}$) and only finitely many values of α are excluded from the range above. Therefore, you can always choose some other α that will not be an eigenvalue of T, making $T - \alpha I$ invertible.

Example 8.10: 5A Exercise 30
Suppose $T \in \mathcal{L}(\mathbb{R}^3)$ and $-4, 5, \sqrt{7}$ are eigenvalues of T. Prove that there exists some $x \in \mathbb{R}^3$ such that $(T - 9I)x = (-4, 5, \sqrt{7})$.

Proof: Since $\dim \mathbb{R}^3 = 3$, the scalars $-4, 5, \sqrt{7}$ are the only eigenvalues of T. Thus, 9 is not an eigenvalue and $T - 9I$ is invertible. Hence, we can always find an $x = (T - 9I)^{-1}(-4, 5, \sqrt{7})$.

Example 8.11: 5A Exercise 29
Let $T \in \mathcal{L}(V)$, $\dim \operatorname{range}(T) = k$ and $\dim V = n$. Prove that T has at most $k + 1$ distinct eigenvalues.

Proof: We will give two proofs of this statement:

1. Suppose that T has j distinct non-zero eigenvalues. Choose the corresponding eigenvectors $v_1, \ldots, v_j$. Let $u_1, \ldots, u_{n-k}$ be a basis of $\ker(T)$ and also the eigenvectors with eigenvalue 0. Then, $v_1, \ldots, v_j, u_1, \ldots, u_{n-k}$ are linearly independent.

   Suppose $c_1 v_1 + \cdots + c_j v_j + d_1 u_1 + \cdots + d_{n-k}u_{n-k} = 0$. Then,

   $$T(c_1 v_1 + \cdots + c_j v_j + d_1 u_1 + \cdots + d_{n-k}u_{n-k}) = T(0)$$
   $$c_1\lambda_1 v_1 + \cdots + c_j\lambda_j v_j = 0$$

   Since $v_1, \ldots, v_j$ are linearly independent, $c_1\lambda_1 = \cdots = c_j\lambda_j = 0$. However, since all $\lambda_i \neq 0$, this further implies that $c_1 = \cdots = c_j = 0$. Thus, $d_1 u_1 + \cdots + d_{n-k}u_{n-k} = 0$. However, as $u_1, \ldots, u_{n-k}$ are linearly independent too, we get that $d_1 = \cdots = d_{n-k} = 0$.

   Thus, the eigenvectors span a $\dim \operatorname{span}(v_1, \ldots, v_j) = j$ dimensional subspace of $V \setminus \ker(T)$. Since $j \leq k$ and there can be no more than k different eigenvectors in $\operatorname{range}(T)$, the operator T has at most $j + 1 \leq k + 1$ distinct eigenvalues (the final +1 comes from the 0 eigenvalue as $\dim \ker(T) > 0$).

2. Let $\lambda_1, \ldots, \lambda_j$ be the distinct non-zero eigenvalues of T. Let $v_1, \ldots, v_j$ be the corresponding eigenvectors. Then,

$$T(v_i) = \lambda_i v_i$$
$$T\left(\frac{v_i}{\lambda_i}\right) = v_i$$

Therefore, $v_1, \ldots, v_j \in \operatorname{range}(T)$ are linearly independent. Thus, $j \leq k$. Since it can also have the 0 eigenvalue for everything in $\ker(T)$, the operator T will have at most $j + 1 \leq k + 1$ distinct eigenvalues.

Example 8.12: 5A Exercise 19

Let $T \in \mathcal{L}(V)$ and define it as $T(x_1, \ldots, x_n) = T(x_1 + \cdots + x_n, \ldots, x_1 + \cdots + x_n)$. What are its eigenvalues and eigenvectors?

Proof: $[T]_e^e$ should look like the following:

$$[T]_e^e = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 \end{bmatrix}$$

Note that $\dim \operatorname{range}(T) = 1$ which, by the rank-nullity theorem, implies that $\dim \ker(T) = n - 1$. Thus, 0 has to be an eigenvalue of T, and there must exist $n - 1$ linearly independent 0-eigenvectors of T in $\ker(T)$. Row-reducing the matrix above will yield

$$\operatorname{RREF}([T]_e^e) = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}$$

Thus,

$$x_1 + x_2 + \cdots + x_n = 0$$
$$x_1 = -x_2 - \cdots - x_n$$
$$\begin{bmatrix} -x_2 - \cdots - x_n \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} -1 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

and the vectors in the linear combination above will form a basis for $\ker(T)$. Thus, their span will include all of the eigenvectors associated with $\lambda = 0$.

From the example above, we know that there can be at most one more distinct eigenvalue of T and since $\operatorname{range}(T) = 1$, only one distinct eigenvector. Observe that

$$\begin{bmatrix} 1 & \ldots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \ldots & 1 \end{bmatrix} \begin{bmatrix} a \\ \vdots \\ a \end{bmatrix} = \begin{bmatrix} na \\ \vdots \\ na \end{bmatrix} = n \begin{bmatrix} a \\ \vdots \\ a \end{bmatrix} \implies \left\{ \begin{bmatrix} a \\ \vdots \\ a \end{bmatrix} \mid a \in \mathbb{F} \right\} = \operatorname{span}\left(\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right)$$

works for $\lambda = n$.

9 Lecture 9

9.1 Polynomials of Linear Maps

Definition 9.1: Integral Powers of Linear Transformations
Suppose $T \in \mathcal{L}(V)$ and m is a positive integer. Define $T^m = T \circ \cdots \circ T$ (T composed with itself m times) and $T^0 = I$, the identity operator on V. If T is invertible, then let $T^{-m} = (T^{-1})^m$.

Note 9.1
One line proofs show that $T^m T^n = T^{m+n}$ and $(T^m)^n = T^{mn}$.

Definition 9.2: Polynomials of an Operator
If $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbb{F})$ is $p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$ for all $z \in \mathbb{F}$, then $p(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_m T^m$.

Note 9.2
If $T \in \mathcal{L}(V)$, then the map $\mathcal{P}(\mathbb{F}) \mapsto \mathcal{L}(V)$ given by $p \mapsto p(T)$ is linear.

Recall that if p, q are polynomials, then $(pq)(z)$ is defined to be $p(z)q(z)$. Following this, some multiplicative properties of $p, q \in \mathcal{P}(\mathbb{F})$ and $T \in \mathcal{L}(V)$ are

1. $(pq)(T) = p(T)q(T)$
2. $p(T)q(T) = q(T)p(T)$

Example 9.1: 5B Exercise 7
Suppose $T \in \mathcal{L}(V)$. Prove that 9 is an eigenvalue of T^2 iff 3 or -3 is an eigenvalue of T.

Proof: If $Tv = \lambda v$ with $v \neq 0$, then $T(Tv) = T(\lambda v) = \lambda T(v) = \lambda(\lambda v) = \lambda^2 v$. Thus, v is an λ^2-eigenvector of T^2. Letting $\lambda = \pm 3$ concludes that 9 is an eigenvalue of T^2.

On the other hand, if 9 is eigenvalue of T^2, then $T^2 - 9I = (T - 3I)(T + 3I)$ is not invertible, which implies that at least one of $T - 3I$ or $T + 3I$ is not invertible. Thus, at least one of -3 or 3 (or both) is an eigenvalue of T.

Example 9.2: 5B Exercise 4
Suppose $P \in \mathcal{L}(V)$ and $P^2 = P$. Prove that $V = \ker(P) \oplus \text{range}(P)$.

Proof: For every $v \in V$, it can be rewritten as $v = v - Pv + Pv$. However, $v - Pv \in \ker(P)$ since $P(v - Pv) = Pv - P^2 v = Pv - Pv = 0$ and $Pv \in \text{range}(T)$ by definition. Thus, $V = \ker(T) + \text{range}(T)$.

To prove that this sum is a direct sum, we also need to show $\ker(T) \cap \text{range}(T) = \{0\}$. Let $y \in \text{range}(P)$ and $y \neq 0$. Then, $y = Px$ for some $x \in V$ and $Py = P^2 x = Px = y \neq 0$. Thus, only $y = 0$ is in both $\ker(P)$ and $\text{range}(P)$. Therefore, $V = \ker(P) \oplus \text{range}(P)$.

Theorem 9.1
Every operator on a finite-dimensional, non-zero, complex vector space has at least one eigenvalue.

Proof: Suppose $\dim V = n > 0$ and $T \in \mathcal{L}(V)$ where V is a vector space over $\mathbb{C}$. Pick some $v \in V$ such that $v \neq 0$. Then, $v, Tv, T^2 v, \ldots, T^n v$ must be linearly dependent as it is a list of length $n + 1$ is an n-dimensional vector space. So, $a_0 v + a_1 Tv + \cdots + a_n T^n v = 0$ for some $a_0, \ldots, a_n \in \mathbb{F}$ such that not all $a_i = 0$. In fact, it must be some a_i for $1 \leq i \leq n$ that is nonzero because $a_1 = \cdots = a_n = 0 \implies a_0 v = 0$. Since $v \neq 0$, it follows that $a_0 = 0$.

Consider $p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$. By the fundamental theorem of algebra, this polynomial factors into a product of linear terms, i.e., $p(z) = c(z - \lambda_1) \ldots (z - \lambda_n)$ for $c, \lambda_1, \ldots, \lambda_n \in \mathbb{C}$, and $c \neq 0$. Thus,

$$a_0 v + a_1 Tv + \cdots + a_n T^n v = 0$$

$$(a_0 I + a_1 T + \cdots + a_n T^n)v = 0$$
$$c(T - \lambda_1 I)\ldots(T - \lambda_n I)v = 0$$

Thus, $T - \lambda_j I$ is non-injective (and consequently non-invertible) for at least one j — all such λ_j will be an eigenvalue of T.

Example 9.3: 5B Exercise 13

Let W be a complex vector space and $T \in \mathcal{L}(W)$ has no eigenvalues. Prove that every T-invariant subspace of W is either $\{0\}$ or infinite dimensional.

Proof: We will proceed with contradiction. Suppose $U \neq \{0\}$ and U is a finite-dimensional T-invariant subspace of W. Then, U is a finite dimensional complex vector space and $T|_U \in \mathcal{L}(U)$. Thus, $T|_U$ has some eigenvalue $\lambda \in \mathbb{C}$ and an associated eigenvector $v \in U$. However, $Tv = T|_U v = \lambda v$ so v is a λ-eigenvector of T as well, which is a contradiction.

9.2 Upper Triangular Matrices

Definition 9.3: Upper Triangular

A matrix is upper triangular if all entries below the main diagonal are equal to 0, i.e., if $i > j$, then $a_{ij} = 0$.

Example 9.4

Some examples of upper-triangular matrices include $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$.

Theorem 9.2

Let $T \in \mathcal{L}(V)$ and $\alpha = v_1, \ldots, v_n$ be a basis of V. Then, the following are equivalent:

1. $[T]_\alpha^\alpha$ is upper triangular
2. $Tv_j \in \operatorname{span}(v_1, \ldots, v_j)$ for each $j = 1, \ldots, n$
3. $\operatorname{span}(v_1, \ldots, v_j)$ is T-invariant for each $j = 1, \ldots, n$

Proof: Suppose

$$[T]_\alpha^\alpha = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{array} \begin{array}{c} \begin{array}{ccccc} T(v_1) & T(v_2) & T(v_3) & \ldots & T(v_n) \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ 0 & a_{22} & a_{23} & \ldots & a_{2n} \\ 0 & 0 & a_{33} & \ldots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & a_{nn} \end{pmatrix} \end{array}$$

Note that statement 1 $\iff$ statement 2 is apparent from the definition of upper triangular matrices and the picture above. Moreover, statement 3 already implies statement 2 so the only implication we really need to show is from statement 2 to statement 3. We know that

$$\begin{aligned} Tv_1 &\in \operatorname{span}(v_1) && \subseteq \operatorname{span}(v_1, \ldots, v_j) \\ Tv_2 &\in \operatorname{span}(v_1, v_2) && \subseteq \operatorname{span}(v_1, \ldots, v_j) \\ &\vdots \\ Tv_j &\in \operatorname{span}(v_1, \ldots, v_j) && \subseteq \operatorname{span}(v_1, \ldots, v_j) \end{aligned}$$

Thus, for any $v \in \operatorname{span}(v_1, \ldots, v_j)$, the vector $Tv \in \operatorname{span}(v_1, \ldots, v_j)$, making it T-invariant as desired.

Theorem 9.3
Let V be a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then, there is a basis α of V such that $[T]_\alpha^\alpha$ is upper triangular, i.e., the operator T is "upper triangular".

Proof: We will proceed by induction on $n = \dim V$:

- Base Case: For $n = 1$, this statement is trivially true since there is an $a \in \mathbb{C}$ such that for any basis α of V, $[T]_\alpha^\alpha = \begin{bmatrix} a \end{bmatrix}$.
- Suppose now that $\dim V > 1$ and the desired result holds for all complex vector spaces with dimension less than $\dim V$. Let λ be an eigenvalue of T (which exists since V is complex and $\dim V > 0$). Moreover, let $U = \text{range}(T - \lambda I)$. Since $T - \lambda I$ is not surjective, U is a proper subspace of V. In other words, $\dim U < \dim V$.

  Note that U is still T-invariant. Why? For $u \in U$, write $Tu = (T - \lambda I)u + \lambda u$. Both parts of the sum are in U so $Tu \in U$. Thus, $T|_U \in \mathcal{L}(U)$.

  By the induction hypothesis, there must be a basis $\gamma = u_1, \ldots, u_m$ of U where $[T|_U]_\gamma^\gamma$ is upper triangular. Following the theorem above, for each $j \le \dim U = m$, we get that $Tu_j = (T|_U)(u_j) \in \text{span}(u_1, \ldots, u_j)$.

  Expand $u_1, \ldots, u_m$ to a basis $\alpha = u_1, \ldots, u_m, v_1, \ldots, v_l$ of V such that $l = n - m$. Write $Tv_k = (T - \lambda I)v_k + \lambda v_k$ and observe that $(T - \lambda I)v_k \in \text{range}(T - \lambda I) = U = \text{span}(u_1, \ldots, u_m)$. Then,

  $$Tv_k \in \text{span}(u_1, \ldots, u_m, v_k) \subseteq \text{span}(u_1, \ldots, u_m, v_1, \ldots, v_k)$$

  for all k. Thus, $[T]_\alpha^\alpha$ is upper triangular.

Theorem 9.4
If $T \in \mathcal{L}(V)$ has an upper triangular matrix representation $[T]_\alpha^\alpha$, then T is invertible iff all the entries along the main diagonal of $[T]_\alpha^\alpha$ are nonzero.

Proof: Let $\alpha = v_1, \ldots, v_n$ be the basis of V with respect to which T has an upper-triangular matrix representation. We will prove both directions of this theorem:

- Let the diagonal entries $\lambda_1, \ldots, \lambda_n$ of $[T]_\alpha^\alpha$ be nonzero. Observe that $Tv_1 = \lambda_1 v_1$ or, equivalently, $v_1 = T(\frac{v_1}{\lambda_1})$. Thus, $v_1 \in \text{range}(T)$.

  Next, $T(\frac{v_2}{\lambda_2}) = av_1 + v_2$ for some $a \in \mathbb{F}$ such that $a = \frac{a_{12}}{\lambda_2}$. Since $T(\frac{v_2}{\lambda_2})$ and av_1 are both in $\text{range}(T)$, then so is v_2 (since $\text{range}(T)$ is a subspace). We can continue this process (because we keep dividing by non-zero λ_i only) and notice that all v_i for $i \le n$ are in $\text{range}(T)$. Thus, $\text{span}(v_1, \ldots, v_n) = \text{range}(T)$ so T is surjective, and thus invertible.
- First, since $Tv_1 = \lambda v_1$ and $v_1 \neq 0$, λ_1 and v_1 are an eigenvalue-eigenvector pair. Since T is invertible, $\lambda_1 \neq 0$.

  Next, let $1 \le j \le n$ and suppose $\lambda_j = 0$. Then, T maps $\text{span}(v_1, \ldots, v_j)$ to $\text{span}(v_1, \ldots, v_{j-1})$. However, this would mean that T restricted to $\text{span}(v_1, \ldots, v_j)$ is not injective, i.e., there is some $v \in \text{span}(v_1, \ldots, v_j)$ such that $v \neq 0$ but $Tv = 0$. This would contradict the invertibility of T, so λ_j must be nonzero.

Theorem 9.5
If $T \in \mathcal{L}(V)$ has some upper triangular matrix representation $[T]_\alpha^\alpha$, then the eigenvalues of T are the diagonal entries of $[T]_\alpha^\alpha$.

Proof: Note that

$$[T]_\alpha^\alpha = \begin{bmatrix} \lambda_1 & * & * & \ldots & * \\ 0 & \lambda_2 & * & \ldots & * \\ 0 & 0 & \lambda_3 & \ldots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$

Let λ be some eigenvalue of T. Then,

$$[T-\lambda I]_\alpha^\alpha = \begin{bmatrix} \lambda_1-\lambda & * & * & \dots & * \\ 0 & \lambda_2-\lambda & * & \dots & * \\ 0 & 0 & \lambda_3-\lambda & \dots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n-\lambda \end{bmatrix}$$

According to the theorem above, this matrix is not invertible iff $\lambda_j - \lambda = 0 \implies \lambda_j = \lambda$ for some j. Since $T-\lambda I = T-\lambda_j I$ is only non-invertible for T's eigenvalues, the diagonal entries $\lambda_1, \dots, \lambda_n$ must indeed be the eigenvalues of T.

Example 9.5: 5B Exercise 14

Give an example of an operator whose matrix with respect to some basis contains only 0s on the diagonal, but the operator is still invertible.

Answer: Consider $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ to be the reflection about the line $y = x$. Then,

$$[T]_e^e = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

For a different example, consider $S : \mathbb{R}^2 \mapsto \mathbb{R}^2$, which is defined as a rotation counterclockwise by $\frac{\pi}{2}$ radians. Then,

$$[S]_e^e = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

This example demonstrates that the upper-triangular condition is necessary to prove non-invertibility.

Example 9.6: 5B Exercise 15

Give an example of an operator whose matrix with respect to some basis contains only non-zero numbers on the diagonal, but the operator is not invertible.

Answer: Consider $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ to be the orthogonal projection onto the line $y = x$. Then,

$$[T]_e^e = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

is not invertible since T is not surjective.

10 Lecture 10

10.1 Diagonal Matrices

Definition 10.1: Diagonal
A matrix $A \in \mathbb{R}^{n\times n}$ is diagonal if $A_{ij} = 0$ for $i \neq j$, i.e., A is of the form

$$A = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

Note 10.1
If $T \in \mathcal{L}(V)$ and $[T]_\alpha^\alpha$ is a diagonal matrix with $\lambda_1, \ldots, \lambda_n$ along its diagonal (refer to the matrix above), then each λ_j is an eigenvalue of T since the basis $\alpha = v_1, \ldots, v_n$ will tell us that the jth column of $[T]_\alpha^\alpha$ yields $Tv_j = \lambda_j v_j$.

Definition 10.2: Eigenspace
If $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T, then

$$E(\lambda, T) = E_\lambda(T) = E_\lambda = \{\lambda - \text{eigenvectors of } T\} \cup \{0\} = \ker(T - \lambda I)$$

is the λ-eigenspace of T

Theorem 10.1
Each $E(\lambda, T)$ is T-invariant.

Proof: There are many proofs of this fact. Here are two for fun:

- If $v \in E(\lambda, T)$, then $Tv = \lambda v$ so $T(T(v)) = T(\lambda v) = \lambda Tv$. So, either $Tv = 0$ or Tv is a λ-eigenvector.
- Another way to view this:

  $$(T - \lambda I)v = 0 \implies T \circ (T - \lambda I)v = 0 \implies (T \circ T - T \circ \lambda I)v = 0 \implies (T \circ T - \lambda I \circ T)v = 0 \implies (T - \lambda I) \circ Tv = 0$$

  so $Tv \in \ker(T - \lambda I) = E(\lambda, T)$.

Definition 10.3: Diagonalizable
$T \in \mathcal{L}(V)$ is called diagonalizable if the operator T has a diagonal matrix with respect to some basis of V, i.e., there is a basis α such that $[T]_\alpha^\alpha$ is diagonal.

Example 10.1
Let $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ be the orthogonal projection onto the plane spanned by linearly independent vectors v_1 and v_2. Let $\text{span}(v_3)$ be the normal line to this plane (i.e., $v_3 = v_1 \times v_2$). Let α be the list of these three vectors. Then,

$$[T]_\alpha^\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Example 10.2

Following the previous example, consider the plane $x - y + 2z = 0$. Then,

$$\alpha = \{v_1, v_2, v_3\} = \left\{ \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \right\}$$

Note that

$$\begin{aligned} [T]_e^e &= [I]_\alpha^e [T]_\alpha^\alpha ([I]_\alpha^e)^{-1} \\ &= \begin{bmatrix} -2 & 1 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -2 & 1 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & 2 \end{bmatrix}^{-1} \end{aligned}$$

is *not* diagonal. In other words, the diagonal matrix representation of an operator depends on the choice of basis.

Theorem 10.2
Let V be finite dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T. Then, the following statements are equal:

1. T is diagonalizable
2. V has a basis consisting of the eigenvectors of T — this is also called the eigenbasis
3. There are 1-dimensional T-invariant subspaces $U_1, \ldots, U_n$ such that $V = U_1 \oplus \cdots \oplus U_n$
4. $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$
5. $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$

Proof: We will prove multiple equivalences:

- statement 1 $\iff$ statement 2

  As already shown above,

  $$[T]_\alpha^\alpha = \begin{bmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_n \end{bmatrix}$$

  is diagonalizable for the basis of eigenvectors $\alpha = v_1, \ldots, v_n$ satisfying $Tv_j = \mu_j v_j$ for each j. There are λ_m distinct eigenvalues but some of them can be repeated, i.e., $Tv_i = \lambda v_i$ and $Tv_j = \lambda v_j$ for the same λ so we will take $\mu_1, \ldots, \mu_n$ to be the sequence of all eigenvalues that permit repetitions.

- statement 2 $\iff$ statement 3

  If $v_1, \ldots, v_n$ is the eigenbasis T, let $U_j = \operatorname{span}(v_j)$. Then, U_j is a 1-dimensional T-invariant subspace of V. Since $v = c_1 v_1 + \cdots + c_n v_n$ for $c_i v_i \in U_i$ for some $c_1, \ldots, c_n \in \mathbb{F}$, we have that $V = U_1 + \cdots + U_n$.

  Now suppose $0 = u_1 + \cdots + u_n$ where $u_j \in U_j$ for each $j \leq n$. Since $u_j = c_j v_j$, we know that $0 = c_1 v_1 + \cdots + c_n v_n$. However, $v_1, \ldots, v_n$ are linearly independent so $c_1 = \cdots = c_n = 0$, and $u_j = 0$. Thus, $U_1 \cap \cdots \cap U_n = \{0\}$ and $V = U_1 \oplus \cdots \oplus U_n$ is a direct sum.

  Lets prove the other direction. If U_j is 1-dimensional and T-invariant, then any $v_j \neq 0$ with $v_j \in U_j$ is an eigenvector of T. Thus, $v_1, \ldots, v_n$ spans V since $V = U_1 + \cdots + U_n$ and the linear independence of eigenvectors implies that $v_1, \ldots, v_n$ is the basis of T.

- statement 2 $\implies$ statement 4

  The eigenbasis of T implies that any vector in V is a linear combination of $v_1, \ldots, v_n$. So, $V = E(\lambda_1, T) + \cdots + E(\lambda_m, T)$. However, this sum is also direct — if $0 = w_1 + \cdots + w_m$ for $w_j \in E(\lambda_j, T)$, then the linear independence of eigenvectors (of distinct eigenvalues) implies that $w_j = 0$ for all j. Thus, $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$.

- statement 4 $\implies$ statement 5

  This is immediately follows from the fact that $V = U \oplus W \implies \dim V = \dim U + \dim W$.

- statement 5 $\implies$ statement 2

  Suppose $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$. Choose bases for each $E(\lambda_j, T)$ and combine them to form a list $v_1, \ldots, v_n$ of the eigenvectors of T such that $\dim V = n$.

  Suppose $c_1 v_1 + \cdots + c_n v_n = 0$. Let u_j be the sum of the terms $c_k v_k$ such that $v_k \in E(\lambda_j, T)$. Then, $u_j \in E(\lambda_j, T)$ and $u_1 + \cdots + u_m = 0$. Since the eigenvectors corresponding to different eigenvalues are linearly independent, $u_j = 0$. However, $0 = u_j = \sum_k c_k v_k$ as defined above, but as all v_k are linearly independent as well (they form the basis of $E(\lambda_j, T)$), all $c_k = 0$. Thus, $0 = u_1 + \cdots + u_m = c_1 v_1 + \cdots + c_n v_n$ implies that all $v_1, \ldots, v_n$ are linearly independent as a whole and are, consequently, a basis.

Example 10.3

Let $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$ such that $T(e_1) = e_1$ and $T(e_2) = e_1 + e_2$. Then,

$$[T]_e^e = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Is T diagonalizable?

Answer: Note that $\lambda = 1$ is the only eigenvalue of T (since $[T]_e^e$ is upper triangular, its eigenvalues lie along its diagonal). Then,

$$E(1, T) = \ker(T - I) = \ker([T - I]_e^e) = \ker([T]_e^e - [I]_e^e) = \ker\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = \operatorname{span}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$$

Thus $\dim \mathbb{R}^2 = 2 \neq 1 = \dim E(1, T)$, making T not diagonalizable.

Theorem 10.3

If $T \in \mathcal{L}(V)$ has $n = \dim V$ distinct eigenvalues, then T is diagonalizable.

Proof: Suppose $\lambda_1, \ldots, \lambda_n$ are n distinct eigenvalues of T. Choose corresponding eigenvectors $v_1, \ldots, v_n$. Since they are linearly independent by the distinctiveness of $\lambda_1, \ldots, \lambda_n$, they form a basis of V and the matrix representation of T with respect to this basis looks like

$$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

Observe that the matrix above is clearly diagonal, and each of its eigenspaces will be 1-dimensional.

Example 10.4

If $T \in \mathcal{L}(\mathbb{R}^2)$ is represented by $[T]_e^e = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$ instead, it is diagonalizable.

10.2 Similar Matrices

Consider the following commutative diagram:

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "{[T]_\alpha^\alpha}"] \arrow[d, "?"'] & \mathbb{R}^n \arrow[d, "?"] \\ \mathbb{R}^n \arrow[r, "{[T]_\beta^\beta}"'] & \mathbb{R}^n \end{tikzcd}$$

The top layer is the matrix representation of T in the α basis, while the bottom layer is with respect to the β basis. Can you find vertical maps (the ?s) to make the following diagram commute? Yes! Consider the following diagram:

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "{[T]_\alpha^\alpha}", "\cong"'] & \mathbb{R}^n \\
V \arrow[u, "\tilde{\alpha}", "\cong"'] \arrow[r, "T"] \arrow[d, "\tilde{\beta}"', "\cong"] & V \arrow[u, "\cong", "\tilde{\alpha}"'] \arrow[d, "\cong"', "\tilde{\beta}"] \\
\mathbb{R}^n \arrow[r, "\cong", "{[T]_\beta^\beta}"'] & \mathbb{R}^n
\end{tikzcd}$$

Notice that

$$\begin{aligned}
[T]_\beta^\beta &= (\tilde{\beta} \circ \tilde{\alpha}^{-1})[T]_\alpha^\alpha(\tilde{\alpha} \circ \tilde{\beta}^{-1}) \\
&= [I]_\alpha^\beta [T]_\alpha^\alpha [I]_\beta^\alpha
\end{aligned}$$

As we have already known so far!

The main idea behind change of basis is as follows: we can convert from a basis β to a basis α, apply some transformation T in basis α, and then convert back to basis β. This composition will achieve the same effect as applying the transformation T in basis β directly.

For the sake of notation, let $P = [I]_\beta^\alpha$ be the change of basis matrix from the "new" basis (the basis in which we are seeking to represent T) to the "old" basis (the basis in which we already know the representation of T). Consequently, the matrix P^{-1} can be interpreted as the map that will retrieve and convert to the "new" basis the result of applying T in the "old" basis. In other words, the linear map above can be rewritten as $[T]_\beta^\beta = P^{-1}[T]_\alpha^\alpha P$.

We can generalize this idea further by the concept of similar matrices.

Definition 10.4: Similar
Let B, A be $n \times n$ matrices. Then, B and A are similar iff there is some invertible matrix P such that $B = P^{-1}AP$.

Theorem 10.4
Matrices A and B are similar iff for any isomorphism ϕ between $V \cong \mathbb{R}^n$, there exists a $T \in \mathcal{L}(V)$ and bases α and β such that $A = [T]_\alpha^\alpha$ and $B = [T]_\beta^\beta$.

Proof: Slight note about notation: let the matrices and their corresponding linear transformations be denoted by the same symbol. For example, P is a matrix but it also denotes the operator $P : \mathbb{R}^n \mapsto \mathbb{R}^n$ such that $P(x) = Px$. We will now prove both directions of the biconditional:

- If $B = [T]_\beta^\beta$ and $A = [T]_\alpha^\alpha$ for some $T \in \mathcal{L}(V)$, then we can simply take $P = [I]_\beta^\alpha$ and let $B = P^{-1}AP$.
- Given the isomorphism ϕ and $B = P^{-1}AP$ for some P, let $\beta = v_1, \ldots, v_n$ where $v_j = \phi^{-1}(e_j)$, i.e., let $\tilde{\beta} = \phi$. Now, let $\tilde{\alpha} = P \circ \tilde{\beta} = P \circ \phi \implies \tilde{\alpha}^{-1} = \phi^{-1} \circ P^{-1}$ (why? it will make sense at the end). With this, let $u_j = \phi^{-1}(P^{-1}e_j)$ and take $\alpha = u_1, \ldots, u_n$. This gives us our choice of bases for a given ϕ and P.

  We want $B = [T]_\beta^\beta$ for some $T \in \mathcal{L}(V)$. Note that setting $T = \tilde{\beta}^{-1} \circ B \circ \tilde{\beta}$ will work based on the following commutative diagram:

$$\begin{tikzcd}
V \arrow[r, "T"] \arrow[d, "\tilde{\beta}"'] & V \arrow[d, "\tilde{\beta}"] \\
\mathbb{R}^n \arrow[r, "{[T]_\beta^\beta}"'] & \mathbb{R}^n
\end{tikzcd}$$

Since $A = PBP^{-1}$ and $[T]_\alpha^\alpha = [I]_\beta^\alpha [T]_\beta^\beta [I]_\alpha^\beta$, we can simply let $A = [T]_\alpha^\alpha$ and $P = [I]_\beta^\alpha$. This further justifies our choice for defining $\tilde{\alpha} = P \circ \tilde{\beta}$ earlier since $[I]_\beta^\alpha$ is precisely defined as $\tilde{\alpha} \circ \tilde{\beta}^{-1}$, as we saw in chapter earlier!

Example 10.5: 5C Exercise 2

Let $V = \ker(T) \oplus \operatorname{range}(T)$. Is $T \in \mathcal{L}(V)$ diagonalizable? No. Consider

$$[T]_e^e = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

It is non-diagonalizable but still invertible. Thus, $\ker(T) = \{0\}$ (from injectivity) and $\operatorname{range}(T) = V$ (from surjectivity), satisfying $V = \ker(T) \oplus \operatorname{range}(T) = \{0\} \oplus V$.

10.3 Discussion Worksheet

Problem 10.1

Prove that if v_1 and v_2 are eigenvectors of T with distinct eigenvalues, then v_1 and v_2 are linearly independent.

Problem 10.2

Show that T and $S^{-1}TS$ have the same eigenvalues. What is the relationship between the eigenvectors of T and those of $S^{-1}TS$?

Problem 10.3

Suppose that u, v and $u + v$ are eigenvectors. Prove that they correspond to the same eigenvalues.

Problem 10.4

Suppose $T \in \mathcal{L}(V)$ and $T^n = 0$ for some integer $n > 0$. Prove that $I - T$ is invertible and that $(I-T)^{-1} = I + T + \cdots + T^{n-1}$.

Problem 10.5

Let $p \in \mathcal{P}(\mathbb{F})$, $S, T \in \mathcal{L}(V)$ and S be invertible. Prove that $p(STS^{-1}) = Sp(T)S^{-1}$.

Problem 10.6

Suppose $T \in \mathcal{L}(V)$ and v be a λ-eigenvector of T. If $p \in \mathcal{P}(\mathbb{F})$, then prove that $p(T)v = p(\lambda)v$.

11 Lecture 11

11.1 Inner Products

Example 11.1
Let $V = \mathbb{R}^n$. Recall that the dot product of two vectors $u, v \in \mathbb{R}^n$ is given by $u \cdot v = u_1 v_1 + \cdots + u_n v_n$.

The usual Euclidean notions of the "length" of a vector and the angle between two non-zero vectors in $\mathbb{R}^n$ can be completely recovered in terms of the dot product:

1. Length of $u \in \mathbb{R}^n$ is $\|u\| = \sqrt{u \cdot u} = \sqrt{u_1^2 + \cdots + u_n^2}$
2. Angle between $u, v \in \mathbb{R}^n$. From the law of cosines, notice that

$$\begin{aligned}
\|v - u\|^2 &= \|u\|^2 + \|v\|^2 - 2\|u\|\|v\| \cos\theta \\
(v - u) \cdot (v - u) &= \|u\|^2 + \|v\|^2 - 2\|u\|\|v\| \cos\theta \\
\|v\|^2 - 2(u \cdot v) + \|u\|^2 &= \|u\|^2 + \|v\|^2 - 2\|u\|\|v\| \cos\theta \\
\|u\|\|v\| \cos\theta &= u \cdot v \\
\theta &= \arccos\left(\frac{u \cdot v}{\|u\|\|v\|}\right)
\end{aligned}$$

Note 11.1
Therefore, u and v are perpendicular (orthogonal) iff $u \cdot v = 0$. We also declare that $u = 0$ is orthogonal to all vectors in a vector space.

However, the length and angle notions determined by the dot product are not the only ones we could impose on $\mathbb{R}^n$. We can generalize the dot product by axiomatizing and abstracting its key features, in a similar way to our abstractions of addition and scalar multiplication (first lecture) to the concept of a vector space over a field. Experimentation, time and reflection lead to the abstract notions of an "inner product" on V and an associated "inner-product space".

Definition 11.1: Inner Product
Let V be a vector space over $\mathbb{F}$. An inner product $\langle \cdot, \cdot \rangle$ on V is a map $\langle \cdot, \cdot \rangle : V \times V \mapsto \mathbb{F}$ satisfying the following properties:

- Positivity: $\langle u, u \rangle \geq 0$ for all $u \in V$
- Definiteness: $\langle u, u \rangle = 0$ iff $u = 0$
- Additivity: $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ for all $u, v, w \in V$
- Homogeneity: $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$ for all $u, v \in V, \lambda \in \mathbb{F}$
- Conjugate Symmetry: $\langle u, v \rangle = \overline{\langle v, u \rangle}$ for all $u, v \in V$

Definition 11.2: Inner Product Space
The collection of a vector space V and its inner product $\langle \cdot, \cdot \rangle$ is known as an inner product space.

Note 11.2
$\langle u, \lambda v \rangle = \overline{\langle \lambda v, u \rangle} = \overline{\lambda \langle v, u \rangle} = \overline{\lambda}\ \overline{\langle v, u \rangle} = \overline{\lambda} \langle u, v \rangle$

Note 11.3
If $\mathbb{F} = \mathbb{R}$, then $z = \overline{z}$. Then, $\langle u, v \rangle = \langle v, u \rangle$ and $\langle u, \lambda v \rangle = \lambda \langle u, v \rangle$.

Example 11.2
Here are some examples of common inner products:

1. For $V = \mathbb{R}^n$, define the regular dot product $\langle u, v\rangle = u \cdot v = u_1 v_1 + \cdots + u_n v_n$
2. For $V = \mathbb{C}^n$, define the complex dot product $\langle u, v\rangle = u \cdot \overline{v} = u_1\,\overline{v_1} + \cdots + u_n\,\overline{v_n}$
3. For $V = l^2(\mathbb{N})$ (sequences with values in $\mathbb{F}$ such that $\sum_{i=1}^{\infty} z_i\,\overline{z_i} < \infty$), define $\langle z, w\rangle = \sum_{i=1}^{\infty} z_i\,\overline{w_i}$
4. For $V = L^2([1,-1]) = \left\{f \in C([-1,1]) \mid \int_{-1}^{1} f(x)\,\overline{f(x)}\,\mathrm{d}x < \infty\right\}$, define $\langle f, g\rangle = \int_{-1}^{1} f(z)\,\overline{g(z)}\,\mathrm{d}z$

Note 11.4
In general, the complex dot product is generally also called the Euclidean inner product.

Example 11.3
Let $V = \mathbb{F}^n$. The Euclidean inner product is given by $\langle z, w\rangle = z_1\,\overline{w_1} + \cdots + z_n\,\overline{w_n}$. Why not define it as simply $\langle z, w\rangle = z_1 w_1 + \cdots + z_n w_n$? Because we want $\langle z, z\rangle \geq 0$ for all $z \in V$ but $z = i$ for $\mathbb{F} = \mathbb{C}$ makes it so that $\langle i, i\rangle = i \cdot i = -1 < 0$.

Example 11.4
Let $V = \mathbb{F}^n$ and $\langle\cdot,\cdot\rangle$ be an inner product on V. Suppose A is a matrix of the form $A = BB^*$ where B is an $n \times n$ matrix, A is invertible and $B^* = \overline{B^T}$ is the conjugate transpose of B (more on this in the next chapter). Then, $\langle u, v\rangle = \langle Au, v\rangle$ is also an inner product on V. Actually, any inner product on V has this form for some A where $\langle\cdot,\cdot\rangle$ can be taken to be the Euclidean inner product on V.

Definition 11.3: Norm and Angle
If $\langle\cdot,\cdot\rangle$ is an inner product on V, then $\|u\| = \sqrt{\langle u, u\rangle}$ is the norm ("magnitude" or "length") of u. Similarly,

$$\theta = \arccos\left(\frac{\langle u, v\rangle}{\|u\|\|v\|}\right)$$

is the angle between u and v for $u, v \neq 0$.

Definition 11.4: Orthogonal
If $\langle u, v\rangle$ is an inner product on V, then u and v are orthogonal if $\langle u, v\rangle = 0$.

Theorem 11.1
0 is the only vector orthogonal to itself.

Proof: $\langle v, v\rangle = 0$ iff $v = 0$ is an axiom for $\langle\cdot,\cdot\rangle$ to be a valid inner product.

Theorem 11.2: Pythagorean Theorem
Let u, v be orthogonal in an inner product space V. Then, $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.

Proof:

$$\begin{aligned}\|u + v\|^2 &= \langle u + v, u + v\rangle \\ &= \langle u + u\rangle + \langle v, u\rangle + \langle u, v\rangle + \langle v, v\rangle \\ &= \|u\|^2 + \langle v, u\rangle + \langle u, v\rangle + \|v\|^2\end{aligned}$$

If u, v are orthogonal, then $\langle u, v\rangle = \overline{\langle v, u\rangle} = 0$. Then, $\langle v, u\rangle = 0$ too. Thus, $\|u+v\|^2 = \|u\|^2 + \|v\|^2$ as expected.

Definition 11.5: Orthogonal Projection
Let $u, v \in V$ and $b \neq 0$ for an inner product space V with inner product $\langle \cdot, \cdot \rangle$. Then, there is a unique $c \in \mathbb{F}$ and a unique $w \in V$ such that $u = cv + w$ and $\langle v, w\rangle = 0$. This unique cv is called the orthogonal projection of u onto $\operatorname{span}(v)$ or $\operatorname{proj}_v u$ or $\operatorname{proj}_{\operatorname{span}(v)} u$. The vector w is, thus, uniquely determined to be $u - \operatorname{proj}_{\operatorname{span}(v)} u$.

Proof: We require that $u = cv + w$ with $\langle v, w\rangle = \langle w, v\rangle = 0$. Then, $w = u - cv$ and,

$$\begin{aligned} \langle w, v\rangle &= 0 \\ \langle u - cv, v\rangle &= 0 \\ \langle u, v\rangle - c\langle v, v\rangle &= 0 \end{aligned}$$

This implies that $c = \frac{\langle u,v\rangle}{\langle v,v\rangle} = \frac{\langle u,v\rangle}{\|v\|^2}$. Then, $\operatorname{proj}_{\operatorname{span}(v)} u = cv = \frac{\langle u,v\rangle}{\langle v,v\rangle} v$.

Theorem 11.3: Cauchy-Schwarz Inequality
Let V be an inner product space with $\langle \cdot, \cdot \rangle$ as its inner product. Then, $|\langle u, v\rangle| \leq \|u\|\|v\|$ for all $u, v \in V$ with equality iff u and v are linearly dependent/parallel.

Proof: If $v = 0$, then $|\langle u, 0\rangle| = 0 = \|u\|\|0\|$. Otherwise, $u = \operatorname{proj}_{\operatorname{span}(v)} u + (u - \operatorname{proj}_{\operatorname{span}(v)} u)$. Let $w = u - \operatorname{proj}_{\operatorname{span}(v)} u$ for convenience. Then,

$$\begin{aligned} \|u\|^2 &= \left\|\operatorname{proj}_{\operatorname{span}(v)} u\right\|^2 + \|w\|^2 \\ &\geq \left\|\operatorname{proj}_{\operatorname{span}(v)} u\right\|^2 \\ &= \left\|\frac{\langle u, v\rangle}{\langle v, v\rangle} v\right\|^2 \\ &= \frac{|\langle u, v\rangle|^2}{\|v\|^2} \end{aligned}$$

The first statement follows from the Pythagorean theorem. Then,

$$\begin{aligned} \|u\|^2\|v\|^2 &\geq |\langle u, v\rangle|^2 \\ \|u\|\|v\| &\geq |\langle u, v\rangle| \end{aligned}$$

We have an equality iff $w = 0$, i.e., $u - \operatorname{proj}_{\operatorname{span}(v)} u = 0 \implies u = \operatorname{proj}_{\operatorname{span}(v)} u$ since this implies that u and v lie on the same line (i.e., they are linearly dependent/parallel).

Theorem 11.4: Triangle Inequality
$\|u + v\| \leq \|u\| + \|v\|$. This is an equality iff u and v are linearly dependent via a non-negative scalar.

Proof: Quick fact: the absolute value of a complex number $z = a + bi$ is $|z| = \sqrt{a^2 + b^2}$ so $\operatorname{Re}(z)^2 = a^2 \leq a^2 + b^2 = |z|^2 \implies \operatorname{Re}(z) \leq |z|$. Now,

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v\rangle \\ &= \|u\|^2 + \langle v, u\rangle + \langle u, v\rangle + \|v\|^2 \\ &= \|u\|^2 + \underbrace{\langle u, v\rangle + \overline{\langle u, v\rangle}}_{2\operatorname{Re}(\langle u,v\rangle)} + \|v\|^2 \\ &\leq \|u\|^2 + 2|\langle u, v\rangle| + \|v\|^2 \\ &\leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \end{aligned}$$

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

Taking the square root, $\|u+v\| \leq \|u\| + \|v\|$. This is an equality iff $\|u\|\|v\| = |\langle u, v\rangle| = \mathrm{Re}(\langle u, v\rangle)$. Cauchy-Schwarz says that $\|u\|\|v\| = |\langle u, v\rangle|$ is true iff u and v are linearly independent. Moreover, $\mathrm{Re}(\langle u, v\rangle) = |\langle u, v\rangle| \geq 0$ is true iff the scalar relating u to v is also non-negative.

Theorem 11.5: Parallelogram Equality/Identity

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

Proof: Just expand both sides:

$$\begin{aligned}\|u+v\|^2 + \|u-v\|^2 &= \langle u+v, u+v\rangle + \langle u-v, u-v\rangle \\ &= \langle u,u\rangle + \langle v,u\rangle + \langle u,v\rangle + \langle v,v\rangle + \langle u,u\rangle - \langle v,u\rangle - \langle u,v\rangle + \langle v,v\rangle \\ &= 2(\langle u,u\rangle + \langle v,v\rangle) \\ &= 2(\|u\|^2 + \|v\|^2)\end{aligned}$$

Example 11.5: 6A Exercise 11

Prove that if a, b, c, d are all ≥ 0, then

$$16 \leq (a+b+c+d)\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)$$

1. Attempt #1

   Let $u = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$ and $v = \begin{bmatrix} 1/a \\ 1/b \\ 1/c \\ 1/d \end{bmatrix}$. Then, $\langle u, v\rangle = 4 \implies |\langle u, v\rangle|^2 = 16$. Cauchy-Schwarz gives us:

   $$\begin{aligned}|\langle u, v\rangle|^2 &\leq \|u\|^2\|v\|^2 \\ &= \langle u,u\rangle\langle v,v\rangle \\ &= (a^2+b^2+c^2+d^2)\left(\frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{c^2} + \frac{1}{d^2}\right)\end{aligned}$$

   This, is not exactly what we want. Lets try something else.

2. Attempt #2

   Now, let $u = \begin{bmatrix} \sqrt{a} \\ \sqrt{b} \\ \sqrt{c} \\ \sqrt{d} \end{bmatrix}$ and $v = \begin{bmatrix} 1/\sqrt{a} \\ 1/\sqrt{b} \\ 1/\sqrt{c} \\ 1/\sqrt{d} \end{bmatrix}$. Then,

   $$\begin{aligned}|\langle u, v\rangle|^2 &\leq \|u\|^2\|v\|^2 \\ \left|\sqrt{a}\cdot\frac{1}{\sqrt{a}} + \sqrt{b}\cdot\frac{1}{\sqrt{b}} + \sqrt{c}\cdot\frac{1}{\sqrt{c}} + \sqrt{d}\cdot\frac{1}{\sqrt{d}}\right|^2 &\leq \langle u, v\rangle\|v, v\| \\ |1+1+1+1|^2 &\leq \left(\sqrt{a}^2 + \sqrt{b}^2 + \sqrt{c}^2 + \sqrt{d}^2\right)\left(\frac{1}{\sqrt{a}^2} + \frac{1}{\sqrt{b}^2} + \frac{1}{\sqrt{c}^2} + \frac{1}{\sqrt{d}^2}\right) \\ 16 &\leq (a+b+c+d)\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)\end{aligned}$$

   So, Cauchy-Schwarz gives us the desired inequality.

Example 11.6: 6A Exercise 10

Find vectors $u, v \in \mathbb{R}^2$ such that u is a scalar multiple of $(1,3)$ and v is orthogonal to $(1,3)$ and $u + v = (1,2)$. Let $u = (c, 3c)$ for some $c \in \mathbb{R}$ and let $v = (a,b)$. Then $\langle v, (1,3)\rangle = (a,b)\cdot(1,3) = a + 3b = 0 \implies a = -3b$. Since, $u + v = (1,2)$, then

$$\begin{aligned}(c,3c) + (-3b,b) &= (1,2)\\ c - 3b &= 1\\ 3c + b &= 2\end{aligned}$$

Solving the system yields $b = -\frac{1}{10}$ and $c = \frac{7}{10}$. The, $u = \left(\frac{7}{10}, \frac{21}{10}\right)$ and $v = \left(\frac{3}{10}, -\frac{1}{10}\right)$.

11.2 Normed Vector Spaces

Definition 11.6: Norm

Let V be a vector space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A norm on V is a map $\|\cdot\| : V \mapsto \mathbb{R}$ such that

1. $\|v\| = 0$ iff $v = 0$, else $\|v\| > 0$ otherwise
2. $\|\lambda v\| = |\lambda|\|v\|$ for all $v \in V, \lambda \in \mathbb{F}$
3. $\|u + v\| \le \|u\| + \|v\|$ for all $u, v \in V$ (triangle inequality)

Example 11.7

If V is an inner product space, then $\|v\| = \sqrt{\langle v, v\rangle}$ makes V a normed vector space as well.

However, there are many interesting examples of normed vector spaces for which the norm is *not* induced by an inner product!

Example 11.8

Let p-norm on $x \in \mathbb{R}^n$ or $x \in \mathbb{C}^n$ for $p \ge 1$ be defined as

$$\begin{aligned}\|x\|_p &= (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}} \text{ for } p \neq \infty\\ \|x\|_\infty &= \max\{|x_1|, \ldots, |x_n|\}\end{aligned}$$

Observe that $\|\cdot\|_p$ is a valid norm on $\mathbb{F}^n$. However, establishing the triangle inequality for $p \neq \infty, p \neq 2$ is not an easy task, so it has a special name: Minkowski's Inequality (proved using Hölder's Inequality).

Note 11.5

Note that $\|\cdot\|_p$ indeed comes from an inner product iff $p = 2$.

Example 11.9: 6A Exercise 18, 6A Exercise 21

Prove the fact above for $\mathbb{R}^2$. Let $u = \begin{bmatrix}1\\0\end{bmatrix}$ and $v = \begin{bmatrix}0\\1\end{bmatrix}$. Then,

$$\begin{aligned}\|u + v\|_p^2 &= (1^p + 1^p)^{\frac{2}{p}}\\ &= 2^{\frac{2}{p}}\\ \|u - v\|_p^2 &= (1^p + |-1|^p)^{\frac{2}{p}}\\ &= 2^{\frac{2}{p}}\end{aligned}$$

$$2(\|u\|_p^2 + \|v\|_p^2) = 2\left((1^p + 0^p)^{\frac{2}{p}} + (0^p + 1^p)^{\frac{2}{p}}\right)$$
$$= 4$$

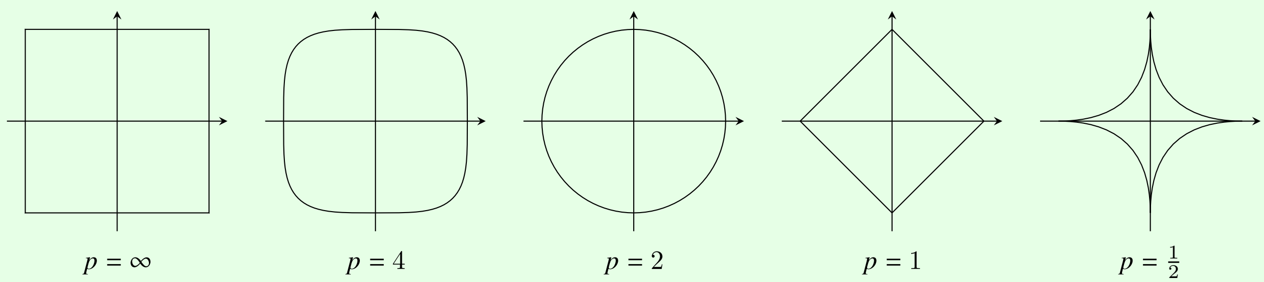
If $\|\cdot\|_p$ is induced by an inner product, then it would satisfy the parallelogram identity:

$$\|u+v\|_p^2 + \|u-v\|_p^2 = 2(\|u\|_p^2 + \|v\|_p^2)$$
$$2^{\frac{2}{p}} + 2^{\frac{2}{p}} = 4$$
$$2^{1+\frac{2}{p}} = 2^2$$

Then, $1 + \frac{2}{p} = 2 \implies p = 2$ is the only norm induced by an inner product, namely the Euclidean inner product.

Example 11.10

Here is a visualization of what different norms look like:



In order, from left to right:

- Unit circle for the ∞-norm given by $\left\{v \in \mathbb{R}^2 \mid \|v\|\infty = 1\right\}$. Note that $\|v\|_\infty = \max_{j\leq 2}\left|v_j\right|$, so this shape lies along $x = \pm 1$, $y = \pm 1$, creating a square.
- Unit circle for the 4-norm is given by $\left\{v \in \mathbb{R}^2 \mid \|v\|_4 = \sqrt[4]{|v_1|^4 + |v_2|^4} = 1\right\}$. It is also called a "hyper ellipse".
- Unit circle for the 2-norm is given by $\left\{v \in \mathbb{R}^2 \mid \|v\|_2 = \sqrt{|v_1|^2 + |v_2|^2} = 1\right\}$. This is the normal unit circle.
- Unit circle for the 1-norm is given by $\left\{v \in \mathbb{R}^2 \mid \|v\|_1 = |v_1| + |v_2| = 1\right\}$. This diamond shape can also be obtained manually by doing casework with absolute values.
- Unit circle for the $\frac{1}{2}$-norm is given by $\left\{v \in \mathbb{R}^2 \mid \|v\|_{\frac{1}{2}} = (\sqrt{v_1} + \sqrt{v_2})^2 = 1\right\}$. However, if $0 < p < 1$, then $\|\cdot\|_p$ actually is *not* a norm because it satisfies $\|u+v\|_p \geq \|u\|_p + \|v\|_p$ (the flipped triangle inequality) instead!

12 Lecture 12

A note on L^2-inner products. Consider $V = L^2([-1,1]) = C([-1,1],\mathbb{R})$, i.e., the set of real-valued functions continuous on the interval $[-1,1]$. Then, $\langle f,g\rangle = \int_{-1}^{1} f(x)g(x)\,\mathrm{d}x$ makes V a valid inner-product space. Here is a generalization of it:

$$V = L^2(\mathbb{R}) = \left\{ f : \mathbb{R} \mapsto \mathbb{R} \mid \int_{-\infty}^{\infty} |f(x)|^2\,\mathrm{d}x < \infty \right\}$$

Therefore, the inner product $\langle f,g\rangle = \int_{-\infty}^{\infty} f(x)g(x)\,\mathrm{d}x$ will make V a valid inner-product space again.

However, how do we know for $f,g \in L^2(\mathbb{R})$, the inner product $|\langle f,g\rangle| < \infty$? Recall that the Cauchy-Schwarz inequality says:

$$|\langle f,g\rangle| \le \|f\|\|g\|$$

$$\left|\int_{-\infty}^{\infty} f(x)g(x)\,\mathrm{d}x\right| \le \left(\int_{-\infty}^{\infty} |f(x)|^2\,\mathrm{d}x\right)^{\frac{1}{2}} \left(\int_{-\infty}^{\infty} |g(x)|^2\,\mathrm{d}x\right)^{\frac{1}{2}}$$

The right hand side is a product of two finite quantities so the inner product must be finite as well!

12.1 Orthonormality

Definition 12.1: Orthogonal and Orthonormal

A list of vectors is orthogonal if each vector in the list is orthogonal to all the other vectors in the list. A list of vectors is orthonormal if it is orthogonal and each vector in the list has norm 1.

Note 12.1

The list of standard basis vectors $e_1,\ldots,e_m$ is orthonormal since $\left\langle e_j, e_k\right\rangle = 0$ for $j \neq k$ and $\left\langle e_j, e_k\right\rangle = 1$ for $j = k$.

Note 12.2

Notation: from now on, the list $e_1,\ldots,e_m$ represents a general orthonormal list, and not necessarily a sub-collection of the standard basis. Everything should be fairly clear with context.

Example 12.1

Examples of orthonormal bases:

1. The standard basis of $\mathbb{F}^n$ (defined for the Euclidean inner product)
2. $\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},0\right), \left(\frac{1}{\sqrt{6}},\frac{1}{\sqrt{6}},-\frac{2}{\sqrt{6}}\right)$ is an orthonormal basis in $\mathbb{R}^3$ (again, with respect to the Euclidean inner product)

Theorem 12.1

If $e_1,\ldots,e_m$ is orthonormal in V, then $\|a_1e_1+\cdots+a_me_m\|^2 = |a_1|^2+\cdots+|a_m|^2$

Proof: Following the Pythagorean theorem,

$$\begin{aligned}\|a_1e_1+\cdots+a_me_m\|^2 &= \|a_1e_1\|^2+\cdots+\|a_me_m\|^2\\ &= |a_1|^2\|e_1\|^2+\cdots+|a_m|^2\|e_m\|^2\\ &= |a_1|^2+\cdots+|a_m|^2\end{aligned}$$

Theorem 12.2
If $e_1, \ldots, e_m$ is orthonormal, then $a_1 e_1 + \cdots + a_m e_m = 0 \implies a_1 = \cdots = a_m = 0$.

Proof: Following the results from above,

$$\begin{aligned} a_1 e_1 + \cdots + a_m e_m &= 0 \\ \|a_1 e_1 + \cdots + a_m e_m\|^2 &= \|0\|^2 \\ |a_1|^2 + \cdots + |a_m|^2 &= 0 \end{aligned}$$

Since $|a_i| \geq 0$ for $j \leq m$, $|a_1|^2 + \cdots + |a_m|^2 = 0 \implies a_1 = \cdots = a_m = 0$.

Some consequences of this theorem:

- A non-zero orthogonal list is always linearly independent.
- If $\dim V = n$ and $e_1, \ldots, e_n$ is orthonormal, then it is a basis.

Theorem 12.3
If $e_1, \ldots, e_m$ are orthonormal and $v = a_1 e_1 + \cdots + a_m e_m$, then $a_j = \langle v, e_j \rangle$ and

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_m \rangle|^2$$

Proof: For each $j \leq n$,

$$\begin{aligned} \langle v, e_j \rangle &= a_1 \langle e_1, e_j \rangle + \cdots + a_j \langle e_j, e_j \rangle + \cdots + a_m \langle e_m, e_j \rangle \\ &= a_1 \cdot 0 + \cdots + a_j \cdot 1 + \cdots + a_m \cdot 0 \\ &= a_j \end{aligned}$$

Then,

$$\begin{aligned} v &= \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m \\ \|v\|^2 &= |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_m \rangle|^2 \end{aligned}$$

The last equality simply follows from theorem 12.1.

Theorem 12.4: Gram-Schmidt Process
If $v_1, \ldots, v_m$ are linearly independent in V, consider the process given by

- Let $e_1 = \frac{v_1}{\|v_1\|}$
- For $j = 2, \ldots, m$, define e_j inductively to be

$$e_j = \frac{v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}}{\|v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}\|}$$

Then, $e_1, \ldots, e_m$ is an orthonormal list of vectors in V such that $\operatorname{span}(v_1, \ldots, v_m) = \operatorname{span}(e_1, \ldots, e_m)$ for each $j \leq m$.

The proof of the validity of this process is given in Axler. However, what is the intuition/idea behind these formulas? Note that e_1 is just v_1 normalized. Consider $v_2 - \operatorname{proj}_{\operatorname{span}(v_1)} v_2$. This is orthogonal to v_1 but $\operatorname{span}(e_1) = \operatorname{span}(v_1)$ so

$$\begin{aligned} v_2 - \operatorname{proj}_{\operatorname{span}(v_1)} v_2 &= v_2 - \operatorname{proj}_{\operatorname{span}(e_1)} v_2 \\ &= v_2 - \frac{\langle v_2, e_1 \rangle e_1}{\|e_1\|^2} \end{aligned}$$

$$= v_2 - \langle v_2, e_1 \rangle e_1$$
$$e_2 = \frac{v_2 - \langle v_2, e_1 \rangle e_1}{\|v_2 - \langle v_2, e_1 \rangle e_1\|}$$

Then, e_2 is a normal vector that is orthogonal to $\operatorname{span}(v_1) = \operatorname{span}(e_1)$. Note that v_2 can be rearranged to be a linear combination of both e_1 and e_2 so $v_2 \in \operatorname{span}(e_1, e_2)$. Moreover, as $v_1 \in \operatorname{span}(e_1) \subset \operatorname{span}(e_1, e_2)$, this further implies that $\operatorname{span}(v_1, v_2) = \operatorname{span}(e_1, e_2)$.

Now, consider the vector $e_3 = \frac{v_3 - \operatorname{proj}_{\operatorname{span}(v_1,v_2)} v_3}{\left\|v_3 - \operatorname{proj}_{\operatorname{span}(v_1,v_2)} v_3\right\|} = \frac{v_3 - \operatorname{proj}_{\operatorname{span}(e_1,e_2)} v_3}{\left\|v_3 - \operatorname{proj}_{\operatorname{span}(e_1,e_2)} v_3\right\|}$ that is a normal vector like e_1 and e_2.

$$\vdots$$

Repeat this process. At each intermediate step, you will consider the vector $e_j = \frac{v_j - \operatorname{proj}_{\operatorname{span}(v_1,\dots,v_{j-1})} v_j}{\left\|v_j - \operatorname{proj}_{\operatorname{span}(v_1,\dots,v_{j-1})} v_j\right\|}$.

However, there is an issue to deal with here. What does $\operatorname{proj}_{\operatorname{span}(v_1,\dots,v_{j-1})} v_j$ even mean in this context? We have so far only defined the concept of orthogonal projections for the case where the span is 1-dimensional.

Theorem 12.5
Suppose $x \in V$ and U is a subspace of V. Then, $x = y + w$ for unique $y \in U$ and unique w such that $\langle y, w \rangle = 0$. Define $\operatorname{proj}_U x = y$ and, thus, $w = x - \operatorname{proj}_U x$.

This is just an extension of the definition of orthogonal projection from earlier but it certainly explains the meaning of projecting onto a subspace.

Theorem 12.6
This is a special case of the theorem above. Suppose that $e_1, \dots, e_m$ is an orthonormal bases of U. Then, $\operatorname{proj}_U x = \langle x, e_1 \rangle e_1 + \cdots + \langle x, e_m \rangle e_m$.

Proof: Suppose that $x = y + w$ with $y \in U$ and $\langle y, w \rangle = 0$. So, $y = c_1 e_1 + \cdots + c_m e_m$. We already know that each $c_j = \langle y, e_j \rangle$. Therefore, if $\operatorname{proj}_U x$ indeed exists, it must be

$$\operatorname{proj}_U x = \langle x, e_1 \rangle e_1 + \cdots + \langle x, e_m \rangle e_m$$

However, this certainly must exist since $e_1, \dots, e_m$ does.

Using a simple algebraic calculation (that I have omitted here), you can verify that $\langle x, e_1 \rangle e_1 + \cdots + \langle x, e_m \rangle e_m$ has the desired property that it is orthogonal to $x - \langle x, e_1 \rangle e_1 - \cdots - \langle x, e_m \rangle e_m$.

Note 12.3
This argument shows that if $e_1, \dots, e_m$ is orthonormal, then

$$\operatorname{proj}_{\operatorname{span}(e_1,\dots,e_m)} = \operatorname{proj}_{\operatorname{span}(e_1)} + \operatorname{proj}_{\operatorname{span}(e_2)} + \cdots + \operatorname{proj}_{\operatorname{span}(e_m)}$$

We have effectively reduced the issue above to the following: we know that U being finite dimensional will imply that it has a basis, but we don't have an argument for whether it has an orthonormal basis. For that, we can either appeal to Gram-Schmidt, which would be circular reasoning, or we need to prove this fact independently. We will just assume for now that this statement holds (and proofs of it that function without Gram-Schmidt do indeed exist out there).

Then, the definition of e_j from earlier can be substituted with

$$e_j = \frac{v_j - \operatorname{proj}_{\operatorname{span}(e_1,\dots,e_{j-1})} v_j}{\left\|v_j - \operatorname{proj}_{\operatorname{span}(e_1,\dots,e_{j-1})} v_j\right\|} = \frac{v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}}{\left\|v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}\right\|}$$

Since $v_j = \operatorname{proj}_{\operatorname{span}(e_1,\dots,e_{j-1})} v_j + (v_j - \operatorname{proj}_{\operatorname{span}(e_1,\dots,e_{j-1})} v_j)$, the vector e_j (its un-normalized version, for now) is orthogonal to all vectors in $e_1, \dots, e_{j-1}$. Since all e_i are normal vectors, this further implies that $e_1, \dots, e_j$ is orthonormal as intended.

We will now prove the existence of an orthonormal basis using the Gram-Schmidt method, which is an easier proof than the ones alluded to above.

Theorem 12.7
Every finite dimensional inner product space V has an orthonormal basis.

Proof: Let $v_1, \dots, v_n$ be a basis of V. Apply the Gram-Schmidt process to obtain an orthonormal list $e_1, \dots, e_n$ such that $\operatorname{span}(e_1, \dots, e_n) = \operatorname{span}(v_1, \dots, v_n) = V$. Since $e_1, \dots, e_n$ is a spanning list of length n in V, it is n-dimensional and hence a basis.

Theorem 12.8
Suppose V is finite-dimensional and $e_1, \dots, e_m$ is an orthonormal list in V. Then, $e_1, \dots, e_m$ can be extended to an orthonormal basis of V.

Proof: Extend $e_1, \dots, e_m$ to a basis $e_1, \dots, e_m, v_1, \dots, v_{n-m}$ of V. Apply the Gram-Schmidt process on this list to obtain an orthonormal basis of V as desired.

Theorem 12.9
Let $T \in \mathcal{L}(V)$. Suppose $[T]_\beta^\beta$ is upper triangular for some basis β. Then, there exists another basis α of V such that $[T]_\alpha^\alpha$ is upper triangular too.

Proof: The matrix $[T]_\beta^\beta$ being upper triangular implies that $\operatorname{span}(v_1, \dots, v_j)$ is T-invariant for all $j \le n$ and $\beta = v_1, \dots, v_n$. Apply the Gram-Schmidt process to B to obtain an orthonormal basis $\alpha = e_1, \dots, e_n$. Since $\operatorname{span}(e_1, \dots, e_j) = \operatorname{span}(v_1, \dots, v_j)$ for all $j \le n$ throughout the process, then $\operatorname{span}(e_1, \dots, e_n)$ is T-invariant too. Then, $[T]_\alpha^\alpha$ is upper-triangular as well.

Theorem 12.10: Schur
If V is a finite-dimensional complex inner product space and $T \in \mathcal{L}(V)$, then there exists an orthonormal basis of V such that $[T]_\alpha^\alpha$ is upper triangular.

Proof: This is a direct consequence of the previous few theorems, and the upper triangular condition covered in lecture 9.

12.2 Linear Functionals

Definition 12.2: Linear Functional
A linear functional on V is a map from V to $\mathbb{F}$.

Definition 12.3: Dual Space
The dual space $V^* = \mathcal{L}(V, \mathbb{F})$ is the set of all linear functionals on V.

Theorem 12.11: Riesz-Representation Theorem
Let V be a finite-dimensional inner-product space and V^* its dual space. If $\phi \in V^*$, then these exists a unique $u \in V$ such that $\phi(v) = \langle u, v \rangle$ for all $v \in V$.

Thus, if we let ϕ_u be the functional given by $\phi_u(v) = \langle v, u \rangle$ for all v, then $u \mapsto \phi_u$ defines an anti-isomorphism, i.e., an isomorphism from V onto the dual space V^* of V.

Definition 12.4: Dual Basis
If V is a finite-dimensional vector space and $v_1, \dots, v_n$ is a basis of V, then V^* has a so-called dual basis to $v_1, \dots, v_n$.

Proof: Define $\delta_1(c_1 v_1 + \cdots + c_n v_n) = c_1, \dots, \delta_n(c_1 v_1 + \cdots + c_n v_n) = c_n$ for $\delta_1, \dots, \delta_n \in V^*$.

- Let's check that $\delta_1, \dots, \delta_n$ is linearly independent: If $c_1\delta_1 + \cdots + c_n\delta_n = 0$, then $(c_1\delta_1 + \cdots + c_n\delta_n)(v) = 0$ for all $v \in V$. However, letting $v = v_j$ gives us $c_j = 0$, implying that the aforementioned functionals are indeed linearly independent.

- Let's check that $\delta_1, \ldots, \delta_n$ spans V^*. Let $v = c_1 v_1 + \cdots + c_n v_n$ and $\phi \in V^*$. Then,

$$\begin{aligned}\phi(v) &= \phi(c_1 v_1 + \cdots + c_n v_n) \\ &= c_1 \phi(v_1) + \cdots + c_n \phi(v_n) \\ &= \phi(v_1)\delta_1(c_1 v_1) + \cdots + \phi(v_n)\delta_n(c_n v_n) \\ &= \phi(v_1)\delta_1(v) + \cdots + \phi(v_n)\delta_n(v) \\ &= (\phi(v_1)\delta_1 + \cdots + \phi(v_n)\delta_n)(v)\end{aligned}$$

for all $v \in V$. So, $\phi = \phi(v_1)\delta_1 + \cdots + \phi(v_n)\delta_n$.

Theorem 12.12
Since V is finite dimensional, then so is V^*. As they have the same dimension, $V \cong V^*$. Similarly, $V^* \cong V^{**}$, which implies that $V \cong V^{**}$.

Let's construct this "natural" (or "canonical") isomorphism.

Proof: Let $v \in V$. Define $\hat{v} \in V^{**}$ such that $\hat{v}(\phi) = \phi(v) \in \mathbb{F}$ for all $\phi \in V^*$. Thus, $\hat{v}(\phi + \phi) = (\phi + \phi)(v) = \phi(v) + \phi(v) = \hat{v}(\phi) + \hat{v}(\phi)$ and $\hat{v}(c\phi) = (c\phi)(v) = c\phi(v) = c\hat{v}(\phi)$.

- Next, we check that the map $v \mapsto \hat{v}$ is linear:

$$\begin{aligned}(v \hat{+} w)(\phi) &= \phi(v + w) \\ &= \phi(v) + \phi(w) \\ &= \hat{v}(\phi) + \hat{w}(\phi)\end{aligned}$$

for all ϕ. Also,

$$(\hat{cv})(\phi) = \phi(cv) = c\phi(v) = c\hat{v}(\phi)$$

Since $\hat{v}$ satisfies both additivity and homogeneity, the map $\hat{} : V \mapsto V^{**}$ is indeed linear.

- Now we will show that $\hat{}$ is injective. Suppose $\hat{v}(\phi) = 0$ for all ϕ. Let $v = c_1 v_1 + \cdots + c_n v_n$. Look at $\hat{v}(\delta_j) = \delta_j(v)$ where each δ_j is defined in the same way as earlier. Therefore, $\hat{v}(\delta_j) = 0 \implies \delta_j(v) = 0 \implies c_j = 0$. Thus, $\hat{v}(\phi) = 0$ for all ϕ implies that $c_1 = \cdots = c_n = 0$ so $v = 0$. Then, $\ker(\hat{v}) = \{0\}$ and $\hat{v} : V \mapsto V^{**}$ is injective.

Based on the grounds of dimensionality, $\hat{v}$ is an isomorphism.

Example 12.2: 6B Exercise 1
For some $\theta \in \mathbb{R}$, show that $(\cos\theta, \sin\theta), (-\sin\theta, \cos\theta)$ and $(\cos\theta, \sin\theta), (\sin\theta, -\cos\theta)$ are orthonormal bases of $\mathbb{R}^2$. Furthermore, show that any orthonormal basis of $\mathbb{R}^2$ has one of these two forms.

- The first part is trivial since the taking the inner product of both vectors for both sets of bases would yield $\cos\theta\sin\theta - \cos\theta\sin\theta = 0$. Moreover, for each vector given above, their norm is $\cos^2\theta + \sin^2\theta = 1$. Thus, they are indeed orthonormal bases.

- For the second part, let $(a, b) \in \mathbb{R}^2$ be a normal vector. Thus, $a^2 + b^2 = 1$, which indicates that it lies on the unit circle. Then,

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

for some unique $0 \leq \theta \leq 2\pi$. Now suppose $(x, y) \in \mathbb{R}^2$ such that it too is a normal vector, but

$$\left\langle \begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} x \\ y \end{bmatrix} \right\rangle = ax + by = 0$$

Thus, (x, y) must also lie on the unit circle and based on the representation of (a, b) shown earlier, it only has the two possibilities given in the first part of the question.

Note 12.4

The matrices representing the two bases above are special:

- $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ represents a rotation in $\mathbb{R}^2$ about the origin, counterclockwise by θ radians.
- $\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$ represents a reflection in $\mathbb{R}^2$ about the line that has an angle of $\frac{\theta}{2}$ with the x-axis.

Thus, the example above essentially tells us that all orthonormal transformations on $\mathbb{R}^2$ are either rotations or reflections. This goes into Lie Group theory, but you can define the orthonormal group $O(2, \mathbb{R})$ to be the set of all 2×2 orthogonal matrices and the special orthonormal group $SO(2, \mathbb{R})$ to be the set of all rotations on $\mathbb{R}^2$ (i.e., all matrices in $O(2, \mathbb{R})$ with determinant 1). This can be generalized to higher dimensions as well. This seemingly abstract concept actually has multiple applications in fields like computer graphics and robotics!

13 Lecture 13

13.1 Review Problems

We went over a practice midterm and a past midterm in preparation for midterm 1.

14 Lecture 14

14.1 Orthogonal Complements

Definition 14.1: Orthogonal Complements
Let U be a subset of the inner product space V. The, $U^\perp = \{v \in V \mid \langle v, u \rangle = 0, \forall u \in U\}$ is the orthogonal complement of U.

Properties of the orthogonal complement:

1. $U^\perp$ is a subspace of V.

   Proof: By linearity of inner product in the first variable.

2. $\{0\}^\perp = V$

   Proof: $\{0\}^\perp = \{v \in V \mid \langle v, 0 \rangle = 0\} = V$ since $\langle v, 0 \rangle$ for all $v \in V$.

3. $\{V\}^\perp = \{0\}$

   Proof: Let $v \in V$. Then, $\langle v, v \rangle \neq 0$ if $v \neq 0$. Thus, if $v \neq 0$, then $v \notin V^\perp$. So, $V^\perp = \{0\}$ since $0 \in V^\perp$ following the definition of a subspace.

4. If U is a subset of V, then $U \cap U^\perp = \{0\}$

   Proof: If $u \in U$, then $\langle u, u \rangle = 0$ iff $u = 0$. So, if $u \in U$ and $u \neq 0$, then $u \neq U^\perp$.

5. If U and W are subsets of V, then $U \subseteq W \implies W^\perp \subseteq U^\perp$

   Proof: If $v \in W^\perp$, then $\langle v, w \rangle = 0$ for all $w \in W$. However, if $u \in U$, then $u \in W$. So, if $v \in W^\perp$, then $\langle v, u \rangle = 0$. So, $v \in U^\perp$ and $W^\perp \subseteq U^\perp$.

Theorem 14.1
Suppose that U is a finite-dimensional subspace of V (where V itself doesn't need to be finite-dimensional). Then, $V = U \oplus U^\perp$.

Proof: By property 4 given above, $U \cap U^\perp = \{0\}$. We will show that $V = U + U^\perp$. Choose any orthonormal basis $e_1, \ldots, e_m$ of U. Write

$$v = \underbrace{\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m}_{u} + \underbrace{(v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m)}_{w}$$

Then, $u \in U$ since $e_1, \ldots, e_m$ span U. Thus, we only need to show that $w \in U^\perp$. Look at $\left\langle w, e_j \right\rangle$:

$$\begin{aligned}
\left\langle w, e_j \right\rangle &= \left\langle v - u, e_j \right\rangle \\
&= \left\langle v, e_j \right\rangle - \left\langle u, e_j \right\rangle \\
&= \left\langle v, e_j \right\rangle - \left\langle v, e_j \right\rangle
\end{aligned}$$

since $e_1, \ldots, e_m$ are orthonormal anyway. This holds for any vector in $\operatorname{span}(e_1, \ldots, e_m)$ so $w \in U^\perp$ as intended.

Note 14.1
Note that the above theorem can fail if U is not finite-dimensional.

Theorem 14.2
If U is a subset of V, then $U \subseteq (U^\perp)^\perp$.

Proof: If $w \in U^\perp$ and $u \in U$, then $\langle u, w\rangle = 0$. However, $\langle w, u\rangle = \overline{\langle u, w\rangle} = 0$ as well. Since $w \in U^\perp$ is arbitrary, $u \in (U^\perp)^\perp$. Thus, $U \subseteq (U^\perp)^\perp$ as intended.

Theorem 14.3
If V is finite-dimensional and U is a subspace of V, then $U = (U^\perp)^\perp$.

Proof: If V is finite-dimensional, then $V = U \oplus U^\perp$ and $V = U^\perp \oplus (U^\perp)^\perp$. Thus, $\dim U = \dim V - \dim U^\perp = \dim(U^\perp)^\perp$. Since $U \subseteq (U^\perp)^\perp$ was just proven, we get that $U = (U^\perp)^\perp$.

Note 14.2
Note that $(U^\perp)^\perp \subseteq U$ can actually fail if V is infinite-dimensional, which is why the finite-dimensional constraint is added here separately.

Theorem 14.4
Let U be a finite-dimensional subspace of V. Then, $V = U \oplus U^\perp$. So, if $v \in V$, then $v = u + w$ for a unique $u \in U$ and a unique $w \in U^\perp$. Define $\operatorname{proj}_U(v) = u = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m$ where $e_1, \ldots, e_m$ is *any* orthonormal basis of U.

Note 14.3
Note that this theorem just restates what we already took for granted in our explanation of the Gram-Schmidt process earlier.

Following the definition of an orthogonal projection, we can now list some its basic properties:

1. $P = \operatorname{proj}_U : V \mapsto U \subseteq V$ is linear
2. $\operatorname{range}(P) = U$ and $\ker(P) = U^\perp$
3. $P^2 = P$
4. $\|\operatorname{proj}_U(v)\| \le \|v\|$ for all $v \in V$

   Proof: $\|v\|^2 = \|\operatorname{proj}_U(v)\|^2 + \|v - \operatorname{proj}_U(v)\|^2 \ge \|\operatorname{proj}_U(v)\|^2$ by the Pythagorean theorem

The proof of the remaining properties should be fairly easy and is, therefore, left as an exercise for the reader.

14.2 Least Squares/Minimization

Theorem 14.5
Let U be a finite dimensional subspace of V and $u \in U, v \in V$. Then, $\|v - \operatorname{proj}_U(v)\|^2 \le \|v - u\|^2$ for all u with equality if $u = \operatorname{proj}_U(v)$.

Proof: Observe that

$$\begin{aligned} \|v - u\|^2 &= \|v - \operatorname{proj}_U(v) + \operatorname{proj}_U(v) - u\|^2 \\ &= \|v - \operatorname{proj}_U(v)\|^2 + \|\operatorname{proj}_U(v) - u\|^2 \end{aligned}$$

by the Pythagorean theorem. So, $\|v - u\|^2 \ge \|v - \operatorname{proj}_U(v)\|^2$ with equality if $u = \operatorname{proj}_U(v)$.

Let A be an $m \times n$ real valued matrix and $b \in \mathbb{R}^n$. Then, the linear system $Ax = b$ may have no solutions or it can even be inconsistent. However, we can always find a "least squares solution" that minimizes $\|Ax - b\|_2^2$. First, lets prove a few claims:

Theorem 14.6
(row space of $A)^{\perp} = \ker(A)$

Proof: Let $r_1, \ldots, r_m$ be the rows of A. Then, $Ax = 0$ iff $r_i \cdot x = 0$ for all $i \leq m$. So, $\text{row}(A) = \ker(A)^{\perp} \iff \text{row}(A)^{\perp} = \ker(A)$.

Theorem 14.7
$\text{im}(A^T) = \text{col}(A^T) = \text{row}(A) = \ker(A)^{\perp}$.

Theorem 14.8
Let $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ be represented by $T(x) = Ax$ where A is an $m \times n$ matrix and $\mathbb{R}^n, \mathbb{R}^m$ are Euclidean inner product spaces. Then, $\ker(A) = \ker(A^T A)$.

Proof: The direction $\ker(A) \subseteq \ker(A^T A)$ is trivially clear since $Ax = 0 \implies A^T Ax = A^T 0 = 0$. Now, suppose that $A^T Ax = 0$. Then, $Ax \in \ker(A^T)$ and $Ax \in \text{im}(A) = \ker(A^T)^{\perp}$. However, $\mathbb{R}^m = \ker(A^T) \oplus \ker(A^T)^{\perp}$ so $Ax = 0 \implies x \in \ker(A)$.

Suppose that v minimizes the least squares error, i.e., $\|Av - b\|_2 \leq \|Ax - b\|_2$ for all $x \in V$. Then $Av = \text{proj}_{\text{im}(A)} b \iff b - Av \in \text{im}(A)^{\perp} = \ker(A^T) = \text{row}(A) \implies A^T(b - Av) = 0 \implies A^T Av = A^T b$.

In conclusion, a "least squares" solution to $Ax = b$ exists and is an exact (i.e. not approximate) solution to the "normal" equation given by $A^T Av = A^T b$.

Example 14.1: 6A Exercise 6
Suppose $u, v \in V$. Prove that $\langle u, v\rangle = 0$ iff $\|u\| \leq \|u + av\|$ for all $a \in \mathbb{F}$.

Proof: Note that

$$\begin{aligned}
\|u + av\|^2 &= \langle u + av, u + av\rangle \\
&= \|u\|^2 + a\langle v, u\rangle + \overline{a}\langle u, v\rangle + |a|^2\|v\|^2
\end{aligned}$$

If $\langle u, v\rangle = \langle v, u\rangle = 0$, then $\|u + av\|^2 = \|u\|^2 + |a|^2\|v\|^2 \geq \|u\|^2$. Conversely, suppose that $\|u\|^2 \leq \|u + av\|^2$ for all $a \in \mathbb{F}$. Then,

$$\begin{aligned}
\|u + av\|^2 - \|u\|^2 &= |a|^2\|v\|^2 + a\langle v, u\rangle + \overline{a}\langle u, v\rangle \\
&\geq 0
\end{aligned}$$

If $v = 0$, then $\langle u, v\rangle = 0$ as desired. However, if $v \neq 0$, then let $a = \frac{-\langle u, v\rangle}{\|v\|^2}$ to get

$$\frac{|-\langle u, v\rangle|^2}{\|v\|^2} - 2\frac{|\langle u, v\rangle|^2}{\|v\|^2} = -\frac{|\langle u, v\rangle|^2}{\|v\|^2} \geq 0 \implies \langle u, v\rangle = 0$$

Example 14.2: 6C Exercise 8
Suppose that V is finite-dimensional and $P \in \mathcal{L}(V)$ such that $P^2 = P$ and $\|Pv\| \leq \|v\|$ for all $v \in V$. Prove that there is a subspace $U \subseteq V$ such that $P = P_U$.

Proof: Let $U = \text{range}(P)$. We claim that $P = P_U$. Now, $V = U \oplus U^{\perp} = \text{range}(P_U) \oplus \text{range}(P_U)^{\perp}$. If $u \in \text{range}(P)$, i.e., $u = Px$ for some $x \in V$, then $Pu = P^2 x = Px = u$ and $P_U u = u$. So, P and P_U agree on $U = \text{range}(P)$ and we need to only prove that $\ker(P) = \ker(P_U) = \text{range}(P_U)^{\perp} = \text{range}(P)^{\perp}$. Towards this, let $u \in \text{range}(P)$ and $w \in \ker(P)$. Then, $\|u\| = \|P(u + aw)\| \leq \|u + aw\|^2$ by the preceding example. Thus, $\langle u, w\rangle = 0$ and $\ker(P) \subseteq \text{range}(P)^{\perp}$. However, as $\dim\ker(P) = \dim\text{range}(P)^{\perp}$, we get that $\ker(P) = \text{range}(P)^{\perp}$. So, if $v = u + w$, then $Pv = Pu = P_U u = P_U v$.

Example 14.3

Let $p \in \mathcal{P}_3(\mathbb{R})$ and $p(0) = 0, p'(0) = 0$. Our goal is to minimize $\int_0^1 (2 + 3x - p(x))^2 \,\mathrm{d}x$.

Let $p = a + bx + cx^2 + dx^3$ and $p' = b + 2cx + 3dx^2$. Then, $p(0) = 0 \implies a = 0$ and $p'(0) = 0 \implies b = 0$. Equivalently, $p = cx^2 + dx^3$, and the basis of this subspace is x^2, x^3. However, we need an orthonormal bases for $U = \operatorname{span}(x^2, x^3)$.

Let the L^2-inner product on this inner product space be $\langle p, q \rangle = \int_0^1 p(x)q(x)\,\mathrm{d}x$. Performing Gram-Schmidt,

$$\begin{aligned}
\left\|x^2\right\| &= \left(\int_0^1 x^4 \,\mathrm{d}x\right)^{\frac{1}{2}} \\
&= \left(\frac{1}{5}\right)^{\frac{1}{2}} \\
&= \frac{1}{\sqrt{5}} \\
e_1 &= \frac{x^2}{\left\|x^2\right\|} \\
&= \sqrt{5}x^2
\end{aligned}$$

Similarly,

$$e_2 = \frac{x^3 - \left\langle x^3, \sqrt{x^2}\right\rangle \sqrt{5}x^2}{\left\|x^3 - \left\langle x^3, \sqrt{x^2}\right\rangle \sqrt{5}x^2\right\|}$$

Since,

$$\begin{aligned}
\left\langle x^3, \sqrt{5}x^2\right\rangle &= \int_0^1 \sqrt{5}x^5 \,\mathrm{d}x \\
&= \frac{\sqrt{5}}{6} \\
e_2 &= \frac{x^3 - \frac{5}{6}x^2}{\left\|x^3 - \frac{5}{6}x^2\right\|} \\
\left\|x^3 - \frac{5}{6}x^2\right\| &= \left(\int_0^1 \left(x^3 - \frac{5}{6}x^2\right)\left(x^3 - \frac{5}{6}x^2\right) \mathrm{d}x\right)^{\frac{1}{2}} \\
&= \left(\int_0^1 \left(x^6 - \frac{10}{6}x^4 + \frac{25}{36}x^4\right) \mathrm{d}x\right) \\
&= \frac{1}{6\sqrt{7}} \\
e_2 &= 6\sqrt{7}x^3 - 5\sqrt{7}x^2
\end{aligned}$$

Note that $\operatorname{proj}_U(2 + 3x)$ minimizes the least-squares error. Thus,

$$\begin{aligned}
\operatorname{proj}_U(2 + 3x) &= \langle 2 + 3x, e_1\rangle e_1 + \langle 2 + 3x, e_2\rangle e_2 \\
&= \left(\int_0^1 (2 + 3x)(\sqrt{5}x^2)\,\mathrm{d}x\right)(\sqrt{5}x^2) + \left(\int_0^1 (2 + 3x)(6\sqrt{7}x^3 - 5\sqrt{7}x^2)\,\mathrm{d}x\right)(6\sqrt{7}x^3 - 5\sqrt{7}x^2) \\
&= -\frac{203}{10}x^3 + 24x^2
\end{aligned}$$

15 Lecture 15

15.1 Adjoint Maps

Definition 15.1: Adjoint

Let $T \in \mathcal{L}(V, W)$. The adjoint of T is the map $T^* : W \mapsto V$ such that $\langle Tv, w\rangle = \langle v, T^*w\rangle$ for all $v \in V$ and $w \in W$.

Proof: We will show that T^* actually exists and is unique. Let $w \in W$ and define $\phi_w(v) = \langle Tv, w\rangle$. Then, $\phi_w \in \mathcal{L}(V, \mathbb{F}) = V^*$ and the Riesz Representation Theorem applies to ϕ_w. Thus, there exists a unique vector in V, call it T^*w, such that $\phi_w(v) = \langle v, T^*w\rangle$ for all $v \in V$. So, $\langle Tv, w\rangle = \langle v, T^*w\rangle$. Therefore, T^* is a well-defined map from W to V and its uniqueness of T^*w guaranteed by the Riesz Representation Theorem.

Theorem 15.1

T^* is linear, i.e., $T^* \in \mathcal{L}(W, V)$.

Proof: Pretty straightforward using the properties of inner products. It is left as an exercise for the reader to verify this.

Example 15.1

Let $T : \mathbb{R}^3 \mapsto \mathbb{R}^2$ be given by $T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1)$. Find a formula for T^*. Let $(y_1, y_2) \in \mathbb{R}^2$. Then,

$$\begin{aligned}
\langle T(x_1, x_2, x_3), (y_1, y_2)\rangle &= \langle (x_1, x_2, x_3), T^*(y_1, y_2)\rangle \\
\langle (x_2 + 3x_3, 2x_1), (y_1, y_2)\rangle &= \langle (x_1, x_2, x_3), T^*(y_1, y_2)\rangle \\
x_2y_1 + 3x_3y_1 + 2x_1y_2 &= \langle (x_1, x_2, x_3), T^*(y_1, y_2)\rangle \\
\langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1)\rangle &= \langle (x_1, x_2, x_3), T^*(y_1, y_2)\rangle
\end{aligned}$$

Thus, $T^*(y_1, y_2) = (2y_2, y_1, 3y_1)$ works for all $(x_1, x_2, x_3) \in \mathbb{R}^3$ and $(y_1, y_2) \in \mathbb{R}^2$.

Theorem 15.2: Properties of the Adjoint

1. $(S + T)^* = S^* + T^*$ for all $S, T \in \mathcal{L}(V, W)$
2. $(\lambda T)^* = \overline{\lambda} T^*$ for all $\lambda \in \mathbb{F}$ and $T \in \mathcal{L}(V, W)$
3. $T^{**} = (T^*)^* = T$ for all $T \in \mathcal{L}(V, W)$
4. $I^* = I$ for the identity operator on V
5. $(ST)^* = T^*S^*$ for all $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$

Proof: Most of these statements are fairly basic but we will prove properties 3 and 5 here.

- Proof of property 3: note that

$$\langle Tv, w\rangle = \langle v, T^*w\rangle = \overline{\langle T^*w, v\rangle} = \overline{\langle w, (T^*)^*v\rangle} = \langle (T^*)^*v, w\rangle$$

  This holds for all $v \in V, w \in W$ so $T = (T^*)^*$. The uniqueness part is implicit from the Riesz Representation Theorem.

- Proof of property 5: note that

$$\begin{aligned}
\langle STv, u\rangle &= \langle v, (ST)^*u\rangle \\
\langle STv, u\rangle &= \langle Tv, S^*u\rangle \\
&= \langle v, T^*S^*u\rangle
\end{aligned}$$

  Again, since this statement holds for all $v \in V, u \in U$, we get that $(ST)^* = T^*S^*$.

Theorem 15.3
If $T \in \mathcal{L}(V, W)$, then,

1. $\ker(T^*) = \operatorname{range}(T)^\perp$
2. $\operatorname{range}(T^*) = \ker(T)^\perp$
3. $\ker(T) = \operatorname{range}(T^*)^\perp$
4. $\operatorname{range}(T) = \ker(T^*)^\perp$

Proof: The proofs of all 4 statements:

1. Suppose $T^*w = 0$. Then, $\langle v, T^*w\rangle = 0 \implies \langle Tv, w\rangle = \overline{\langle w, Tv\rangle} = 0 \implies \langle w, Tv\rangle = 0$. Since v is arbitrary, we conclude that $w \in \operatorname{range}(T)^\perp \implies \ker(T^*) \subseteq \operatorname{range}(T)^\perp$. This argument can be reversed to obtain $\operatorname{range}(T)^\perp \subseteq \operatorname{range}(T^*)$ and together, $\ker(T^*) = \operatorname{range}(T)^\perp$.
2. Following proof 1, $\ker(T^*) = \operatorname{range}(T)^\perp \implies \operatorname{range}(T) = \ker(T^*)^\perp \implies \operatorname{range}(T^*) = \ker((T^*)^*)^\perp = \ker(T)^\perp$.
3. Replace T^* with T in proof 1.
4. Shown as an intermediate step in proof 2.

Definition 15.2: Conjugate Transpose
Let $A \in \mathcal{M}_{m,n}(\mathbb{F})$. The conjugate transpose of A (usually denoted by A^*) is the $n \times m$ matrix B such that $B_{i,j} = \overline{A_{j,i}}$.

Example 15.2
If $A = \begin{bmatrix} 2+i & 1-i \\ -2 & 3 \end{bmatrix}$, then $B = \begin{bmatrix} 2-i & -2 \\ 1+i & 3 \end{bmatrix}$

Theorem 15.4
If $T \in \mathcal{L}(V, W)$ and V has an orthonormal basis $e_1, \dots, e_n$ while W has an orthonormal basis $f_1, \dots, f_m$, then $\mathcal{M}(T^*, (f_1, \dots, f_m), (e_1, \dots, e_n))$ is the conjugate transpose of $\mathcal{M}(T, (e_1, \dots, e_n), (f_1, \dots, f_m))$.

Proof: Note that $T^* f_j = \langle T^* f_j, e_1\rangle e_1 + \cdots + \langle T^* f_j, e_n\rangle e_n$. So, the jth column of $\mathcal{M}(T^*)$ is

$$\begin{bmatrix} \langle T^* f_j, e_1\rangle \\ \vdots \\ \langle T^* f_j, e_n\rangle \end{bmatrix}$$

Similarly, since $Te_j = \langle Te_j, f_1\rangle f_1 + \cdots + \langle Te_j, f_m\rangle f_m$, the jth column of $\mathcal{M}(T)$ is

$$\begin{bmatrix} \langle Te_j, f_1\rangle \\ \vdots \\ \langle Te_j, f_m\rangle \end{bmatrix}$$

The (i, j)th entry of $\mathcal{M}(T^*)$ is $\langle T^* f_j, e_i\rangle$ and the (j, i)th entry of $\mathcal{M}(T)$ is $\langle Te_i, f_j\rangle$. Since $\overline{\langle Te_i, f_j\rangle} = \overline{\langle e_i, T^* f_j\rangle} = \langle T^* f_j, e_i\rangle$ for any two sets of bases, $\mathcal{M}(T)$ and $\mathcal{M}(T^*)$ are indeed conjugate transposes of each other.

Definition 15.3: Self-Adjoint
If V is a finite-dimensional inner product space and $T \in \mathcal{L}(V)$, then T is self-adjoint iff $T = T^*$, i.e., $\langle Tv, w\rangle = \langle v, Tw\rangle$ for all $v, w \in V$.

Note 15.1
If T is self-adjoint and $\beta = e_1, \ldots, e_n$ is an orthonormal basis of V, then $[T^*]_\beta^\beta = ([T]_\beta^\beta)^*$.

Theorem 15.5
If T is self-adjoint, then all of its eigenvalues are real. This holds for T defined over both real and complex vector spaces.

Proof: Let v be a λ-eigenvector of T and $v \neq 0$ (by definition). Then,

$$\begin{aligned} Tv &= \lambda v \\ \langle Tv, v \rangle &= \langle \lambda v, v \rangle \\ &= \lambda \|v\|^2 \\ \langle v, T^* v \rangle &= \langle v, Tv \rangle \\ &= \langle v, \lambda v \rangle \\ &= \overline{\lambda}\,\|v\|^2 \end{aligned}$$

Since $v \neq 0$, we get that $\overline{\lambda} = \lambda$ which implies that λ is real.

Theorem 15.6
Let V is a complex inner product space and $T \in \mathcal{L}(V)$. If $\langle Tv, v \rangle = 0$ for all $v \in V$, then $T = 0$.

Proof: Note that

$$\langle Tu, w \rangle = \frac{\langle T(u+w), u+w \rangle - \langle T(u-w), u-w \rangle}{4} + \frac{\langle T(u+iw), u+iw \rangle - \langle T(u-iw), u-iw \rangle}{4} i$$

Since $\langle Tv, v \rangle = 0$, it follows that $\langle Tu, w \rangle = 0$ for all $u, w \in V$. Thus, $\langle Tu, Tu \rangle = 0 \implies Tu = 0 \implies T = 0$ since this holds for any arbitrary u.

Note 15.2
If V is a real inner product space instead, this result can fail. Consider $V = \mathbb{R}^2$ with the normal dot product as its inner product. Then, the counterclockwise rotation by $\frac{\pi}{2}$ will not work.

Theorem 15.7
If V is a complex inner product space and $T \in \mathcal{L}(V)$, then $T = T^*$ iff $\langle Tv, v \rangle \in \mathbb{R}$ for all $v \in V$.

Proof: For some $v \in V$, observe that $\langle Tv, v \rangle - \overline{\langle Tv, v \rangle} = \langle Tv, v \rangle - \langle v, Tv \rangle = \langle Tv, v \rangle - \langle T^* v, v \rangle = \langle (T - T^*)v, v \rangle$. If $\langle Tv, v \rangle \in \mathbb{R}$ for all $v \in V$, then $\langle Tv, v \rangle - \overline{\langle Tv, v \rangle} = 0 = \langle (T - T^*)v, v \rangle$ for all $v \in V$. By the previous theorem, $T - T^* = 0 \implies T = T^*$.
Conversely, $T = T^* \implies \langle Tv, v \rangle = \langle v, T^* v \rangle = \langle v, Tv \rangle = \overline{\langle Tv, v \rangle}$. So, $\langle Tv, v \rangle$ is real.

Note 15.3
This result will fail for real inner product spaces too. As a counterexample again, consider the counterclockwise rotation by $\frac{\pi}{2}$. In fact, you can consider any real operator that is not self-adjoint.

Theorem 15.8
If V is an inner product space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ and $T = T^*$, then $\langle Tv, v \rangle = 0$ for all $v \in V$ implies that $T = 0$.

Proof: Essentially the same proof as that of theorem 15.6.

Note 15.4
This result just confirms the two counterexamples above since there is no non-zero operator on a real product space that satisfies $\langle Tv, v\rangle = 0$ for all $v \in V$ *and* is self-adjoint.

Definition 15.4: Normal
$T \in \mathcal{L}(V)$ is normal iff $T^*T = TT^*$.

Example 15.3
Let $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be such that $T(c_1e_1 + c_2e_2) = c_2e_1$. Then, $T^* : \mathbb{R}^2 \mapsto \mathbb{R}^2$ is given by $T^*(c_1e_1 + c_2e_2) = c_1e_2$. Note that if $e = e_1, e_2$ is the standard basis, then

$$[T]_e^e = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$
$$[T^*]_e^e = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Since $T^*T(e_1) = 0 \neq e_1 = TT^*(e_1)$, this linear map is not normal.

Example 15.4
Let $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be such that $[T]_e^e = \begin{bmatrix} 2 & -3 \\ 3 & 2 \end{bmatrix}$. Then, T is normal but not self-adjoint.

Note 15.5
Following the example above, all self-adjoint operators are normal but not all normal operators are self-adjoint.

Theorem 15.9
T is normal iff $\|Tv\| = \|T^*v\|$ for all $v \in V$.

Proof: If T is normal, then

$$\begin{aligned} TT^* = T^*T &\iff TT^* - T^*T = 0 \\ &\iff \langle (TT^* - T^*T)v, v\rangle = 0\,\forall v \in V \\ &\iff \langle TT^*v, v\rangle = \langle T^*Tv, v\rangle\,\forall v \in V \\ &\iff \langle T^*v, T^*v\rangle = \langle Tv, Tv\rangle\,\forall v \in V \\ &\iff \|T^*v\|^2 = \|Tv\|^2\,\forall v \in V \end{aligned}$$

The second line follows since $TT^* - T^*T$ is self-adjoint.

Theorem 15.10
Let T be normal. If $v \in V$ is a λ-eigenvector of T, then v is a $\overline{\lambda}$-eigenvector of T^*.

Proof: If T is normal, then $T - \lambda I$ is normal. Then, the theorem above implies $\|(T - \lambda I)v\| = 0 \implies \|(T - \lambda I)^*v\| = 0 \implies \|(T^* - \overline{\lambda}\,I)v\| = 0 \implies (T^* - \overline{\lambda}\,I) = 0$. Thus, v is a $\overline{\lambda}$-eigenvector of T^*.

Theorem 15.11
If T is normal, then the eigenvectors of T corresponding to distinct eigenvalues are orthogonal.

Proof: Suppose that $Tv_1 = \lambda_1 v_1$ and $Tv_2 = \lambda_2 v_2$ for $v_1, v_2 \neq 0$ and $\lambda_1 \neq \lambda_2$. Then, T being normal implies that $T^* v_2 = \overline{\lambda_2}\, v_2$. Thus,

$$\begin{aligned} 0 &= \langle Tv_1, v_2\rangle - \langle v_1, T^* v_2\rangle \\ &= \langle \lambda_1 v_1, v_2\rangle - \langle v_1, \overline{\lambda_2}\, v_2\rangle \\ &= (\lambda_1 - \lambda_2)\langle v_1, v_2\rangle \end{aligned}$$

Since the eigenvalues are distinct, $\langle v_1, v_2\rangle = 0$, making the eigenvectors orthogonal. This can be generalized to vector spaces with dimensions greater than 2.

Example 15.5: 7A Exercise 14

Let T be normal, $\|v\| = \|w\| = 2$, $Tv = 3v$ and $Tw = 4w$. Prove that $\|T(v+w)\| = 10$.

Proof: We have $\langle v, w\rangle = 0$ since v and w are eigenvectors for distinct eigenvalues of the normal operator T. So,

$$\begin{aligned} \|T(v+w)\|^2 &= \langle T(v+w), T(v+w)\rangle \\ &= \|Tv\|^2 + \langle Tw, Tv\rangle + \langle Tv, Tw\rangle + \|Tw\|^2 \\ &= \|3v\|^2 + \langle 4w, 3v\rangle + \langle 3v, 4w\rangle + \|4w\|^2 \\ &= |3|^2\|v\|^2 + 12\langle w, v\rangle + 12\langle v, w\rangle + |4|^2\langle w\rangle^2 \\ &= 9 * 4 + 12 * 0 + 12 * 0 + 16 * 4 \\ &= 100 \end{aligned}$$

Thus, $\|T(v+w)\| = 10$ as needed.

Example 15.6: 7A Exercise 19

Let $T \in \mathcal{L}(\mathbb{C}^3)$ be normal and $T(1,1,1) = (2,2,2)$. Suppose $(z_1, z_2, z_3) \in \ker(T)$. Prove that $z_1 + z_2 + z_3 = 0$.

Proof: If T is normal, then $\ker(T^*) = \ker(T)$ since $\|Tv\| = 0 \implies Tv = 0$ and $\|Tv\| = \|T^*v\| \implies T^*v = 0$. Then, $(z_1, z_2, z_3) \in \ker(T^*) = \operatorname{range}(T)^\perp$. However, $(1,1,1) \in \operatorname{range}(T)$ since $T(0.5, 0.5, 0.5) = (1,1,1)$. Then, $(z_1, z_2, z_3) \cdot (1,1,1) = 0 \implies z_1 + z_2 + z_3 = 0$.

16 Lecture 16

16.1 Spectral Theorem

Theorem 16.1: Complex Spectral Theorem
If V is a finite-dimensional complex inner product space and $T \in \mathcal{L}(V)$, then the following are equivalent

1. T is normal
2. V has an orthonormal basis consisting of the eigenvalues of T
3. T has a diagonal matrix with respect to some orthonormal basis of V

Proof: We will prove directions $1 \implies 3$, $3 \implies 1$ and $2 \iff 3$.

- statement $3 \implies$ statement 1

  Suppose $\beta = e_1, \dots, e_n$ is an orthonormal basis of V and

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

  is diagonal. Since β is orthonormal,

$$[T^*]_\beta^\beta = ([T]_\beta^\beta)^* = \begin{bmatrix} \overline{\lambda_1} & & \\ & \ddots & \\ & & \overline{\lambda_n} \end{bmatrix}$$

  So,

$$[T^*T]_\beta^\beta = [T^*]_\beta^\beta [T]_\beta^\beta = \begin{bmatrix} |\lambda_1|^2 & & \\ & \ddots & \\ & & |\lambda_n|^2 \end{bmatrix} = [T]_\beta^\beta [T^*]_\beta^\beta = [TT^*]_\beta^\beta$$

  which implies that $T^*T = TT^*$, i.e., T is normal.

- statement $1 \implies$ statement 3

  Suppose that T is normal. Since V is complex, by Schur's Theorem, there is an orthonormal basis $\beta = e_1, \dots, e_n$ of V with respect to which T has an upper triangular matrix. Then,

$$[T]_\beta^\beta = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ & a_{22} & \dots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{bmatrix}$$

  We show that $[T]_\beta^\beta$ is actually diagonal as well. From the matrix above, note that

$$T(e_1) = a_1 e_1$$
$$\|T(e_1)\|^2 = |a_{11}|^2$$

  By the Pythagorean theorem,

$$T^*(e_1) = \overline{a_{11}}\, e_1 + \overline{a_{12}}\, e_2 + \dots + \overline{a_{1n}}\, e_n$$
$$\|T^*(e_1)\|^2 = |a_{11}|^2 + |a_{12}|^2 + \dots + |a_{1n}|^2$$

However, since T is normal, $\|T(e_1)\|^2 = \|T^*(e_1)\|^2 \implies |a_{11}|^2 = |a_{11}|^2 + |a_{12}|^2 + \cdots + |a_{1n}|^2 \implies a_{12} = \cdots = a_{1n} = 0$. So,

$$T(e_2) = 0e_1 + a_{22}e_2 = a_{22}e_2 \implies \|T(e_2)\|^2 = |a_{22}|^2$$

while

$$T^*(e_2) = \overline{a_{22}}\, e_2 + \overline{a_{23}}\, e_3 + \cdots + \overline{a_{2n}}\, e_n$$

Again, due to T being normal, $\|T(e_2)\|^2 = |a_{22}|^2 = |a_{22}|^2 + |a_{23}|^2 + \cdots + |a_{2n}|^2 = \|T^*(e_2)\|^2 \implies a_{23} = \cdots = a_{2n} = 0$. Continuing this process n times, note that all of the non-diagonal elements of $[T]_\beta^\beta$ will just reduce to a 0.

- statement 2 $\iff$ statement 3

  This follows immediately from the fact that $[T]_\beta^\beta$ is diagonal iff β is a basis of V consisting of the eigenvectors of T.

Now, we want to show that if V is a finite-dimensional real inner product space and $T \in \mathcal{L}(V)$ is self-adjoint, then T has a real eigenvalue, i.e., there exists some $\lambda \in \mathbb{R}$, $v \in V$, $v \neq 0$ with $Tv = \lambda v$. First, we will define complexification.

Definition 16.1: Complexification

Let V be a real vector space. We define $V_\mathbb{C}$ to be the complexification of V. That is, $V_\mathbb{C}$ will be a complex vector space.

The underlying set of $V_\mathbb{C}$ is $V \times V = \{(u, v) \mid u, v \in V\}$ and we write $(u, v) \in V_\mathbb{C}$ as $u + iv$. Then, addition in $V_\mathbb{C}$ can simply be defined as

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2)$$

for $u_1, u_2, v_1, v_2 \in V$ and complex scalar multiplication can be defined as

$$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$

for $a, b \in \mathbb{R}$ and $u, v \in V$. It is easy to check that

- $V_\mathbb{C}$ is a complex vector space
- If $v_1, \ldots, v_n$ is a basis of V (as a real vector space), then $v_1, \ldots, v_n$ is a basis of $V_\mathbb{C}$ (as a complex vector space)
- If $V = \mathbb{R}^n$, then $V_\mathbb{C} = \mathbb{C}^n$

If V is a real inner product space with inner product $\langle \cdot, \cdot \rangle$, then $V_\mathbb{C}$ is a complex inner product space with inner product

$$\langle u_1 + iv_1, u_2 + iv_2 \rangle = \langle u_1, u_2 \rangle + i\langle v_1, u_2 \rangle - i\langle u_1, v_2 \rangle + \langle v_1, v_2 \rangle$$

Note that

$$\begin{aligned} \langle u + iv, u + iv \rangle &= \langle u, u \rangle + i\langle v, u \rangle - i\langle u, v \rangle + \langle v, v \rangle \\ &= \|u\|^2 + \|v\|^2 \\ &\geq 0 \end{aligned}$$

as expected from the notion of a norm.

Definition 16.2: Operator Complexification

If $T \in \mathcal{L}(V)$ and V is a real vector space, then $T_\mathbb{C}(u + iv) = Tu + iTv$ where $T_\mathbb{C} \in \mathcal{L}(V_\mathbb{C})$.

Theorem 16.2

If T is self-adjoint for $\langle \cdot, \cdot \rangle$ on V, then $T_\mathbb{C}$ is self-adjoint for $\langle \cdot, \cdot \rangle$ on $V_\mathbb{C}$.

Proof: This proof is left as an exercise for the reader.

Theorem 16.3
Let V be a finite-dimensional real inner product space with $T \in \mathcal{L}(V)$ self-adjoint. Then, there is a $\lambda \in \mathbb{R}$ and $v \in V, v \neq 0$ such that $Tv = \lambda v$.

Proof: Since $V_{\mathbb{C}}$ is complex, $T_{\mathbb{C}}$ has a at least one eigenvalue $\lambda \in \mathbb{C}$ and associated eigenvector $w \in V_{\mathbb{C}}$. Moreover, $T_{\mathbb{C}}$ is self-adjoint so $\lambda \in \mathbb{R}$. If $w = u + iv$ for $u, v \in V$, then $T_{\mathbb{C}}(w) = \lambda w \implies Tu + iTv = \lambda u + i\lambda v \implies Tu = \lambda u$ and $Tv = \lambda v$. Therefore, at least one of u, v is nonzero since w is nonzero, so λ has a valid eigenvector associated with it.

Theorem 16.4
Let V be an inner product space and $T \in \mathcal{L}(V)$. If U is a T-invariant subspace of V, then $U^{\perp}$ is a T^*-invariant subspace of V.

Proof: Suppose $w \in U^{\perp}$ and $u \in U$. Then, $Tu \in U$ and $\langle Tu, w\rangle = 0 \implies \langle u, T^*w\rangle = 0$. So, $\langle T^*w, u\rangle = 0$. Since $u \in U$ was arbitrary, $T^*w \in U^{\perp}$. However, $w \in U^{\perp}$ was also arbitrary so $U^{\perp}$ is T^*-invariant.

Theorem 16.5
Let V be an inner product space with $T \in \mathcal{L}(V)$ self-adjoint. If U is a T-invariant subspace of V, then $U^{\perp}$ is also T-invariant. Also, $T|_U \in \mathcal{L}(U)$ is self-adjoint, as is $T|_{U^{\perp}} \in \mathcal{L}(U^{\perp})$.

Proof: $U^{\perp}$ being T-invariant immediately follows from the theorem above. If $u, w \in U$, then $\langle T|_U u, w\rangle = \langle Tu, w\rangle = \langle u, Tw\rangle = \langle u, T|_U w\rangle$. Thus, $T|_U$ is self-adjoint. We can show that $T|_{U^{\perp}}$ is self-adjoint using a very similar calculation.

Theorem 16.6: Real Spectral Theorem
Let V be a finite-dimensional real inner product space with $T \in \mathcal{L}(V)$. Then, T is self-adjoint iff T has a diagonal matrix (with real entries) with respect to some orthonormal basis of V.

Proof: We will prove both directions

- Let's proceed by induction for the forward direction. The base case $\dim V = 1$ is trivial.

  Suppose $\dim V = n$. Then, T being self-adjoint implies that there is a $\lambda_1 \in \mathbb{R}, v_1 \in V, v \neq 0$ with $Tv_1 = \lambda_1 v_1$. WLOG, let $\|v_1\| = 1$. Note that $T|_U$ and $T|_{U^{\perp}}$ are self-adjoint operators on spaces of dimension $\leq n$, following the previous theorem. So, the induction hypothesis applies here and $T|_{U^{\perp}}$ has a diagonal matrix with respect to some orthonormal basis α of $U^{\perp} = \operatorname{span}(v_1)^{\perp}$. Let β be a basis consisting of v_1 and the vectors in α. Then, β is orthonormal and $[T]_{\beta}^{\beta}$ has the form

$$[T]_{\beta}^{\beta} = \left[\begin{array}{c|c} \lambda_1 & 0 \quad 0 \quad 0 \\ \hline 0 & \\ \vdots & \left[T|_{\operatorname{span}(v_1)^{\perp}}\right]_{\alpha}^{\alpha} \\ 0 & \end{array}\right]$$

  Since the lower right block matrix above is also diagonal, the matrix representation of T is diagonal with respect to β.

- Suppose that $\beta = e_1, \dots, e_n$ is the orthonormal basis of V mentioned above and $[T]_{\beta}^{\beta}$ is diagonal with real entries. Then,

$$([T]_{\beta}^{\beta})^* = \begin{bmatrix} \overline{\lambda_1} & & \\ & \ddots & \\ & & \overline{\lambda_n} \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = [T]_{\beta}^{\beta}$$

  since all $\lambda_j \in \mathbb{R}$. Therefore, $[T]_{\beta}^{\beta} = [T^*]_{\beta}^{\beta} \implies T = T^*$, i.e., T is self-adjoint.

Example 16.1: 7B Exercise 1

True of False: there exists some $T \in \mathcal{L}(\mathbb{R}^3)$ such that T is not self-adjoint (with respect to the usual inner product) and there is a basis of $\mathbb{R}^3$ consisting of the eigenvectors of T.

Answer: True — Let $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ be given by

$$[T]_e^e = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Then, T has three distinct eigenvalues, namely 1, 2, and 3. So, $\mathbb{R}^3$ has a basis consisting of the eigenvectors of T. However,

$$[T^*]_e^e = ([T]_e^e)^* = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \neq [T]_e^e$$

Thus, $T^* \neq T$, i.e., T is not self-adjoint.

Example 16.2: 7B Exercise 2

Let T be self-adjoint on a finite-dimensional inner product space and that 2 and 3 are the only eigenvalues of T. Prove that $T^2 - 5T + 6I = 0$.

Proof: By the (real/complex) spectral theorem, there is a basis $v_1, \dots, v_k$ of 2-eigenvectors of T and a basis $v_{k+1}, \dots, v_n$ of 3-eigenvectors of T such that $v_1, \dots, v_n$ is a basis of V. Write $v = c_1v_1 + \cdots + c_nv_n$. Then,

$$\begin{aligned}(T^2 - 5T + 6I)(v) &= (T - 3I)(T - 2I)(v) \\ &= (T - 3I)(T - 2I)(c_1v_1 + \cdots + c_nv_n) \\ &= (T - 3I)(c_{k+1}v_{k+1} + \cdots + c_nv_n)\end{aligned}$$

since $(T - 2I)(v_j) = 0$ for $1 \leq j \leq k$ and $(T - 2I)(v_j) = 3v_j - 2v_j = v_j$ for $k + 1 \leq j \leq n$. Also,

$$\begin{aligned}(T - 3I)(c_{k+1}v_{k+1} + \cdots + c_nv_n) &= 0 \\ (T^2 - 5T + 6I)(v) &= 0\end{aligned}$$

since $(T - 3I)(v_j) = 0$ for $k + 1 \leq j \leq n$. The last equality holds for all $v \in V$ so we indeed have that $T^2 - 5T + 6I = 0$.

Example 16.3: 7B Exercise 3

Given an example of a $T \in \mathcal{L}(\mathbb{C}^3)$ such that 2 and 3 are the only eigenvalues of T and $T^2 - 5T + 6I \neq 0$. Choose a non-diagonalizable operator on $\mathbb{C}^3$, like

$$[T]_e^e = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Then,

$$\begin{aligned}\left[T^2 - 5T + 6I\right]_e^e &= \left[T^2\right]_e^e - 5[T]_e^e + 6I \\ &= \begin{bmatrix} 4 & 4 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} - \begin{bmatrix} 10 & 5 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 15 \end{bmatrix} + \begin{bmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\end{aligned}$$

Clearly, $T^2 - 5T + 6I \neq 0$.

Example 16.4: 7B Exercise 11

Prove or give a counterexample: every self-adjoint operator on V has a cube root.

Proof: If T is self-adjoint, then there is an orthonormal basis β of V such that

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

is diagonal with $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. Consider $S \in \mathcal{L}(V)$ such that

$$[S]_\beta^\beta = \begin{bmatrix} \sqrt[3]{\lambda_1} & & \\ & \ddots & \\ & & \sqrt[3]{\lambda_n} \end{bmatrix}$$

Then, $\left[S^3\right]_\beta^\beta = ([S]_\beta^\beta)^3 = [T]_\beta^\beta$. Thus, $S^3 = T$ and S is the desired cube root of T.

Example 16.5: 7B Exercise 14

Suppose that U is a finite-dimensional real vector space and $T \in \mathcal{L}(U)$. Prove that U has a basis consisting of the eigenvectors of T iff there is an inner product on U that makes T into a self-adjoint operator.

Proof: We will prove both directions:

- If there is an inner product that makes T self adjoint, then U has a basis consisting on the eigenvectors of T. This just follows from the real spectral theorem.
- On the other hand, suppose U has a basis consisting of the eigenvalues of T, given by $v_1, \ldots, v_n$. Define $\langle c_1 v_1 + \cdots + c_n v_n, d_1 v_1 + \cdots + d_n v_n \rangle = c_1 d_1 + \cdots + c_n d_n$. According to this inner product, $v_1, \ldots, v_n$ has to be an orthonormal basis of V. By the real spectral theorem, T is self-adjoint for this inner product.

17 Lecture 17

17.1 Positive Operators

Definition 17.1: Positive
The operator $T \in \mathcal{L}(V)$ is positive if T is self-adjoint and $\langle Tv, v\rangle \geq 0$ for all $v \in V$.

Example 17.1
If U is a subspace of V and $T = P_U = \text{proj}_U$ is the orthogonal projection onto U, then P_U is positive.

Proof: Orthogonally decompose $v, w \in V$ as $v = P_U v + (v - P_U v)$ and $w = P_U w + (w - P_U w)$. Then,

$$\langle P_U v, w\rangle = \langle P_U v, P_U w + (w - P_U w)\rangle = \langle P_U v, P_U w\rangle = \langle P_U v + (v - P_U v), P_U w\rangle = \langle v, P_U w\rangle$$

for all $v, w \in V$. So, $P_U^* = P_U$, i.e., the orthogonal projection operator is self adjoint. Also,

$$\langle P_U v, v\rangle = \langle P_U^2 v, v\rangle = \langle P_U v, P_U^* v\rangle = \langle P_U v, P_U v\rangle = \|P_U v\|^2 \geq 0$$

for all $v \in V$ so P_U is positive as well.

Theorem 17.1
T is positive iff T is self-adjoint and all eigenvalues of T are non-negative.

Proof: We will prove both directions:

- If T is positive, then it is self-adjoint by definition. Suppose that $Tv = \lambda v$ for $v \neq 0$. Then, $\langle Tv, v\rangle \geq 0 \implies \langle \lambda v, v\rangle = \lambda \|v\|^2 \geq 0$. Since $v \neq 0$, we get that $\lambda \geq 0$ instead.
- Conversely, assume that $T = T^*$ and all eigenvalues of T are nonnegative. We need to show that $\langle Tv, v\rangle \geq 0$ for all $v \in V$. By the Spectral Theorem, V has an orthonormal basis $e_1, \ldots, e_n$ consisting of the eigenvectors of T. Then,

$$v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n \implies Tv = \langle v, e_1\rangle \lambda_1 e_1 + \cdots + \langle v, e_n\rangle \lambda_n e_n$$

where $\lambda_1, \ldots, \lambda_n \geq 0$ are the eigenvalues of T. So, $\langle Tv, v\rangle = |\langle v, e_1\rangle|^2 \lambda_1 + \cdots + |\langle v, e_n\rangle|^2 \lambda_n \geq 0$, making T a positive operator.

Theorem 17.2
Let $T \in \mathcal{L}(V)$. The following statements are equivalent:

1. T is positive
2. T is self-adjoint and all eigenvalues of T are nonnegative
3. T has a positive square root
4. T has a self-adjoint square root
5. There exists some $R \in \mathcal{L}(V)$ such that $T = R^* R$

Proof: We have already shown statement 1 $\iff$ statement 2. We will now prove statement 2 $\implies$ statement 3 $\implies$ statement 4 $\implies$ statement 5 $\implies$ statement 1

- statement 2 $\implies$ statement 3

By the Spectral Theorem, there is an orthonormal basis $e_1, \ldots, e_n$ of V such that $[T]_e^e$ is diagonal with $\lambda_1, \ldots, \lambda_n \geq 0$. Let $S \in \mathcal{L}(V)$ such that

$$[S]_e^e = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix}$$

Then, $\left[S^2\right]_e^e = ([S]_e^e)^2 = [T]_e^e$, so $S^2 = T$. Since S is self-adjoint (by the Spectral Theorem), and its eigenvalues are $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n} \geq 0$, it is also positive.

- statement 3 $\implies$ statement 4

  This is just a restatement of what was said at the end of last proof.

- statement 4 $\implies$ statement 5

  If $T = R^2$ with $R = R^*$, then $T = RR = R^*R$.

- statement 5 $\implies$ statement 1

  If $T = R^*R$, then $T^* = (R^*R)^* = R^*R^{**} = R^*R$, so T is self-adjoint. Also, $\langle Tv, v\rangle = \langle R^*Rv, v\rangle = \langle Rv, R^{**}v\rangle = \langle Rv, Rv\rangle = \|Rv\|^2 \geq 0$, making T positive as well.

Example 17.2

Given an example of a 2×2 matrix A such that $A \neq I_2$ but $A^2 = I_2$.

There are two possible ways to think about it:

- Any matrix representing a reflection about a line through the origin. One special case of this is a rotation by π radians, but that's equivalent to a reflection across $y = x$.
- What about other kinds of matrices? Note that

$$A^2 = I \implies A = A^{-1} \implies A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = A^{-1}$$

So,

$$a = \frac{d}{\det(A)}$$
$$d = \frac{a}{\det(A)}$$
$$b = -\frac{b}{\det(A)}$$
$$c = -\frac{c}{\det(A)}$$

Then, $a = \frac{a}{(\det(A))^2}$ and $d = \frac{d}{(\det(A))^2}$ so either $a = d = 0$ or $\det(A) = \pm 1$.

If $\det(A) = 1$, then $b = c = 0$ and $a = d = 1$ or $a = d = -1$. Thus, if $A^2 = I$ and $\det(A) = 1$, then $A = \pm I$.

However, if $\det(A) = -1$, then $a = -d$ and

$$A = \begin{bmatrix} a & b \\ c & -a \end{bmatrix}$$

with $-a^2 - bc = -1$ or $a^2 + bc = 1$. There are many solution to an equation of this form:

  - If $a \neq \pm 1$, then b can be anything non-zero and c can be uniquely determined
  - If $a = \pm 1$, then either b or c must be 0 and the other can be anything

In fact, any matrix similar to a solution is also a solution. If $A^2 = I$ and $B = S^{-1}AS$, then $B^2 = S^{-1}ASS^{-1}AS = S^{-1}A^2S = S^{-1}IS = I$.

Note 17.1
The example above also shows that the only positive square root of I is I itself.

Theorem 17.3
Every positive operator on V has a unique positive square root.

17.2 Isometries

Definition 17.2: Isometry
Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be two normed vector spaces over $\mathbb{F}$. Then, $T \in \mathcal{L}(V, W)$ is an isometry if $\|Tv\|_W = \|v\|_V$ for all $v \in V$.

Note 17.2
Note that this is a more general definition than the one given in Axler. Often, the surjectivity of T is also required.

Theorem 17.4
Let $(V, \langle\cdot,\cdot\rangle_V)$ and $(W, \langle\cdot,\cdot\rangle_W)$ be two inner product spaces over $\mathbb{F}$ (either $\mathbb{R}$ or $\mathbb{C}$), both of dimension n. Then an isometric isomorphism exists between V and W.

Proof: Choose an orthonormal basis $e_1, \ldots, e_n$ of V and $f_1, \ldots, f_n$ of W. Then, $T(c_1e_1 + \cdots + c_ne_n) = c_1f_1 + \cdots + c_nf_n$ defines an isomorphism from V onto W. Since any $v \in V$ can be written as $v = c_1e_1 + \cdots + c_ne_n$, observe that

$$\begin{aligned}
\|v\|_V &= \langle v, v\rangle_V \\
&= \langle c_1e_1 + \cdots + c_ne_n, c_1e_1 + \cdots + c_ne_n\rangle \\
&= c_1^2 + \cdots + c_n^2 \\
\|Tv\|_W &= \langle Tv, Tv\rangle_W \\
&= \langle c_1f_1 + \cdots + c_nf_n, c_1f_1 + \cdots + c_nf_n\rangle \\
&= c_1^2 + \cdots + c_n^2
\end{aligned}$$

Since this holds for any arbitrary $v \in V$, the map T is an isometry.

Note 17.3
The proposition does not hold for normed vector spaces. For example, $\mathbb{R}^n$ with a p-norm and $\mathbb{R}^n$ with a p'-norm with $p \neq p'$ are not isometric.

Here is the more specific definition of an Isometry that Axler gives in his textbook:

Definition 17.3: Isometry
The operator $S \in \mathcal{L}(V)$, where V is a finite-dimensional inner product space, is an isometry iff $\|Sv\| = \|v\|$ for all $v \in V$.

Example 17.3
Here are two basis examples:

- If $V = \mathbb{R}^n$ and $\langle\cdot,\cdot\rangle$ is the dot product, then S is an isometry iff $[S]_e^e$ is an orthogonal matrix, i.e., $([S]_e^e)^T[S]_e^e = I$ (the columns of S are orthonormal).
- If $V = \mathbb{C}^n$ and $\langle\cdot,\cdot\rangle$ is the Euclidean inner product, then S is an isometry iff $[S]_e^e$ is a unitary matrix, i.e., $([S]_e^e)^*[S]_e^e = I$ (the columns of S are orthonormal).

Theorem 17.5
Suppose $S \in \mathcal{L}(V)$. Then, the following are equivalent:

1. S is an isometry
2. $\langle Su, Sv\rangle = \langle u, v\rangle$ for all $u, v \in V$
3. $Se_1, \dots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \dots, e_n$
4. There is an orthonormal basis $e_1, \dots, e_n$ of V such that $Se_1, \dots, Se_n$ is also orthonormal
5. $S^*S = I$
6. $SS^* = I$
7. S^* is an isometry
8. S is invertible with $S^{-1} = S^*$

Proof: We will not prove all of these statements. However, here is a sketch of two of the proofs:

- statement 1 $\implies$ statement 2

  Use the fact that

$$\langle Su, Sv\rangle = \frac{\|Su + Sv\|^2 - \|Su - Sv\|^2}{4}$$

- statement 4 $\implies$ statement 5

  Following the proof above, $\langle e_i, e_j\rangle = \langle Se_i, Se_j\rangle$ for an orthonormal $e_1, \dots, e_n$. So, $\langle S^*Se_i, e_j\rangle = \langle e_i, e_j\rangle$. Express $u, v \in V$ in terms of the orthonormal basis $e_1, \dots, e_n$ and expand to get that $\langle S^*Su, v\rangle = \langle u, v\rangle$ holds for all $u, v \in V$. Thus, $S^*S = I$.

Example 17.4
S is an isometry iff S is normal and all eigenvalues of S have absolute value 1.

Example 17.5: 7C Exercise 2
Suppose T is a positive operator on V. Suppose $v, w \in V$ are such that $Tv = w$ and $Tw = v$. Prove that $v = w$.

Proof: Note that $T(Tv) = Tw = v$ so $T^2v = v$. Either $v = 0$ and $w = 0$ or v is a 1-eigenvector of T^2. Since T is diagonalizable, the eigenvectors of T^2 are precisely those of T. So, v is either a $+1$-eigenvector or a -1-eigenvector of T. However, since T is positive, v must be a 1-eigenvector and $v = Tv = w$.

Example 17.6: 7C Exercise 7
Suppose T is a positive operator on T. Prove that T is invertible iff $\langle Tv, v\rangle > 0$ for all $v \neq 0$ in V.

Proof: If T is not invertible, then there is a vector $v \neq 0$ such that $Tv = 0$. For this, $\langle Tv, v\rangle = \langle 0, v\rangle = 0$. So, $\langle Tv, v\rangle > 0$ for all $v \neq 0$ implies that T is invertible (contrapositive).

If T is invertible and positive, then the positive square root S of T is also invertible. So, $\langle Tv, v\rangle = \langle S^*Sv, v\rangle = \langle Sv, Sv\rangle = \|Sv\|^2 > 0$ if $v \neq 0$ since $Sv \neq 0$.

Example 17.7: 7C Exercise 13
Prove or give a counterexample: If $S \in \mathcal{L}(V)$ and there is an orthonormal basis $e_1, \dots, e_n$ of V such that $\|Se_i\| = 1$ for each e_i, then S is an isometry.

Answer: This statement is False. Let $V = \mathbb{R}^2$ with $\langle \cdot, \cdot \rangle$ as the usual dot product, e_1, e_2 as the standard basis and

$$[S]_e^e = \begin{bmatrix} 1 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \end{bmatrix}$$

That is, $Se_1 = e_1 \implies \|e_1\| = 1$ and $Se_2 = \frac{1}{\sqrt{2}}e_1 + \frac{1}{\sqrt{2}}e_2 \implies \|Se_2\| = \frac{1}{2} + \frac{1}{2} = 1$. However, S is clearly not an isometry (it's not even normal).

Example 17.8: 7C Exercise 9

Prove or disprove: the identity operator on $\mathbb{R}^2$ has infinitely many self-adjoint square roots.

Proof: WLOG, let $\mathbb{F} = \mathbb{R}$ and $V = \mathbb{R}^2$ with $\langle \cdot, \cdot \rangle$ as the usual dot product. Then, $A^2 = I \implies A^T A = I$ implies that A is an orthogonal 2×2 matrix. Then, A is either a rotation or a reflection matrix. The only rotation matrices that satisfy $A^2 = I$ are rotations by 0 and π radians, i.e., $A = \pm I$. However, any reflection matrix satisfies $A^2 = I$. Then, the self-adjoint square roots of the identity operator have matrices, with respect to the standard basis, of the forms

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

where the last matrix is a reflection about the line $y = \tan\left(\frac{\theta}{2}\right)x$.

18 Lecture 18

18.1 Polar Decomposition

Theorem 18.1: Polar Decomposition
If $T \in \mathcal{L}(V)$, then there is an isometry $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$

Proof: This is a long proof so we will break it down into digestible parts:

- For any $v \in V$,

$$\begin{aligned}\|Tv\|^2 &= \langle Tv, Tv\rangle \\ &= \langle T^*Tv, v\rangle \\ &= \langle \sqrt{T^*T}\sqrt{T^*T}v, v\rangle \\ &= \langle \sqrt{T^*T}v, \sqrt{T^*T}v\rangle \\ &= \|\sqrt{T^*T}v\|^2\end{aligned}$$

  This follows since T^*T should be a positive operator with a positive square root.

- Define a linear map $S_1 : \text{range}(\sqrt{T^*T}) \mapsto \text{range}(T)$ by $S_1(\sqrt{T^*T}v) = Tv$. The main idea is that we are trying to extend S_1 to be an isometry on V such that $Tv = S_1\sqrt{T^*T}v$ for all $v \in V$.

- We first need to check that S_1 is well defined. Suppose that $\sqrt{T^*T}v_1 = \sqrt{T^*T}v_2$. We need to show that $Tv_1 = Tv_2$. Observe that $\ker(T) \subseteq \ker(\sqrt{T^*T}) \subseteq \ker(T^*T) = \ker(T)$. So, $\sqrt{T^*T}v_1 = \sqrt{T^*T}v_2 \implies v_1 - v_2 \in \ker(\sqrt{T^*T}) = \ker(T)$. Thus, $T(v_1 - v_2) = 0 \implies Tv_1 = Tv_2$.

- By the linearity of $\sqrt{T^*T}$ and T, the map S_1 is linear. Moreover, $\|S_1(\sqrt{T^*T}v)\| = \|Tv\| = \|\sqrt{T^*T}v\|$ (from above). This implies that S_1 is indeed an from $\text{range}(\sqrt{T^*T}) \mapsto \text{range}(T)$.

- We need to extend S_1 to be an isometry from $V \mapsto V$. Since V is finite-dimensional and $\ker(\sqrt{T^*T}) = \ker(T)$, then $\dim\text{range}(\sqrt{T^*T}) = \dim\text{range}(T) \implies \dim\text{range}(\sqrt{T^*T})^\perp = \dim\text{range}(T)^\perp$. Therefore, we can choose an orthonormal basis $e_1, \ldots, e_m$ of $\text{range}(\sqrt{T^*T})^\perp$ and an orthonormal basis $f_1, \ldots, f_m$ of $\text{range}(T)^\perp$. Define the map $S_2 : \text{range}(\sqrt{T^*T})^\perp \mapsto \text{range}(T)^\perp$ as

$$S_2(a_1e_1 + \cdots + a_me_m) = a_1f_1 + \cdots + a_mf_m$$

  Thus, S_2 maps $\text{range}(\sqrt{T^*T})^\perp$ isometrically onto $\text{range}(T)^\perp$ since

$$\|a_1e_1 + \cdots + a_me_m\|^2 = |a_1|^2 + \cdots + |a_m|^2 = \|a_1f_1 + \cdots + a_mf_m\|^2$$

- Let $v = u + w$ where $u \in \text{range}(\sqrt{T^*T})$ and $w \in \text{range}(\sqrt{T^*T})^\perp$. Define $Sv = S_1u + S_2w$. Since

$$\|Sv\|^2 = \|S_1u + S_2w\|^2 = \|S_1u\|^2 + \|S_2w\|^2 = \|u\|^2 + \|w\|^2 = \|u + w\|^2$$

  the map S is an isometry on V.

- Now, we only need to check that $T = S\sqrt{T^*T}$. First, note that $\sqrt{T^*T}v \in \text{range}(\sqrt{T^*T})$ satisfies $\sqrt{T^*T}v = u + w$ where $u = \sqrt{T^*T}v \in \text{range}(\sqrt{T^*T})$ and $w = 0 \in \text{range}(\sqrt{T^*T})^\perp$. Therefore, $S\sqrt{T^*T}v = S_1(\sqrt{T^*T}v) + S_2(0) = Tv$ for all $v \in V$.

18.2 Singular Values

Definition 18.1: Singular Values
Suppose that $T \in \mathcal{L}(V)$. The singular values of T are the eigenvalues of $\sqrt{T^*T}$.

Note 18.1
Actually, if $T \in \mathcal{L}(V, W)$ and W is also a finite-dimensional inner product space over $\mathbb{F}$, then the eigenvalues of $\sqrt{T^*T}$ are still defined (since $\sqrt{T^*T} \in \mathcal{L}(V)$). The singular values of $T \in \mathcal{L}(V, W)$ are the eigenvalues of $\sqrt{T^*T}$ so they are defined as well.

Note 18.2
It is a universal convention to list singular values in non-increasing order. If $\dim = n$, then common notations for the singular values of T are $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ or $s_1 \geq s_2 \geq \cdots \geq s_n$.

Note 18.3
If $T \in \mathcal{L}(V)$ and $\dim V = n$, then $\sqrt{T^*T}$ is a positive operator and V has n non-negative real eigenvalues $s_1 \geq \cdots \geq s_n \geq 0$ (some s_j possibly repeated) by the Spectral Theorem.

Theorem 18.2: Singular Value Decomposition
Suppose $T \in \mathcal{L}(V)$ has singular values $s_1, \ldots, s_n$. Then, there is an orthonormal bases $e = e_1, \ldots, e_n$ and $f = f_1, \ldots, f_n$ in V such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

i.e.,

$$[T]_e^f = \begin{bmatrix} s_1 & 0 & \ldots & 0 \\ 0 & s_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & s_n \end{bmatrix}$$

Proof: Note that $\sqrt{T^*T}$ is a positive operator. So, by the spectral theorem, there is an orthonormal basis $e_1, \ldots, e_n$ of V such that $\sqrt{T^*T}e_j = s_je_j$ for all $j \leq n$. Moreover, $v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n$ since $e_1, \ldots, e_n$ is orthonormal. Then,

$$\sqrt{T^*T}v = s_1\langle v, e_1\rangle e_1 + \cdots + s_n\langle v, e_n\rangle e_n$$

Following the polar decomposition of T,

$$Tv = S\sqrt{T^*T}v = s_1\langle v, e_1\rangle Se_1 + \cdots + s_n\langle v, e_n\rangle Se_n$$

Since S is an isometry, the list $Se_1, \ldots, Se_n$ is also an orthonormal basis. Taking $f_j = Se_j$, we get the desired result.

Theorem 18.3
If $T \in \mathcal{L}(V)$, then the singular values of T are the non-negative square roots of the eigenvalues of T^*T.

Proof: By the spectral theorem, there is an orthonormal basis $e_1, \ldots, e_n$ of V with $\sqrt{T^*T}e_k = s_ke_k$ for all $k \leq n$. However, that also makes e_k an s_k^2-eigenvector of T^*T. In other words, $E(s_k, \sqrt{T^*T}) = E(s_k^2, T^*T)$. Moreover, since $\sqrt{T^*T}$ and T^*T are both diagonalizable, the multiplicities of their eigenvalues must also be the same.

Definition 18.2: Operator Norm
Let V be a normed vector space and $T \in \mathcal{L}(V)$. Define

$$\|T\| = \max_{x \neq 0} \frac{\|Tx\|}{\|x\|} = \max_{x \neq 0} \left\| T\left(\frac{x}{\|x\|}\right) \right\| = \max_{\|x\|=1} \|T(x)\|$$

In fact, if $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$, then $\|T\|$ will equal σ_1, the largest singular value of T.

Theorem 18.4
Here are some properties of an operator norm:

1. $\|T\| = 0$ iff T is the zero operator
2. $\|\alpha T\| = |\alpha|\|T\|$ if $\alpha \in \mathbb{F}$
3. $\|T + S\| \leq \|T\| + \|S\|$

Proof: Proofs of statements 1 and 2 is left as an exercise for the reader. We will prove statement 3 now:

$$\begin{aligned} \|T + S\| &= \max_{\|x\|=1} \|(T + S)x\| \\ &= \max_{\|x\|=1} \|Tx + Sx\| \\ &\leq \max_{\|x\|=1} \|Tx\| + \|Sx\| \\ &\leq \max_{\|x\|=1} \|Tx\| + \max_{\|x\|=1} \|Sx\| \\ &= \|T\| + \|S\| \end{aligned}$$

Example 18.1: 7D Exercise 12
Prove or give a counterexample: If $T \in \mathcal{L}(V)$, then the singular values of T^2 are the squares of the singular values of T.

Answer: This statement is false. Consider the backward shift $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ defined over the usual dot product. Then,

$$\begin{aligned} [T]_e^e &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \\ [T^*]_e^e &= \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \\ [T^*T]_e^e &= \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

Thus, the singular values of T are $\sqrt{1} = 1$ and $\sqrt{0} = 0$. However,

$$\begin{aligned} \left[T^2\right]_e^e &= ([T]_e^e)^2 \\ &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}^2 \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

Thus, T^2 is the zero operator and its singular values are just 0 and 0.

Example 18.2: 7D Exercise 13
Prove that T is invertible iff 0 is not a singular value of T.

Proof: Note that

$$\begin{aligned} 0 \text{ is not a singular value of } T &\iff 0 \text{ is not an eigenvalue of } T^*T \\ &\iff \ker(T^*T - 0I) = \{0\} \\ &\iff \ker(T^*T) = \{0\} \end{aligned}$$

$$\iff \ker(T) = \{0\}$$
$$\iff T \text{ is invertible}$$

Theorem 18.5
If $T \in \mathcal{L}(V)$, then $\|T\|$ exists in the case that $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.

Proof: During the existence proof of the polar decomposition, we showed that $\|Tv\|^2 = \|\sqrt{T^*T}v\|^2$ for all $v \in V$. Moreover, $\sqrt{T^*T}v = s_1\langle v, e_1\rangle e_1 + \cdots + s_n\langle v, e_n\rangle e_n$, i.e.,

$$\left[\sqrt{T^*T}\right]_e^e = \begin{bmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{bmatrix}$$

where $s_1 \geq \cdots \geq s_n \geq 0$. Then,

$$\begin{aligned} \|\sqrt{T^*T}v\|^2 &= s_1^2|\langle v, e_1\rangle|^2 + \cdots + s_n^2|\langle v, e_n\rangle|^2 \\ \|v\|^2 &= |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2 \\ &= c_1^2 + \cdots + c_n^2 \end{aligned}$$

If $\|v\| = 1$, then $c_1^2 + \cdots + c_n^2 = 1$.

We can vary $c_1, \ldots, c_n$ over all possible coefficients such that $c_1^2 + \cdots + c_n^2 = 1$ and $\|\sqrt{T^*T}v\|^2$ is maximized. Letting $c_1^2 = 1$, i.e., taking $v = e_1$ works. Why? Maximizing $\|\sqrt{T^*T}v\|$ is equivalent to maximizing $s_1^2c_1^2 + \cdots + s_n^2c_n^2$ constrained to $c_1^2 + \cdots + c_n^2 = 1$ and $s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$. Note that

$$\begin{aligned} s_1^2c_1^2 + s_2^2c_2^2 + \cdots + s_n^2c_n^2 &\leq s_1^2c_1^2 + s_1^2c_2^2 + \cdots + s_1^2c_n^2 \\ &= s_1^2(c_1^2 + \cdots + c_n^2) \\ &= s_1^2 \end{aligned}$$

Thus, choosing $s_1^2c_1^2 + s_2^2c_2^2 + \cdots + s_n^2c_n^2 = s_1^2$ will maximize the given norm. Therefore, $\|T\| = \|\sqrt{T^*T}\| = s_1$ is the largest singular value of T.

Note 18.4
Just like $\max_{\|x\|=1} \|T(x)\| = s_1$, the largest singular value of T, we can show that $\min_{\|x\|=1} \|T(x)\| = s_n$, the smallest singular value of T.

Note 18.5
Since the singular values of T are the eigenvalues of $\sqrt{T^*T}$ and T^* depends on the inner product on V, talking about the singular values independently of some choice of inner product doesn't make sense.

Note 18.6
The singular values of T are not similarity invariant of T, i.e., if S is invertible, then T and $S^{-1} \circ T \circ S$ might not have the same singular values.

Example 18.3
Let

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & 0 \\ a & 0 \end{bmatrix}$$

for $a > 0$ be defined over $\mathbb{R}^2$ with the regular dot product as the inner product. Then,

$$\begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}}_{A} \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix}^{-1} = \underbrace{\begin{bmatrix} 1 & 0 \\ a & 0 \end{bmatrix}}_{B}$$

Thus, A and B are similar matrices. Note that

$$A^T A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

so the singular values of A are $\sqrt{2}$ and 0. However, B is a skew projection onto the lines $y = ax$ and

$$B^T B = \begin{bmatrix} 1 & a \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ a & 0 \end{bmatrix} = \begin{bmatrix} 1 + a^2 & 0 \\ 0 & 0 \end{bmatrix}$$

so the singular values of B are $\sqrt{1 + a^2}$ and 0 instead.

19 Lecture 19

19.1 Singular Value Decomposition Cont.

We considered SVD at a transformation (specifically, operator) level before. Now consider it at a matrix level: let A be an $m \times n$ matrix over $\mathbb{R}$. Then, we can decompose it as $A = U\Sigma V^T$ where U is an $m \times m$ orthogonal matrix U, V is an $n \times n$ orthogonal matrix and Σ is an $m \times n$ matrix with the singular values of A along its diagonal. How can we arrive at this decomposition?

1. Consider $B = A^T A$. Then, B is a symmetric (or self-adjoint since it has real entries) $n \times n$ matrix. The spectral theorem applies here and B has real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ with corresponding eigenvectors $v_1, \ldots, v_n$ that form an orthonormal basis of $\mathbb{R}^n$. We claim that all λ_k's are non-negative, i.e., $A^T A$ is positive:
$$\begin{aligned}\left\langle A^T A v_k, v_k \right\rangle &= \langle \lambda_k v_k, v_k \rangle \\ &= \lambda_k \|v_k\|^2 \\ &= \lambda_k \\ \left\langle A^T A v_k, v_k \right\rangle &= \left\langle A v_k, (A^T)^T v_k \right\rangle \\ &= \langle A v_k, A v_k \rangle \\ &= \|A v_k\|^2\end{aligned}$$
Thus, $\lambda_k = \|A v_k\|^2 \geq 0$ for all $j \leq n$.
2. Let $\sigma_k = \sqrt{\lambda_k}$. These σ_k will form the singular values of A.

Note 19.1
If $\lambda_1 \geq \cdots \geq \lambda_n$ and $\sigma_1 \geq \cdots \geq \sigma_n$, then $\sigma_k > 0$ iff $k \leq r = \operatorname{rank}(A^T A) = \operatorname{rank}(A)$. Since $A^T A$ is symmetric, its eigenvectors form an orthogonal matrix S such that
$$S^T A^T A S = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$
where all λ_k are arranged in non-increasing order and $\lambda_k = 0$ for $k > r = \operatorname{rank}(A^T A) = \operatorname{rank}(A)$. In other words, the rank of a matrix is the number of nonzero entries along the main diagonal of its diagonalized matrix representation.

3. Consider $Av_1, \ldots, Av_r$ where $r = \operatorname{rank}(A) = \operatorname{rank}(A^T A)$. Let $u_k = \frac{Av_k}{\sigma_k}$ so that $Av_k = \sigma_k u_k$.
We claim that all Av_k's are orthogonal to each other and $\|Av_k\| = s_k$ for all $k \leq n$. In other words, $\|u_k\| = 1$ for $k \leq r$. Why? Note that
$$\langle Av_k, Av_l \rangle = \left\langle v_k, A^T A v_l \right\rangle = \langle v_k, \lambda_l v_l \rangle = \lambda_l \langle v_k, v_l \rangle$$
Since $\langle v_k, v_l \rangle = 0$, the vectors Av_k, Av_l are also orthogonal. Moreover,
$$\|Av_k\|^2 = \lambda_k \|v_k\|^2$$
so $\|Av_k\| = \sqrt{\lambda_k} = \sigma_k$. Therefore, $\|u_k\| = 1$ as desired.
4. We have defined u_k for $k \leq r$. What about $k > r$? We can simply extend $u_1, \ldots, u_r$ to an orthonormal basis $u_1, \ldots, u_m$ of $\mathbb{R}^m$ using the Gram-Schmidt process.

In summary, if A is an $m \times n$ real matrix, then we found an orthonormal basis $v_1, \ldots, v_n$ of $\mathbb{R}^n$ based on the singular values of A, and used these to define an orthonormal basis $u_1, \ldots, u_m$ of $\mathbb{R}^m$ such that $Av_k = \sigma_k u_k$ for $k \leq r$ and $Av_k = 0$ for $k > r$, where $r = \operatorname{rank}(A^T A) = \operatorname{rank}(A) \leq n$. How can we express this in a matrix form?
$$A \underbrace{\begin{bmatrix} v_1 & \ldots & v_n \end{bmatrix}}_{V} = \begin{bmatrix} Av_1 & \ldots & Av_r & Av_{r+1} & \ldots & Av_n \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1 u_1 & \dots & \sigma_r u_r & 0 & \dots & 0 \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} u_1 & \dots & u_r & u_{r+1} & \dots & u_m \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}}_{\Sigma}$$

where $u_{r+1}, \dots, u_m$ are multiplied by the 0 entries in Σ. Thus, $AV = U\Sigma$. Since V is a square matrix whose columns are orthonormal vectors, $V^T V = VV^T = I_n$. Thus, $V^{-1} = V^T$ and $A = U\Sigma V^T$.

Example 19.1
Let

$$A = \begin{bmatrix} p & -q \\ q & p \end{bmatrix} \implies A^T A = \begin{bmatrix} p & q \\ -q & p \end{bmatrix} \begin{bmatrix} p & -q \\ q & p \end{bmatrix} = \begin{bmatrix} p^2 + q^2 & 0 \\ 0 & p^2 + q^2 \end{bmatrix}$$

So, the singular values of A are $\sqrt{p^2+q^2}$ and $\sqrt{p^2+q^2}$. Geometrically, A is just a rotation matrix times a scalar stretching, i.e.,

$$A = \sqrt{p^2+q^2} \begin{bmatrix} \frac{p}{\sqrt{p^2+q^2}} & -\frac{q}{\sqrt{p^2+q^2}} \\ \frac{q}{\sqrt{p^2+q^2}} & \frac{p}{\sqrt{p^2+q^2}} \end{bmatrix}$$

Note that the SVD of A is

$$A = \begin{bmatrix} \frac{p}{\sqrt{p^2+q^2}} & -\frac{q}{\sqrt{p^2+q^2}} \\ \frac{q}{\sqrt{p^2+q^2}} & \frac{p}{\sqrt{p^2+q^2}} \end{bmatrix} \begin{bmatrix} \sqrt{p^2+q^2} & 0 \\ 0 & \sqrt{p^2+q^2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

This is actually the polar decomposition of A as well! Since $\|Av_1\| = \|Ae_1\| = \sqrt{p^2+q^2}$ and $\|Av_2\| = \|Ae_2\| = \sqrt{p^2+q^2}$, the matrix A takes the unit square to a square with area $|\det(A)| = p^2 + q^2 = \sigma_1 \cdot \sigma_2$.

Example 19.2
Let

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \implies A^T A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

The eigenvalues of $A^T A$ are $\lambda_1 = 3, \lambda_2 = 1$ so $\sigma_1 = \sqrt{3}, \sigma_2 = 1$. Moreover,

$$E(3, A^T A) = \text{span}\left(\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right) = \text{span}(v_1)$$
$$E(1, A^T A) = \text{span}\left(\begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right) = \text{span}(v_2)$$

And,

$$Av_1 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 2/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$u_1 = \frac{Av_1}{\sigma_1} = \begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$$
$$Av_2 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}$$
$$u_2 = \frac{Av_1}{\sigma_1} = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}$$

We need to find a third u_3 that is orthonormal to u_1, u_2. Note that

$$u_3 = \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

works. Thus, the SVD of A is given by

$$V = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$
$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$
$$U = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}$$
$$A = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Observe that V^T is a rotation matrix, Σ is an embedding into $\mathbb{R}^3$ that also stretches/scales along the x-coordinate and U is another orthogonal matrix that is neither a rotation nor a reflection!

Theorem 19.1
Let A be an $m \times n$ matrix and $v \in \mathbb{R}^n$. Then, $\sigma_1 \|v\| \geq \|Av\| \geq \sigma_n \|v\|$ where $\sigma_1, \ldots, \sigma_n$ are the singular values of A.

Proof: Let $v_1, \ldots, v_n$ be an orthonormal basis of the eigenvectors of $A^T A$. Write $v = c_1 v_1 + \cdots + c_n v_n$. Then, $\|v\|^2 = c_1^2 + \cdots + c_n^2$ and

$$\begin{aligned} \|Av\|^2 &= \|c_1 Av_1 + \cdots + c_n Av_n\|^2 \\ &= c_1^2 \|Av_1\|^2 + \cdots + c_n^2 \|Av_n\|^2 \\ &= c_1^2 \sigma_1^2 + \cdots + c_n^2 \sigma_n^2 \end{aligned}$$

by the Pythagorean theorem. Then,

$$\begin{gathered} c_1^2\sigma_1^2 + \cdots + c_n^2\sigma_1^2 \geq c_1^2\sigma_1^2 + \cdots + c_n^2\sigma_n^2 \geq c_1^2\sigma_n^2 + \cdots + c_n^2\sigma_n^2 \\ \sigma_1^2(c_1^2 + \cdots + c_n^2) \geq \|Av\|^2 \geq \sigma_n^2(c_1^2 + \cdots + c_n^2) \\ \sigma_1^2\|v\|^2 \geq \|Av\|^2 \geq \sigma_n^2\|v\|^2 \end{gathered}$$

Thus, $\sigma_1\|v\| \geq \|Av\| \geq \sigma_n\|n\|$.

Just like we defined operator norms earlier, if we extend this definition to matrices, then

$$\begin{aligned}\|A\| &= \max_{\|x\|\neq 0} \frac{\|Ax\|}{\|x\|} \\ &= \max_{\|x\|\neq 0} \left\| A\frac{x}{\|x\|} \right\| \\ &= \max_{\|v\|=1} \|Av\|\end{aligned}$$

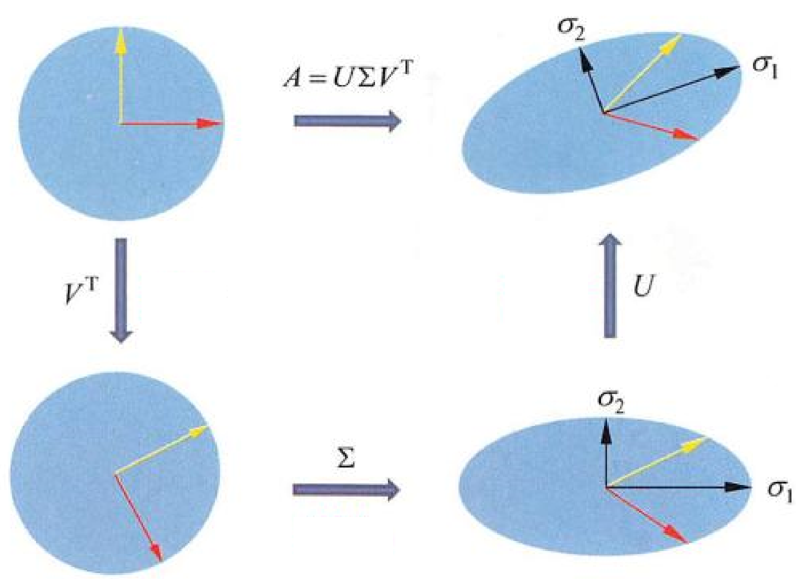
We showed that $\|Av\| \leq \sigma_1 \|v\|$ for all $v \in \mathbb{R}^n$. So,

$$\frac{\|Av\|}{\|v\|} \leq \sigma_1$$

if $\|v\| \neq 0$. Then, $\|A\| \leq \sigma_1$. However, $\|Av\| = \sigma_1$ and $\|v\| = 1$ so $\|A\| = \sigma_1$ is satisfied with equality as expected.

19.2 Geometry of the SVD

What happens if we look at each matrix in a singular value decomposition of an operator separately? The figure below depicts an example of a 2×2 matrix:



Theorem 19.2
Let $A : \mathbb{R}^2 \mapsto \mathbb{R}^2$ (for $V = \mathbb{R}^2$ with the usual dot product) and A is invertible. The image of the unit circle in $\mathbb{R}^2$ under A is an ellipse. The length of the semi-major axis of the ellipse is σ_1, and the length of the semi-minor axis is σ_2. More generally, if $A : \mathbb{R}^n \mapsto \mathbb{R}^n$ is invertible, then the image of the unit sphere in $\mathbb{R}^n$ under A is an n-dimensional ellipsoid with axes of lengths $\sigma_1, \ldots, \sigma_n$.

- The matrix $V^T = V^{-1}$ is a rotation matrix
- The matrix Σ is a scaling/stretching matrix
- The matrix U is an orthogonal matrix (can be a rotation, reflection or neither)

Theorem 19.3
Let λ be a real eigenvalue of an $n \times n$ matrix A. Then, $\sigma_n \leq |\lambda| \leq \sigma_1$.

Proof: Let v be a unit length λ-eigenvector of A. Thus, $\sigma_n\|v\| \leq \|Av\| \leq \sigma_1\|v\| \implies \sigma_1\|v\| \leq \|\lambda v\| \leq \sigma_1\|v\| \implies \sigma_1\|v\| \leq |\lambda|\|v\| \leq \sigma_1\|v\|$. Then, $\sigma_n \leq |\lambda| \leq \sigma_1$ as intended.

Note 19.2
The result above is still true if λ is a complex eigenvalue of A.

Example 19.3
Let

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \implies A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

Both eigenvalues of A are 1 but its singular values are $\frac{3 \pm \sqrt{5}}{2}$ instead. Note that $\frac{3-\sqrt{5}}{2} \leq 1 \leq \frac{3+\sqrt{5}}{2}$ as expected.

Theorem 19.4
Let A be an $m \times n$ matrix. Then, $A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$ where $r = \text{rank}(A)$.

Proof: Σ is an $m \times n$ matrix with $\Sigma_{j,j} = \sigma_j$ if $j \leq r$, and all of the other entries are 0. Then, break it down as $\Sigma = \Sigma_1 + \cdots + \Sigma_r$ where Σ_j has the (j, j)th entry as σ_j but all other entries are 0. Consider V_k^T to be the kth column of V^T. Then,

$$U\Sigma_j V_k^T = (V^T)_{1k}(U\Sigma_j)_1 + \cdots + (V^T)_{nk}(U\Sigma_j)_n = (V^T)_{j,k}(U\Sigma_j)_j$$

since all of the other terms will get cancelled by the 0s in Σ_j. We just showed that $U\Sigma_j(V^T)_k = (V^T)_{j,k}\sigma_j u_j$. Then,

$$\begin{aligned} U\Sigma_j V^T &= \begin{bmatrix} (V^T)_{j1}\sigma_j u_j & (V^T)_{j2}\sigma_j u_j & \ldots & (V^T)_{jn}\sigma_j u_j \end{bmatrix} \\ &= \sigma_j u_j \begin{bmatrix} (V^T)_{j1} & \ldots & (V^T)_{jn} \end{bmatrix} \\ &= \sigma_j u_j v_j^T \end{aligned}$$

This follows since $\begin{bmatrix} (V^T)_{j1} & \ldots & (V^T)_{jn} \end{bmatrix}$ is the jth row of V^T, i.e., the transpose of the jth column of V. Thus,

$$\begin{aligned} A &= U\Sigma V^T \\ &= U(\Sigma_1 + \Sigma_2 + \cdots + \Sigma_n)V^T \\ &= U\Sigma_1 V^T + \cdots + U\Sigma_n V^T \\ &= \sigma_1 u_1 v_1^T + \cdots + \sigma_n u_n v_n^T \end{aligned}$$

Example 19.4
Let

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{V^T}$$

Then,

$$\begin{aligned} A &= \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T \\ &= 2\begin{bmatrix} 0 \\ 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \end{bmatrix} + 1\begin{bmatrix} 1 \\ 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \end{bmatrix} \end{aligned}$$

Example 19.5

A satellite transmits a picture containing 1000×1000 pixels. If the color of each pixel is digitized, this information can be represented in 1000×1000 matrix A with 1,000,000 entries.

If $A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$, even though r may be 1000 as well, many singular values of A will most likely be very small compared to first few singular values, and will contribute little to the actual image. If we ignore all but say, $k = 10$ of the singular values and transmit $\sigma_1 u_1 v_1^T + \cdots + \sigma_k u_k v_k^T$ as an approximation to A, then we only used about $k(1000 + 1000) = 10 \cdot 2000 = 20000$ numbers instead of a million entries to obtain a hopefully reasonably good reconstruction of the original image. This is precisely the main idea behind dimensionality reduction via low-rank approximation, and it has various applications in computer science and engineering!

Note 19.3

The SVD is hardly the only way to write A as a sum of rank 1 outer products. It just happens to have a very special property that the kth partial sum $A_k = \sum_{i=1}^k \sigma_i u_i v_i^T$ carries as much "energy" of A as possible in the sense that $\|A - B\| \geq \|A - A_k\| = \sigma_{k+1}$ (look at the theorem below) for all matrices B of rank $\leq k$. Basically, A_k gives the best rank k approximation to A in this sense.

This implies, for example, that if A is invertible, then the closest non-invertible matrix B to A is obtained by changing the smallest σ_n of A to 0, and otherwise leaving the SVD unchanged. In other words, if $A = U\Sigma V^T$ and Σ' is Σ but with $\Sigma'_{n,n} = 0$, then $B = U\Sigma' V^T$ will be non-invertible.

Theorem 19.5: Eckart-Young-Misrky Theorem

Let A be an $m \times n$ with $\operatorname{rank}(A) = r$, such that $A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$. If B is $m \times n$ and $\operatorname{rank}(B) \leq k < r$ and $A_k = \sigma_1 u_1 v_1^T + \cdots + \sigma_k u_k v_k^T$, then $\|A - B\| \geq \|A - A_k\| = \sigma_{k+1}$.

Proof: First of all,

$$A - A_k = \sigma_{k+1} u_{k+1} v_{k+1}^T + \cdots + \sigma_r u_r v_r^T$$

The expression on the RHS is an SVD of $A - A_k$ with σ_{k+1} as the largest singular value so $\|A - A_k\| = \sigma_{k+1}$.

Now, for the sake of contradiction, assume that $\|A - B\| < \|A - A_k\|$ for some B with $\operatorname{rank}(B) \leq k$. Then, following rank-nullity, $\dim\ker(B) = n - \operatorname{rank}(B) \geq n - k$. So, there is an $n - k$ dimensional subspace $W \subseteq \ker(B) \subseteq \mathbb{R}^n$ such that $Bw = 0$ for all $w \in W$. Therefore, $\|Aw\| = \|(A - B)w\| \leq \|A - B\|\|w\| < \sigma_{k+1}\|w\|$ for $w \neq 0 \in W$.

Now, let's consider all w such that $\|Aw\| \geq \sigma_{k+1}\|w\|$. Note that $V_{k+1} = \operatorname{span}(v_1, \ldots, v_{k+1})$ is a $k+1$ dimensional subspace of $\mathbb{R}^n$ such that $\|Aw\| \geq \sigma_{k+1}\|w\|$ for all $w \in V_{k+1}$. In fact, this is the largest subspace that will yield the lower bound above — $w \in V_{k+2}$ will yield an even smaller lower bound of $\sigma_{k+2}\|w\|$ while V_k will not capture any w such that $\sigma_k\|w\| > \|Aw\| \geq \sigma_{k+1}\|w\|$. Since $\dim W + \dim V_{k+1} = (n - k) + (k + 1) = n + 1$, there must be a non-zero vector $w \in W \cap V_{k+1}$ that will satisfy both inequalities above — this is a contradiction!

Note 19.4

If A is not normal, then by perturbing A slightly, it is possible to change the eigenvalues (or some subset of the eigenvalues) of A fairly significantly. However, if A is normal, then a small change in A to another normal matrix will result in a very small change to its eigenvalues. Since $A^T A$ is always normal, even if A isn't, the singular values of A don't change much with a slight perturbation of A. This makes the SVD highly robust and stable.

20 Lecture 20

20.1 Review Problems

We went over a practice midterm and a past midterm in preparation for midterm 2.

21 Lecture 21

21.1 Generalized Eigenvectors

Theorem 21.1
If $T \in \mathcal{L}(V)$, then $\{0\} \subset \ker(T) \subset \ker(T^2) \subset \cdots \subset \ker(T^k) \subset \ker(T^{k+1}) \subset \ldots$

Proof: If $T^k v = 0$, then $T^{k+1}(v) = T(T^k v) = T(0) = 0$ by the linearity of T.

Theorem 21.2
Let $T \in \mathcal{L}(V)$. If $\ker(T^m) = \ker(T^{m+1})$, then $\ker(T^m) = \ker(T^{m+1}) = \ker(T^{m+2}) = \ldots$.

Proof: We need to show that if $T^{m+j+1}(v) = 0$, then $T^{m+j}(v) = 0$ for all $j \geq 0$. Suppose $T^{m+j+1}(v) = 0$. Then,

$$\begin{aligned} 0 &= T^{m+j+1}(v) \\ &= T^{m+1}(T^j(v)) \end{aligned}$$

Thus, $T^j(v) \in \ker(T^{m+1}) = \ker(T^m)$ and $T^m(T^j(v)) = T^{m+j}(v) = 0$.

Theorem 21.3
Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then, $\ker(T^n) = \ker(T^{n+1}) = \ker(T^{n+2}) = \ldots$.

Proof: First off, the chain $\{0\} = \ker(T^0) \subset \ker(T) \subset \ker(T^2) \subset \ldots$ must eventually stabilize, as otherwise V would be infinite-dimensional. Suppose the chain stabilizes when $k = j$, i.e., $\{0\} \subset \ker(T) \subset \cdots \subset \ker(T^j) = \ker(T^{j+1}) = \ldots$. Then, $\dim \ker(T^j) \geq j$ since the dimension of each kernel increases by 1 as the chain continues (If it did not, then the chain would stabilize before j). Since you can't have a dimension greater than n, we must have that $j \leq n$.

Recall that $V = \ker(T) \oplus \operatorname{range}(T)$ can easily fail (though if T is normal, this statement is true), as evident by

$$[T]_e^e = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

In this case, $\ker(T) = \operatorname{range}(T) = \operatorname{span}(e_1)$ so $\mathbb{R}^2 \neq \ker(T) + \operatorname{range}(T)$. However,

Theorem 21.4
If $\dim V = n$ and $T \in \mathcal{L}(V)$, then $V = \ker(T^n) \oplus \operatorname{range}(T^n)$.

Proof: If $T^n x = 0$ and $x = T^n y$ for some $y \in V$ (i.e., if $x \in \ker(T^n) \cap \operatorname{range}(T^n)$), then $T^n x = 0 \implies T^{2n} y = 0$. Then, $y \in \ker(T^{2n}) = \ker(T^n)$ and $x = T^n y = 0$ so $\ker(T^n) \cap \operatorname{range}(T^n) = \{0\}$. Moreover, applying the rank-nullity theorem to T^n also yields $\dim V = \dim \ker(T^n) + \dim \operatorname{range}(T^n)$. Thus, $V = \ker(T^n) \oplus \operatorname{range}(T^n)$.

Definition 21.1: Generalized Eigenvectors
If $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T, then if $v \neq 0$ satisfies $(T - \lambda I)^j v = 0$ for some $j \geq 1$, then v is a generalized eigenvector of T (for the eigenvalue λ, of course).

Definition 21.2: Generalized Eigenspace
If λ is an eigenvalue of $T \in \mathcal{L}(V)$, then the generalized eigenspace of T for the eigenvalue λ, denote by $G(\lambda, T)$, is the set of all generalized λ-eigenvectors of T, including 0.

Note 21.1
Clearly $E(\lambda, T) \subseteq G(\lambda, T)$ and it is easy to check that $G(\lambda, T)$ is a subspace of V.

Theorem 21.5
If $\dim V = n$ and λ is an eigenvalue of T, then $G(\lambda, T) = \ker((T - \lambda I)^n)$.

Proof: The vector $v \in G(\lambda, T)$ if $(T - \lambda I)^j v = 0$ for some $j \geq 0$, i.e., if $v \in \ker((T - \lambda I)^j)$. However, for all $j \leq n$, we have that $\ker((T - \lambda I)^j) \subseteq \ker((T - \lambda I)^n)$ so $v \in \ker((T - \lambda I)^n)$. Thus, $G(\lambda, T) \subseteq \ker((T - \lambda I)^n)$ and, trivially, $\ker((T - \lambda I)^n) \subseteq G(\lambda, T)$.

Example 21.1
Let $T \in \mathcal{L}(\mathbb{R}^3)$ such that

$$[T]_e^e = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Since $[T]_e^e$ is upper triangular, the only eigenvalue of T is $\lambda = 1$. Then,

$$E(1, T) = \ker(T - I) = \ker \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \operatorname{span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right)$$

What about $\ker((T - I)^j)$? For $j = 2$,

$$\ker((T - I)^2) = \ker \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \operatorname{span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right)$$

Finally,

$$G(1, T) = \ker((T - I)^3) = \ker \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \operatorname{span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right)$$

Of course, if we only wanted $G(1, T)$, we could have computed it as soon as we learned that $\lambda = 1$ was an eigenvalue.

Theorem 21.6
Let $T \in \mathcal{L}(V)$ and $\dim V = n$. Suppose that $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T and that $v_1, \ldots, v_m$ are corresponding generalized eigenvectors. Then, $v_1, \ldots, v_m$ are linearly independent.

Proof: Suppose that $a_1, \ldots, a_m$ are scalars such that $0 = a_1 v_1 + \cdots + a_m v_m$. Let k be the max j such that $(T - \lambda_1 I)^j v_1 \neq 0$ and let $w = (T - \lambda_1 I)^k v_1$. Then, $(T - \lambda_1 I) w = (T - \lambda_1 I)^{k+1} v_1 = 0$ by the definition of k. Thus, $Tw = \lambda_1 w$ and $w \neq 0$. Since $(T - \lambda I) w = (\lambda_1 - \lambda) w$ for all $\lambda \in \mathbb{F}$, then $(T - \lambda I)^n w = (\lambda_1 - \lambda)^n w$ for all $\lambda \in \mathbb{F}$ too.
Applying $(T - \lambda_1 I)^k (T - \lambda_2 I)^n \ldots (T - \lambda_m)^n$ to both sides of

$$\begin{aligned} 0 &= a_1 v_1 + \cdots + a_m v_m \\ &= a_1 (T - \lambda_1 I)^k (T - \lambda_2 I)^n \ldots (T - \lambda_m)^n v_1 \\ &= a_1 (T - \lambda_2 I)^n \ldots (T - \lambda_m)^n w \end{aligned}$$

All $(T - \lambda_j I)^n$s commute with each other, so all $a_j v_j$ are erased except $j = 1$. However, then

$$\begin{aligned} 0 &= a_1 (T - \lambda_2 I)^n \ldots (T - \lambda_m)^n w \\ &= a_1 (\lambda_1 - \lambda_2)^n \ldots (\lambda_1 - \lambda_m)^n w \end{aligned}$$

Since all $\lambda_1, \ldots, \lambda_m$ are distinct, $a_1 = 0$. Similarly, by repeating the same process by considering for each j, we can set each $a_j = 0$ for all $j \leq m$. Therefore, $v_1, \ldots, v_m$ are linearly independent as expected.

21.2 Nilpotent Operators

Definition 21.3: Nilpotent
$T \in \mathcal{L}(V)$ is nilpotent if $T^k = 0$ for some $k \geq 0$.

Example 21.2
Let $D : \mathcal{P}_n \mapsto \mathcal{P}_n$ be the differentiation operator. This is nilpotent as the $n+1$th derivative of any $f \in \mathcal{P}_n$ is the zero polynomial, i.e., $D^{n+1} = 0$.

Example 21.3
As we saw in a previous example, $[T]_e^e = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ is nilpotent since $T^3 = 0$.

Theorem 21.7
If $\dim V = n$ and $T \in \mathcal{L}(V)$ is nilpotent, then $T^n = 0$.

Proof: The chain $\{0\} = \ker(T^0) \subset \ker(T) \subset \cdots \subset \ker(T^k) = \ker(T^{k+1}) = \ldots$ stabilizes at the latest when $k = n$. Hence, $T^j = 0$ for some j implies that $T^k = 0$ for all $k \geq n$.

Theorem 21.8
If $T \in \mathcal{L}(V)$ is nilpotent, then there is a basis β of V such that

$$[T]_\beta^\beta = \begin{bmatrix} 0 & a_{12} & a_{13} & \ldots & a_{1n} \\ 0 & 0 & a_{23} & \ldots & a_{2n} \\ 0 & 0 & 0 & \ldots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 \end{bmatrix}$$

i.e., $[T]_\beta^\beta$ is upper triangular with all diagonal entries 0.

Proof: Choose a basis of $\ker(T)$. Extend that to a basis of $\ker(T^2)$. Extend that to a basis of $\ker(T^3)$. Keep extending and we eventually just get the basis of $\ker(T^n) = V$. We claim that the matrix representation of T with respect to this basis has the required form.

The first $\dim\ker(T) \geq 1$ columns are filled with 0s since those are all members of $\ker(T)$. Now, consider a basis element $v \in \ker(T^2)$ such that $v \notin \ker(T)$. Then, $Tv \in \ker(T)$ and Tv is a linear combination of the basis elements of $\ker(T)$. That is, the entries of that corresponding column will have non-zero numbers above the main diagonal.

Similarly, basis elements coming from $\ker(T^3)$ but not $\ker(T^2)$ are such that applying T to them will give us elements in $\ker(T^2)$ and so on.

Example 21.4: 8A Exercise 5
Let $T \in \mathcal{L}(V)$, $m > 0$, $v \in V$ with $T^{m-1}v \neq 0$ but $T^m v = 0$. Prove that $v, Tv, \ldots, T^{m-1}v$ is linearly independent.

Proof: Suppose that $c_0 v + c_1 Tv + c_2 T^2 v + \cdots + c_{m-1}T^{m-1}v = 0$. Apply T^{m-1} to both sides to get $c_0 T^{m-1}v = 0$. But $T^{m-1}v \neq 0$ so $c_0 = 0$. Thus, $c_1 Tv + c_2 T^2 v + \cdots + c_{m-1}T^{m-1}v = 0$. Apply T^{m-2} to both sides to get $c_1 T^{m-1} v = 0$. But $T^{m-1}v \neq 0$ so $c_1 = 0$. Repeat this process to get $c_0 = c_1 = \cdots = c_{m-1} = 0$, making the list $v, Tv, T^2v, \ldots, T^{m-1}v$ linearly independent.

Example 21.5: 8A Exercise 6
Let $T \in \mathcal{L}(\mathbb{C}^3)$ be defined by $T(z_1, z_2, z_3) = (z_2, z_3, 0)$. Prove that T has no square root S, i.e., that there does not exist $S \in \mathcal{L}(\mathbb{C}^3)$ such that $S^2 = T$.

Proof:

$$[T]_e^e = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Note that $T^3 = 0$. So, if $S^2 = T$, then $S^6 = 0$. Then, $S^3 = 0$. As $S^2 = T \neq 0$, we must have the chain $\ker(S) \subset \ker(S^2) \subset \ker(S^3) = \mathbb{C}^3$. This is only possible if $\dim\ker(S) = 1$ and $\dim\ker(S^2) = 2$. But, $S^2 = T$ has rank 2, contradicting $\dim\ker(S^2) = 2$. So, $S^2 = T$ is impossible.

Example 21.6: 8A Exercise 13
Let V be an inner product space and $N \in \mathcal{L}(V)$ a normal and nilpotent operator. Prove that $N = 0$.

Proof: Since N is nilpotent, there is a basis β of V such that $[N]_\beta^\beta$ is upper triangular with all 0s along the main diagonal.

Apply Gram Schmidt to β to obtain an orthonormal basis α of V for which $[N]_\alpha^\alpha$ still has the same form. However, now the same argument as the proof of the complex spectral theorem shows that $[N]_\alpha^\alpha$ has off-diagonal elements that are 0. So, every element of N is 0, i.e., $N = 0$.

22 Lecture 22

22.1 Block Diagonal Matrices

Theorem 22.1
Let $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbb{F})$. Then, $\ker(p(T))$ and $\operatorname{range}(p(T))$ are both T-invariant.

Proof: We will prove both parts:

- Suppose $p(T)v = 0$, i.e., $0 \in \ker(p(T))$. Then, $(p(T))(Tv) = T(p(T)v) = T(0) = 0$. So, $Tv \in \ker(p(T))$ and $\ker(p(T))$ is T-invariant.
- Suppose $v \in \operatorname{range}(p(T))$, i.e., there is a $u \in V$ such that $v = p(T)u$. Then, $Tv = T(p(T)u) = p(T)(Tu)$. So, $Tv \in \operatorname{range}(p(T))$ and $\operatorname{range}(p(T))$ is T-invariant.

Theorem 22.2
Suppose V is a complex finite dimensional vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. Then,

1. $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$
2. Each $G(\lambda_j, T)$ is T-invariant
3. Each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent.

Proof: We will prove all three conclusions:

- If $\dim V = n$, then $G(\lambda_j, T) = \ker((T - \lambda_j I)^n)$. So, statement 2 follows from the first theorem by letting $p(z) = (z - \lambda_j)^n$.
- If $v \in G(\lambda_j, T)$, then $(T - \lambda_j I)^n v = 0$. Thus, $(T - \lambda_j I)^n|_{G(\lambda_j, T)} = 0$, making $(T - \lambda_j T)|_{G(\lambda_j, T)}$ nilpotent so statement 3 also holds.
- Statement 1 can be proven by induction on $n = \dim V$. A basis with $n = 1$ is trivial, so assume the result holds for all vector spaces of dimension less than n. Since V is a complex vector space, T has an eigenvalue $\lambda_1 \in \mathbb{C}$. Thus, $m \geq 1$. We also have that $V = G(\lambda_1, T) \oplus \operatorname{range}(T - \lambda_1 T)^n$ by theorem 21.4. Call the second term U.

  Note that U is T-invariant by the theorem above (consider $p(z) = (z - \lambda_1)^2$). Since $\dim G(\lambda_1, T) \geq 1$ and $\dim U < n$, our induction hypothesis applies to $T|_U$. None of the generalized eigenvectors of $T|_U$ correspond to eigenvalue λ_1 since a generalized λ_1-eigenvector of $T|_U$ would also be in $G(\lambda_1, T) \cap \{0\}$. Thus, each eigenvalue of $T|_U$ is in the list $\lambda_2, \ldots, \lambda_m$.

  Therefore, the induction hypothesis can be applied to the subspace U to yield $U = G(\lambda_2, T|_U) \oplus \cdots \oplus G(\lambda_m, T|_U)$. So, we basically need to show that $G(\lambda_j, T|_U) = G(\lambda_j, T)$ for each $j = 2, \ldots, m$. Observe that $G(\lambda_j, T|_U) \subseteq G(\lambda_j, T)$ is trivial. We want to show that $G(\lambda_j, T) \subseteq G(\lambda_j, T|_U)$ now.

  Suppose $v \in G(\lambda_j, T)$. Since $V = G(\lambda_1, T) \oplus U$, each $v \in V$ can be written as $v_1 + u$ for some $v_1 \in G(\lambda_1, T)$ and $u \in U$. However, as $U = G(\lambda_2, T|_U) \oplus \cdots \oplus G(\lambda_m, T|_U)$, note that $u = v_2 + \cdots + v_m$ where $v_l \in G(\lambda_l, T|_U) \subseteq G(\lambda_l, T)$. Therefore, $v = v_1 + v_2 + \cdots + v_m \in G(\lambda_j, T)$. As generalized eigenvectors corresponding to distinct eigenvalues are linearly independent, all of the v_is above are 0 except for possibly v_j (in which case, we have $v = v_j$ here). In particular, $v_1 = 0$ so v must be in U. Thus, $v \in G(\lambda_j, T|_U)$ as well.

Theorem 22.3
Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then, there is a basis of V consisting of the generalized eigenvectors of T.

Proof: Let $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$. Choose a basis of each $G(\lambda_j, T)$ and collate to obtain a basis of V consisting of the generalized eigenvectors of T.

Definition 22.1: Algebraic Multiplicity

Let $T \in \mathcal{L}(V)$ and $n = \dim V$. The multiplicity of an eigenvalue λ of T is $\dim G(\lambda, T)$, i.e., $\dim \ker(T - \lambda I)^n$.

Example 22.1

Let $T : \mathbb{R}^5 \mapsto \mathbb{R}^5$ be defined such that

$$[T]_e^e = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 6 & 1 \\ 0 & 0 & 0 & 0 & 6 \end{bmatrix}$$

Then, $\dim E(2, T) = 1$, $\dim E(6, T) = 1$, $\dim G(2, T) = 3$ and $\dim G(6, T) = 2$.

Theorem 22.4

If V is a complex vector space and $T \in \mathcal{L}(V)$, then the sum of the algebraic multiplicities of T equals $\dim V$.

Proof: Immediately follows from $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T) \implies \dim V = \dim G(\lambda_1, T) + \cdots + \dim G(\lambda_m, T)$.

Definition 22.2: Block Diagonal Matrices

A square matrix A is block diagonal if it has the form

$$A = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_m \end{bmatrix}$$

where $A_1, \ldots, A_m$ are each square matrices lying along the diagonal and all other entries equal 0.

Example 22.2

$$A = \begin{bmatrix} 4 & & & & \\ & 2 & 6 & & \\ & 0 & 1 & & \\ & & & 1 & 2 \\ & & & 8 & 4 \end{bmatrix}$$ is a block diagonal matrix (the empty space, like always, is assumed to be filled with 0s).

Theorem 22.5

Let V be a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the the distinct eigenvalues of T with multiplicities $d_1, \ldots, d_m$. Then, there is a basis of V with respect to which T has a block diagonal matrix of the form

$$\begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_m \end{bmatrix}$$

where each A_j is a $d_j \times d_j$ upper triangular matrix of the form

$$A_j = \begin{bmatrix} \lambda_j & \ldots & * \\ & \ddots & \vdots \\ & & \lambda_j \end{bmatrix}$$

Proof: Let $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$ where each $G(\lambda_j, T)$ is T-invariant. Observe that

$$T|_{G(\lambda_j,T)} = (\lambda_j I + T - \lambda_j I)|_{G(\lambda_j,T)} = \lambda_j I|_{G(\lambda_j,T)} + (T - \lambda_j I)|_{G(\lambda_j,T)}$$

where $(T - \lambda_j T)|_{G(\lambda_j,T)}$ is nilpotent following theorem 22.2. Thus, there is a basis β_j of $G(\lambda_j, T)$ with respect to which the matrix representation of $(T - \lambda_j T)|_{G(\lambda_j,T)}$ is an upper triangular nilpotent matrix. Then,

$$\left[T|_{G(\lambda_j,T)}\right]_{\beta_j}^{\beta_j} = \left[(T - \lambda_j, T)|_{G(\lambda_j,T)} + \lambda_j I|_{G(\lambda_j,T)}\right]_{\beta_j}^{\beta_j} = \begin{bmatrix} 0 & \dots & * \\ & \ddots & \vdots \\ & & 0 \end{bmatrix} + \begin{bmatrix} \lambda_j & \dots & * \\ & \ddots & \vdots \\ & & \lambda_j \end{bmatrix} = \begin{bmatrix} \lambda_j & \dots & * \\ & \ddots & \vdots \\ & & \lambda_j \end{bmatrix}$$

Now, collate the bases $\beta_1, \beta_2, \dots, \beta_m$ to get the desired basis of V.

Theorem 22.6
If $N \in \mathcal{L}(V)$ is nilpotent, then $I + N$ is invertible.

Proof: We will give two proofs of this statement:

1. What are the eigenvalues of $I + N$? If λ is an one and $v \neq 0$ an eigenvector, then $(I + N)v = \lambda$. However, for such λ and v,

$$v + Nv = (I + N)v = \lambda v \implies Nv = (\lambda - 1)v$$

Thus, the eigenvalues of $I + N$ are the eigenvalues of N shifted by 1. However, N only has eigenvalue 0. Thus, 1 is the only eigenvalue of $I + N$, so $I + N$ is invertible.

2. Recall the formula for an infinite geometric series:

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \dots$$

for $|x| < 1$. Then,

$$\frac{I}{I+N} = I - N + N^2 - N^3 + \dots$$

However, as N is nilpotent, the series converges (terminates since $N^{\dim V} = 0$). Thus, $(I + N)^{-1} = I - N + N^2 - \dots$. Actually, for a "Banach algebra" like $M_n(\mathbb{C})$, if $\|A\| < 1$, then $I - A$ is invertible and $(I - A)^{-1} = I + A + A^2 + A^3 + \dots$ "converges." (see any book on Banach algebras for a detailed proof).

Theorem 22.7
If $N \in \mathcal{L}(V)$ is nilpotent, then $I + N$ has a square root, i.e., there is an $S \in \mathcal{L}(V)$ such that $I + N = S^2$.

Proof: Look at the Taylor series to $\sqrt{1+x}$ (about $x = 0$). Then,

$$\sqrt{1+x} = 1 + f'(0)x + \frac{f''(0)}{2}x^2 + \dots$$

where $f(x) = \sqrt{1+x}$. Then,

$$\sqrt{I+N} = I + f'(0)N + \frac{f''(0)}{2}N^2 + \dots$$

Since N is nilpotent, this is a finite series (all terms beyond the $\dim V$st are zero).

Theorem 22.8
If V is complex vector space and $T \in \mathcal{L}(V)$ is invertible, then T has a square root.

Proof: Look at

$$T|_{G(\lambda_j,T)} = \underbrace{\lambda_j I|_{G(\lambda_j,T)}}_{\neq 0} + \underbrace{(T - \lambda_j I)|_{G(\lambda_j,T)}}_{\text{nilpotent}}$$

This has a square root, when restricted to $G(\lambda_j, T)$, by a scaling modification to the identity matrix in the theorem above. You can then assemble all the individual square roots (for each generalized eigenspace) to get a global square root.

23 Lecture 23

23.1 Characteristic Polynomial

Definition 23.1: Characteristic Polynomial
Let V be a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T with multiplicities $d_1, \ldots, d_m$. The polynomial $c_T(z) = (z - \lambda_1)^{d_1} \ldots (z - \lambda_m)^{d_m}$ is called the characteristic polynomial of T.

Note 23.1
Some things to note:

1. The characteristic polynomial of T has degree $d_1 + \cdots + d_m = \dim V$.
2. The roots of the characteristic polynomial are the eigenvalues of T

Theorem 23.1: Cayley-Hamilton Theorem
Let V be a complex vector space and $T \in \mathcal{L}(V)$. Then, $c_T(T) = 0$.

Proof: Note that $c_T(T) = (T - \lambda_1 I)^{d_1} \ldots (T - \lambda_m I)^{d_m}$. Write $v = c_1 v_1 + \cdots + c_m v_m$ where $v_j \in G(\lambda_j, T)$. As $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent and $d_j = \dim G(\lambda_j, T)$, then $(T - \lambda_j I)^{d_j}|_{G(\lambda_j, T)} = 0|_{G(\lambda_j, T)}$. Since the $(T - \lambda_j I)^{d_j}$s commute, we can see that $c_T(T)|_{G(\lambda_j, T)} v_j = 0$. This is true for all $j = 1, \ldots, m$ and V is partitioned into multiple generalized eigenspaces so $c_T(T)v = 0$ for all $v \in V$. In other words, $c_T(T)$ is the zero operator on V.

Note 23.2
Beware of the following invalid "proof" of the Cayley-Hamilton Theorem: $c_T(z) = \det(zI - T)$ is a polynomial in z. Thus, $c_T(T) = \det(TI - T) = \det(0) = 0$.

What's the flaw here? While it is true that $\det(zI - T) = c_T(z) = (z - \lambda_1)^{d_1} \ldots (z - \lambda_m)^{d_m}$ is a polynomial in z with coefficients in $\mathbb{C}$, the expression $zI - T$ occurring in $\det(zI - T)$ is *not* a polynomial in z with coefficients in $\mathbb{C}$. So, what does it even mean to plug in T for z in $zI - T$? It is $\det(zI - T)$ that is a polynomial z, not $zI - T$, so any valid proof has to go through plugging in T for z in $(z - \lambda_1)^{d_1} \ldots (z - \lambda_m)^{d_m}$.

Definition 23.2: Monic Polynomial
A monic polynomial is a polynomial whose highest degree coefficient equals 1.

Example 23.1
Note that $p(z) = z^{11} + 6z^3 + 3$ is a monic polynomial of degree 11.

Definition 23.3: Minimal Polynomial
Suppose $T \in \mathcal{L}(V)$. There is a unique monic polynomial p of the smallest degree such that $p(T) = 0$. This is called the minimal polynomial and is denoted by $m_T(z)$.

Proof: We will establish the existence, followed by the uniqueness, of the minimal polynomial:

- Let $n = \dim V$. The list $I, T, T^2, \ldots, T^{n^2}$ is not linearly independent in $\mathcal{L}(V)$ as it is a list of length $n^2 + 1$ and $\dim \mathcal{L}(V) = n^2$. By the Linear Dependence Lemma, one of the operators in the list can be written as a linear combination of the preceding ones, say T^m, such that $a_0 I + a_1 T + \cdots + a_{m-1} T^{m-1} + T^m = 0$ for some scalars $a_1, \ldots, a_{m-1} \in \mathbb{F}$. Define $p(z) = a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} + z^m$. Then, p is monic and $p(T) = 0$.
- The choice of m (from the LDL) implies no monic polynomial q of degree smaller than m can satisfy $q(T) = 0$. Suppose q is a monic polynomial of degree m and $q(T) = 0$. Then, $p(T) = 0 = q(T)$ so $(p - q)(T) = 0$ and $\deg(p - q) < m$.

Dividing $p - q$ by its leading coefficient would result in a monic polynomial with degree $< m$, yet $(p - q)(T) = 0$. However, that is a contradiction unless $p - q$ is the zero polynomial in the first place.

Theorem 23.2
Suppose $T \in \mathcal{L}(V)$ and $q \in \mathcal{P}(\mathbb{F})$. Then, $q(T) = 0$ iff q is a polynomial multiple of the minimal polynomial of T.

Proof: Let $p = m_T(z)$ denote the minimal polynomial of T. Suppose q is a polynomial multiple of m_T, i.e., there is a polynomial $s \in \mathcal{P}(\mathbb{F})$ such that $q = m_T s$. Then, $q(T) = m_T(T)s(T) = 0 \cdot s(T) = 0$.

For the converse, suppose that $q(T) = 0$. By the definition of the division algorithm for polynomials, there are polynomials $s, r \in \mathcal{P}(\mathbb{F})$ such that $q = ps + r$ with degree $\deg(r) < \deg(p)$ and $p = m_T$. Thus,

$$0 = q(T) = p(T)s(T) + r(T) = r(T)$$

But, r must be the zero polynomial, as otherwise we could divide it by its leading coefficient to obtain a monic polynomial that kills T and has a degree less than that of p. That would again be a contradiction so $r = 0$ and $q = ps$.

Theorem 23.3
If V is a complex vector space and $T \in \mathcal{L}(V)$, then the characteristic polynomial $c_T(z)$ is a polynomial multiple of the minimal polynomial $m_T(z)$ of T.

Proof: It follows directly from the theorem above.

Theorem 23.4
Let $T \in \mathcal{L}(V)$. The roots of the minimal polynomial of T are the eigenvalues of T.

Proof: Let $p(z) = m_T(z) = a_0 + a_1 z + \cdots + a_{m-1}z^{m-1} + z^m$ and let $\lambda \in \mathbb{F}$ be a zero of p. Then, $p(z) = (z - \lambda)q(z)$ where q is a monic polynomial in $\mathcal{P}(\mathbb{F})$. Note that $p(T) = m_T(T) = 0 \implies (T - \lambda I)q(T)v = 0$ for all $v \in V$. However, as $\deg(q) < \deg(p)$ and $q(T) \neq 0$, there is a $v \in V$ with $v \neq 0$ such that $q(T)v \neq 0$. Thus, λ is an eigenvalue and $q(T)v$ is a λ-eigenvector of T.

To prove the other direction, suppose λ is an eigenvalue of T. We need to show that it is a zero of the minimal polynomial. If v is a λ-eigenvector of T, then v is a $p(\lambda)$-eigenvector of $p(T)$ for any polynomial p, including the minimal polynomial. That is, we have $p(T)v = p(\lambda)v$. However, $p(T) = 0$. So, $0 = p(\lambda)v$. Since, $v \neq 0$, this implies that $p(\lambda) = 0$ and λ is a zero of p.

Example 23.2: 8C Exercise 1
Let $T \in \mathcal{L}(\mathbb{C}^4)$ such that its eigenvalues are $3, 5, 8$. Prove that $(T - 3I)^2(T - 5I)^2(T - 8I)^2 = 0$.

Proof: We know that $\mathbb{C}^4 = G(3,T) \oplus G(5,T) \oplus G(8,T)$. As $1 \leq \dim G(\lambda, T) \leq 2$ by the pigeonhole principle, for all three eigenvalues, $(T - \lambda I)^2|_{G(\lambda,T)} = 0|_{G(\lambda,T)}$. Thus, $(T - 3I)^2(T - 5I)^2(T - 8I)^2 = 0$.

Example 23.3: 8C Exercise 5
Give an example of an operator on $\mathbb{C}^4$ whose characteristic polynomial and minimal polynomial both equal $z(z-1)^2(z-3)$. Consider

$$[T]_e^e = \begin{bmatrix} 0 & & & \\ & 1 & 1 & \\ & 0 & 1 & \\ & & & 3 \end{bmatrix}$$

Example 23.4: 8C Exercise 6
Give an example of an operator on $\mathbb{C}^4$ whose characteristic polynomial equals $z(z - 1)^2(z - 3)$ and whose minimal

polynomial equals $z(z-1)(z-3)$. Consider

$$[T]_e^e = \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 3 \end{bmatrix}$$

Why does this work? Note that

$$\begin{aligned} \mathbb{C}^4 &= E(0,T) \oplus E(1,T) \oplus E(3,T) \\ &= G(0,T) \oplus G(1,T) \oplus G(3,T) \end{aligned}$$

Thus, $c_T(z) = z(z-1)^2(z-3)$.

Let $\operatorname{span}(v_1) = E(0,T)$, $\operatorname{span}(v_2, v_3) = E(1,T)$ and $\operatorname{span}(v_4) = E(3,T)$. Then, v_1, v_2, v_3, v_4 is a basis of $\mathbb{C}^4$. If $v = c_1v_1 + c_2v_2 + \cdots + c_3v_3 + c_4v_4$, then $T(T-I)(T-3I)v = 0$ as T kills v_1, $T - I$ kills v_2 and v_3, and $T - 3I$ kills v_4 — all of those operators commute. As each eigenvalue $0, 1, 3$ of T must be a zero of m_T, the polynomial $z(z-1)(z-3)$ must divide $m_T(z)$. However, that in itself is already monic so $m_T(z) = z(z-1)(z-3)$.

Example 23.5: 8C Exercise 8

Let $T \in \mathcal{L}(V)$. Prove T is invertible iff the constant term of $m_T(z)$ is nonzero.

Proof: If $p(z)$ is any polynomial, then $p(0)$ is the constant term of p. Thus, 0 is a zero of $m_T(z)$ iff the constant of $m_T(z)$ is 0. However, the zeroes of $m_T(z)$ are precisely the eigenvalues of T. Thus, T is invertible iff the constant term of $m_T(z)$ is nonzero.

Example 23.6

Characteristic polynomials can also aid in finding, say, the SVD of a matrix. For example, let $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$. Then,

$$\begin{aligned} A^TA &= \begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix} \\ \det\left(A^TA - \lambda I\right) &= (1-\lambda)(5-\lambda) - 1 \\ 0 &= \lambda^2 - 6\lambda + 4 \end{aligned}$$

Thus, $\lambda = \frac{6 \pm \sqrt{20}}{2} = 3 \pm \sqrt{5}$. So, $\sigma_1, \sigma_2 = \sqrt{3+\sqrt{5}}, \sqrt{3-\sqrt{5}}$. Also,

$$E(3+\sqrt{5}, A^TA) = \ker \begin{bmatrix} -2-\sqrt{5} & 1 \\ 1 & 2-\sqrt{5} \end{bmatrix} = \operatorname{span}\left(\begin{bmatrix} -2+\sqrt{5} \\ 1 \end{bmatrix}\right) = \operatorname{span}\left(\begin{bmatrix} \frac{-2+\sqrt{5}}{\sqrt{10-4\sqrt{5}}} \\ \frac{1}{\sqrt{10-4\sqrt{5}}} \end{bmatrix}\right) = \operatorname{span}(v_1)$$

$$E(3-\sqrt{5}, A^TA) = \ker \begin{bmatrix} -2+\sqrt{5} & 1 \\ 1 & 2+\sqrt{5} \end{bmatrix} = \operatorname{span}\left(\begin{bmatrix} -2-\sqrt{5} \\ 1 \end{bmatrix}\right) = \operatorname{span}\left(\begin{bmatrix} \frac{-2-\sqrt{5}}{\sqrt{10+4\sqrt{5}}} \\ \frac{1}{\sqrt{10+4\sqrt{5}}} \end{bmatrix}\right) = \operatorname{span}(v_2)$$

Then,

$$Av_1 = \begin{bmatrix} \frac{-1+\sqrt{5}}{\sqrt{10-4\sqrt{5}}} \\ \frac{2}{\sqrt{10-4\sqrt{5}}} \end{bmatrix}$$

$$u_1 = \frac{Av_1}{\sigma_1} = \begin{bmatrix} \frac{-1+\sqrt{5}}{\sqrt{10-2\sqrt{5}}} \\ \frac{2}{\sqrt{10-2\sqrt{5}}} \end{bmatrix}$$

$$Av_2 = \begin{bmatrix} \frac{-1-\sqrt{5}}{\sqrt{10+4\sqrt{5}}} \\ \frac{2}{\sqrt{10+4\sqrt{5}}} \end{bmatrix}$$

$$u_2 = \frac{Av_2}{\sigma_2} = \begin{bmatrix} \frac{-1-\sqrt{5}}{\sqrt{10+2\sqrt{5}}} \\ \frac{2}{\sqrt{10+2\sqrt{5}}} \end{bmatrix}$$

Finally,

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} \frac{-1+\sqrt{5}}{\sqrt{10-2\sqrt{5}}} & \frac{-1-\sqrt{5}}{\sqrt{10-2\sqrt{5}}} \\ \frac{2}{\sqrt{10-2\sqrt{5}}} & \frac{2}{\sqrt{10+2\sqrt{5}}} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sqrt{3+\sqrt{5}} & 0 \\ 0 & \sqrt{3-\sqrt{5}} \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \frac{-2+\sqrt{5}}{\sqrt{10-4\sqrt{5}}} & \frac{1}{\sqrt{10-4\sqrt{5}}} \\ \frac{-2-\sqrt{5}}{\sqrt{10+4\sqrt{5}}} & \frac{1}{\sqrt{10+4\sqrt{5}}} \end{bmatrix}}_{V^T}$$

How does it compare to the diagonalization of A? Since A is upper triangular, its eigenvalues are 1 and 2. Then,

$$E(2,T) = \operatorname{span}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \operatorname{span}\left(\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}\right)$$
$$E(1,T) = \operatorname{span}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$$

Thus,

$$\begin{aligned} \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} &= \begin{bmatrix} \frac{1}{\sqrt{2}} & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & \sqrt{2} \\ 1 & -1 \end{bmatrix} \end{aligned}$$

In $A = PDP^{-1}$ and $A = U\Sigma V^T$, the biggest difference between the two is that U and V are orthogonal matrices while P isn't. Furthermore, there is no direct relationship between the eigenvalues and singular values of A.

24 Lecture 24

24.1 Jordan Forms

Consider the following two examples as a preview of what is to come later in the lecture!

Example 24.1
Let $V = \mathbb{F}^4$ and $N \in \mathcal{L}(V)$ such that $N(z_1, z_2, z_3, z_4) = (0, z_1, z_2, z_3)$. In other words,

$$[N]_e^e = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Let $v = e_1 = (1, 0, 0, 0)$ and choose the basis $\beta = N^3v, N^2v, Nv, v = e_4, e_3, e_2, e_1$. Then,

$$[N]_\beta^\beta = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Example 24.2
Let $V = \mathbb{F}^6$ and $N = \mathcal{L}(V)$ such that $N(z_1, z_2, z_3, z_4, z_5, z_6) = (0, z_1, z_2, 0, z_4, 0)$. In other words,

$$[N]_e^e = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Let $v_1 = e_1$, $v_2 = e_4$, $v_3 = e_6$ and define the basis $\beta = N^2v_1, Nv_1, v_1, Nv_2, v_2, v_3$. Then,

$$[N]_\beta^\beta = \begin{bmatrix} 0 & 1 & 0 & & & \\ 0 & 0 & 1 & & & \\ 0 & 0 & 0 & & & \\ & & & 0 & 1 & \\ & & & 0 & 0 & \\ & & & & & 0 \end{bmatrix}$$

Theorem 24.1
Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then, there are vectors $v_1, \dots, v_n \in V$ and nonnegative integers $m_1, \dots, m_n$ such that

1. $N^{m_1}v_1, \dots, Nv_1, v_1, \dots, N^{m_n}v_n, \dots, Nv_n, v_n$ is a basis of V
2. $N^{m_1+1}v_1 = \cdots = N^{m_n+1}v_n = 0$

Proof: We will prove this theorem by induction on $\dim V$. The $\dim V = 1$ case is trivial as N is the 0 operator. The induction hypothesis states that the desired result holds on all vector spaces of dimension less than $\dim V$.

Since N is nilpotent, N is not injective and, hence, not surjective. Then, either N is the 0 operator or we can apply the induction hypothesis to $N|_{\text{range}(N)} \in \mathcal{L}(\text{range}(N))$ so that there are vectors $v_1, \dots, v_n \in \text{range}(N)$ and nonnegative integers $m_1, \dots, m_n$ such that

$$N^{m_1}v_1, \dots, Nv_1, v_1, \dots, N^{m_n}v_n, \dots, Nv_n, v_n$$

is a basis of $\mathrm{range}(N)$ and

$$N^{m_1+1}v_1 = \cdots = N^{m_n+1}v_n = 0$$

As each $v_j \in \mathrm{range}(N)$, for each j, there is a $u_j \in V$ with $v_j = Nu_j$. Thus, $N^{k+1}u_j = N^k v_j$ for each j and k. We claim that the list

$$N^{m_1+1}u_1, \ldots, Nu_1, u_1, \ldots, N^{m_n+1}u_n, \ldots, Nu_n, u_n$$

is linearly independent. Suppose a linear combination of these vectors is equal to 0. Apply N to both sides. This results in a linear combination of $N^{m_1}v_1, \ldots, Nv_1, v_1, \ldots, N^{m_n}v_n, \ldots, Nv_n, v_n$ that equals 0 (actually it would also contain the $N^{m_1+1}v_1, \ldots, N^{m_n+1}v_n$ terms but those are all already 0). However, this list is linearly independent because of the induction hypothesis. So, all of the coefficients in the linear combination above must be 0. Then, all coefficients in the linear combination of $N^{m_1+1}u_1, \ldots, Nu_1, u_1, \ldots, N^{m_n+1}u_1$ except for possibly those of $N^{m_1+1}u_1, \ldots, N^{m_n+1}u_n$ are 0 (because $N^{m_1+1}v_1, \ldots, N^{m_n+1}v_n$ are 0 and can essentially be multiplied with any constant). However, $N^{m_1}v_1, \ldots, N^{m_n}v_n$ are linearly independent as well, so those coefficients must be 0 too. Thus, the list of vectors mentioned above will all be linearly independent as claimed.

So, we extend this list to a basis $N^{m_1+1}u_1, N^{m_1}u_1, \ldots, Nu_1, u_1, \ldots, N^{m_n+1}u_n, N^{m_n}u_n, \ldots, Nu_n, u_n, w_1, \ldots, w_p$ of V. Each Nw_j is in $\mathrm{range}(N)$, so they are in the span of $N^{m_1}v_1, \ldots, Nv_1, v_1, \ldots, N^{m_n}v_n, \ldots, Nv_n, v_n$. All of the vectors in that list are a result of applying N to some subset of vectors in $N^{m_1+1}u_1, \ldots, Nu_1, u_1, \ldots, N^{m_n+1}u_n, \ldots, Nu_n, u_n$ so there is an $x_j \in \mathrm{span}(N^{m_1+1}v_1, \ldots, Nv_1, v_1, \ldots, N^{m_n+1}v_n, \ldots, Nv_n, v_n)$ such that $Nw_j = Nx_j$. Let $u_{n+j} = w_j - x_j$. Then, $Nu_{n+j} = 0$ and $N^{m_1+1}u_1, \ldots, Nu_1, u_1, \ldots, N^{m_n+1}u_n, \ldots, Nu_n, u_n, u_{n+1}, \ldots, u_{n+p}$ spans V because it contains each x_j and u_{n+j} (so essentially each w_j). However, we have a spanning list of length equal to the extended basis given above. This is clearly a basis too and it also has the required form.

Definition 24.1: Jordan Basis

Suppose $T \in \mathcal{L}(V)$. A basis of V is called a Jordan basis for T if, with respect to this basis, T has a block diagonal matrix representation

$$\begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_p \end{bmatrix}$$

where each A_j (a "Jordan block") is an upper-triangular matrix of the form

$$\begin{bmatrix} \lambda_j & 1 & & & \\ & \lambda_j & \ddots & & \\ & & \ddots & \ddots & \\ & & & \lambda_j & 1 \\ & & & & \lambda_j \end{bmatrix}$$

There are λ_js on the diagonal and 1s on the "super-diagonal" while everything else is 0.

Theorem 24.2

Let V be a complex vector space. If $T \in \mathcal{L}(V)$, then there is a Jordan basis for T.

Proof: Let $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$ where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T. Then, $N_j = (T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent and each $G(\lambda_j, T)$ has a basis β_j of the type asserted in the theorem above. The matrix representation of a single

Jordan block with respect to such a basis will look like

$$\begin{array}{c|c}
 & \begin{matrix} N_j^{m_j+1} v_j & N_j^{m_j} v_j & \dots & N_j v_j \end{matrix} \\
[N_j]_{\beta_j}^{\beta_j} = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix} & \begin{matrix} N_j^{m_j} v_j \\ \vdots \\ N_j v_j \\ v_j \end{matrix}
\end{array}$$

Thus, $T|_{G(\lambda_j,T)} = \lambda_j I|_{G(\lambda_j,T)} + (T - \lambda_j I)|_{G(\lambda_j,T)}$ has matrix representation given by

$$\begin{bmatrix} \lambda_j & & & & \\ & \lambda_j & & & \\ & & \ddots & & \\ & & & \lambda_j & \\ & & & & \lambda_j \end{bmatrix} + \begin{bmatrix} 0 & 1 & & & \\ & 0 & \ddots & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix} = \begin{bmatrix} \lambda_j & 1 & & & \\ & \lambda_j & \ddots & & \\ & & \ddots & \ddots & \\ & & & \lambda_j & 1 \\ & & & & \lambda_j \end{bmatrix}$$

with respect to this basis β_j. Collate all the $\beta_1, \dots, \beta_m$ to obtain the Jordan basis for T.

Example 24.3: 8D Exercise 1

Consider $N(z_1, z_2, z_3, z_4) = (0, z_1, z_2, z_3)$. Then, as we showed in example 24.1,

$$\begin{array}{cc}
\begin{matrix} N^4 v & N^3 v & N^2 v & N v \end{matrix} & \\
\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \begin{matrix} N^3 v \\ N^2 v \\ N v \\ v \end{matrix}
\end{array}$$

Note that $c_N(z) = z^4$ since $\dim G(0, N) = 4$. Observe that $N^4 v = 0$ but $N^3 v \neq 0$ and $N^3 v, N^2 v, N v, v$ form a basis of $\mathbb{F}^4$. Thus, $N^4 = 0$ and $N^3 \neq 0$, so z^4 is also the minimal polynomial of N.

Example 24.4: 8D Exercise 2

Consider $N = \mathcal{L}(V)$ such that $N(z_1, z_2, z_3, z_4, z_5, z_6) = (0, z_1, z_2, 0, z_4, 0)$ from example 24.2. Let $v_1 = e_1$, $v_2 = e_4$, $v_3 = e_6$ and define the basis $\beta = N^2 v_1, N v_1, v_1, N v_2, v_2, v_3$. Then,

$$[N]_\beta^\beta = \begin{bmatrix} 0 & 1 & 0 & & & \\ 0 & 0 & 1 & & & \\ 0 & 0 & 0 & & & \\ & & & 0 & 1 & \\ & & & 0 & 0 & \\ & & & & & 0 \end{bmatrix}$$

Since $N^3 v_1 = 0$, $N^2 v_2 = 0$ and $N v_3 = 0$, the minimal polynomial of N is z^3. However, the characteristic polynomial of N is z^6 since $\dim G(0, T) = 6$.

Example 24.5: 8D Exercise 3

Let $N \in \mathcal{L}(V)$ be nilpotent. The minimal polynomial of N is z^{m+1} where m is the length of the largest string of consecutive 1s that appear on the super-diagonal in any Jordan form matrix representation of N.

Example 24.6

Suppose T has the matrix representation, with respect to its Jordan basis, given by

$$\begin{bmatrix} 3 & 1 & 0 & & & & & & \\ 0 & 3 & 1 & & & & & & \\ 0 & 0 & 3 & & & & & & \\ & & & 3 & 1 & & & & \\ & & & 0 & 3 & & & & \\ & & & & & 6 & 1 & 0 & 0 \\ & & & & & 0 & 6 & 1 & 0 \\ & & & & & 0 & 0 & 6 & 1 \\ & & & & & 0 & 0 & 0 & 6 \end{bmatrix}$$

Then, T has two Jordan blocks for eigenvalue $\lambda = 3$ and one for eigenvalue $\lambda = 6$. The sum of the sizes of the Jordan blocks for eigenvalue λ is the algebraic multiplicity of λ, i.e., $\dim G(\lambda, T)$. The minimal polynomial of T has a λ root with multiplicity equal to the size of the largest Jordan block for λ. Therefore, the minimal polynomial of T is $(z-3)^3(z-6)^4$ while the characteristic polynomial is $(z-3)^5(z-6)^4$.

Example 24.7

Note, however, that a characteristic polynomial and minimal polynomial are not enough to uniquely determine the Jordan form of T. The following matrix representation has the same characteristic and minimal polynomial as the example above:

$$\begin{bmatrix} 3 & 1 & 0 & & & & & & \\ 0 & 3 & 1 & & & & & & \\ 0 & 0 & 3 & & & & & & \\ & & & 3 & & & & & \\ & & & & 3 & & & & \\ & & & & & 6 & 1 & 0 & 0 \\ & & & & & 0 & 6 & 1 & 0 \\ & & & & & 0 & 0 & 6 & 1 \\ & & & & & 0 & 0 & 0 & 6 \end{bmatrix}$$

Example 24.8: 8D Exercise 4

Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a Jordan basis for T. Describe the matrix with respect to basis $v_n, \ldots, v_1$ obtained by reversing the order of v_js. Let T be such that it has the following as the possible Jordan form matrix representation:

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

Then, reversing the basis will yield

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

Example 24.9: 8D Exercise 5

Let $T \in \mathcal{L}(V)$ and let $v_1, \ldots, v_n$ be Jordan basis for T. What is the matrix representation of T^2 for this basis? Here is what the single block case looks like:

$$\begin{bmatrix}\lambda\end{bmatrix}\begin{bmatrix}\lambda\end{bmatrix} = \begin{bmatrix}\lambda^2\end{bmatrix}$$
$$\begin{bmatrix}\lambda & 1 \\ 0 & \lambda\end{bmatrix}\begin{bmatrix}\lambda & 1 \\ 0 & \lambda\end{bmatrix} = \begin{bmatrix}\lambda^2 & 2\lambda \\ 0 & \lambda^2\end{bmatrix}$$
$$\begin{bmatrix}\lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda\end{bmatrix}\begin{bmatrix}\lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda\end{bmatrix} = \begin{bmatrix}\lambda^2 & 2\lambda & 0 \\ 0 & \lambda^2 & 2\lambda \\ 0 & 0 & \lambda^2\end{bmatrix}$$
$$\begin{bmatrix}\lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda\end{bmatrix}\begin{bmatrix}\lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda\end{bmatrix} = \begin{bmatrix}\lambda^2 & 2\lambda & 0 & 0 \\ 0 & \lambda^2 & 2\lambda & 0 \\ 0 & 0 & \lambda^2 & 2\lambda \\ 0 & 0 & 0 & \lambda^2\end{bmatrix}$$

This pattern continues for Jordan blocks of any size. For the general case, one can find the representation of each Jordan block separately and assemble them at the end.

Example 24.10

Suppose that $T \in \mathcal{L}(V)$ is invertible. Prove that there is a polynomial $p \in \mathcal{P}(\mathbb{F})$ such that $T^{-1} = p(T)$.

Proof: Let $z^m + a_{m-1}z^{m-1} + \cdots + a_1 z + a_0$ be the minimal polynomial of T. So, $T^m + a_{m-1}T^{m-1} + \cdots + a_1 T + a_0 I = 0 \implies a_0 I = -T^m - a_{m-1}T^{m-1} - \cdots - a_1 T$. Since T is invertible, $a_0 \neq 0$ so

$$\begin{aligned} T^{-1} &= a_0 I \frac{T^{-1}}{a_0} \\ &= -\frac{T^{m-1}}{a_0} - \frac{a_{m-1}}{a_0}T^{m-2} - \cdots - \frac{a_1}{a_0} I \end{aligned}$$

25 Lecture 25

25.1 Jordan Forms Cont.

In general, the knowledge of all the eigenvalues of T and the dimensions of $\ker(T - \lambda_j I)^k$ for each eigenvalue λ_j and each $k \geq 1$ is necessary to determine the Jordan form of T.

Theorem 25.1

Let $T \in \mathcal{L}(V)$. If T has a Jordan basis and λ is an eigenvalue of T, then the number of λ eigenvalue Jordan blocks of T of size at least k is exactly $\dim\ker(T - \lambda I)^k - \dim\ker(T - \lambda I)^{k-1}$. Thus, the number of Jordan blocks for the eigenvalue λ of size exactly k is exactly

$$\dim\ker(T - \lambda I)^{k+1} - 2\dim\ker(T - \lambda I)^k + \dim\ker(T - \lambda I)^{k-1}$$

and the number of Jordan blocks of T for λ is the geometric multiplicity $\dim\ker(T - \lambda)$, i.e., $\dim E(\lambda, T)$.

Example 25.1

Let $V = \mathbb{C}^4$ and $T \in \mathcal{L}(\mathbb{C}^4)$. Let λ be the only eigenvalue of T with Jordan basis N^3v, N^2v, Nv, v where $N = T - \lambda I$. The Jordan form of T is

$$\begin{bmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix}$$

The nilpotent operator $T - \lambda I$ has the matrix representation

$$\begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}$$

with respect to the given Jordan basis.

- Since $N^4v = 0$, the list N^3v, N^2v, Nv, v is a basis of $\ker(T - \lambda I)^4 = \ker(T - \lambda I)^5 = \ker(T - \lambda I)^6 = \dots$ and $\dim\ker(T - \lambda I)^4 = 4$.
- Again, as $N^4v = 0$ but $N^3v \neq 0$, we have that N^3v, N^2v, Nv is a basis of $\ker(T - \lambda I)^3$ (since $N^3(Nv) = N^4v = 0$, $N^3(N^2v) = N^5v = 0$ and $N^3(N^3v) = N^6v = 0$) and $\dim\ker(T - \lambda I)^3 = 3$.
- Similarly, N^3v, N^2v is a basis of $\ker(T - \lambda I)^2$ (same logic as before) and $\dim\ker(T - \lambda I)^2 = 2$.
- Finally, N^3v is a basis of $\ker(T - \lambda I) = E(\lambda, T)$ and $\dim\ker(T - \lambda T) = 1$.

The proof is that if we restrict our attention to the sequence of subspaces $\ker(T - \lambda I) \subset \ker(T - \lambda I)^2 \subset \ker(T - \lambda I)^3 \subset \dots$, where T has a single Jordan block for the eigenvalue λ, say of size l, then we must have

$$\begin{aligned} \dim\ker(T - \lambda I)^1 &= 1 \\ \dim\ker(T - \lambda I)^2 &= 2 \\ \dim\ker(T - \lambda I)^3 &= 3 \\ &\vdots \\ \dim\ker(T - \lambda I)^l &= l \\ \dim\ker(T - \lambda I)^{l+1} &= l \\ &\vdots \end{aligned}$$

So, if $k \leq l$, then $\dim\ker(T-\lambda I)^k - \dim\ker(T-\lambda I)^{k-1} = 1$, which is exactly the number of Jordan blocks of T for eigenvalues λ of size at least k. If $k \geq l+1$, then $\dim\ker(T-\lambda I)^k - \dim\ker(T-\lambda I)^{k-1} = k-k = 0$ so that checks out as well.

This establishes the theorem for the case in which T has a single Jordan block for the eigenvalue λ. Suppose, instead, that T has Jordan basis say, $N^3v_1, N^2v_1, Nv_1, v_1, Nv_2, v_2$ where $N = T - \lambda I$. In other words, N has the Jordan form

$$\begin{bmatrix} 0 & 1 & 0 & 0 & & \\ 0 & 0 & 1 & 0 & & \\ 0 & 0 & 0 & 1 & & \\ 0 & 0 & 0 & 0 & & \\ & & & & 0 & 1 \\ & & & & 0 & 0 \end{bmatrix}$$

By the same logic as before ($N^4v_1 = 0$, $N^3v_1 \neq 0$, $N^2v_2 = 0$, $Nv_2 \neq 0$),

$$\begin{aligned} \ker(T-\lambda I)^1 &= \operatorname{span}(N^3v_1, Nv_2) \\ \ker(T-\lambda I)^2 &= \operatorname{span}(N^3v_1, N^2v_1, Nv_2, v_2) \\ \ker(T-\lambda I)^3 &= \operatorname{span}(N^3v_1, N^2v_1, Nv_1, Nv_2, v_2) \\ \ker(T-\lambda I)^4 &= \operatorname{span}(N^3v_1, N^2v_1, Nv_1, v_1, Nv_2, v_2) \\ \ker(T-\lambda I)^5 &= \operatorname{span}(N^3v_1, N^2v_1, Nv_1, v_1, Nv_2, v_2) \\ &\vdots \end{aligned}$$

and the number of Jordan blocks of size at least k for the eigenvalue λ is $\dim\ker(T-\lambda I)^k - \dim\ker(T-\lambda I)^{k-1}$ (which can be confirmed from the matrix above). It isn't hard to turn this into a formal completely general proof of the theorem.

Example 25.2

Suppose $T \in \mathcal{L}(\mathbb{C}^9)$ has eigenvalues 3 and 6 and $\dim(T-3I)^3 = 5$, $\dim(T-3I) = 3$, $\dim(T-6I)^4 = 4$. Is the Jordan form of T completely determined by this information? What are the characteristic and minimal polynomials of T?

As $\dim E(3,T) = \dim\ker(T-3I) = 3$, the operator T has exactly three Jordan blocks for the eigenvalue $\lambda = 3$. The algebraic multiplicity (sum of the size of all Jordan blocks) of $\lambda = 3$ is at least 5 since $\dim\ker(T-3I)^3 = 5$. Similarly, the algebraic multiplicity of $\lambda = 6$ is at least 4. However, $9 = 5+4$ so the algebraic multiplicities of $\lambda = 3$ and $\lambda = 6$ are exactly 5 and 4 respectively. Then, the characteristic polynomial of T is $c_T(z) = (z-3)^5(z-6)^4$.

Thus, we either have one Jordan block of size 3 and two of size 1, or two of size 2 and one of size 1 for $\lambda = 3$. On the other hand, we either have four Jordan blocks of size 1, one of size 2 and two of size 1, two of size 2, one of size 3 and one of size 1, or one of size 4 for $\lambda = 6$. This information is not enough to determine the minimal polynomial of T though $m_T(z)$ is one of the following:

- $(z-3)^3(z-6)^4$
- $(z-3)^3(z-6)^3$
- $(z-3)^3(z-6)^2$
- $(z-3)^3(z-6)$

Example 25.3

Let V be a complex vector space of dimension 8 and $T \in \mathcal{L}(V)$ such that $3, -1$ are its two eigenvalues. Suppose $\dim\ker(T-3I) = 3$, $\dim\ker(T+I) = 2$, $\dim\ker(T-3I)^2 = 4$ and $\dim\ker(T+I)^2 = 4$. What can we say about the Jordan form of T?

Since $\dim\ker(T-3I) = 3$, there are three Jordan blocks for $\lambda = 3$. As $\dim\ker(T-3I)^2 - \dim\ker(T-3I) = 4-3 = 1$, two of the Jordan blocks for $\lambda = 3$ must be 1×1 and the last one should be 2×2 or greater.

Now, as $\dim\ker(T+I) = 2$ and $\dim\ker(T+I)^2 - \dim\ker(T+I) = 4-2 = 2$, there are two Jordan blocks for $\lambda = -1$ and both are at least 2×2.

The only way this is possible is if there are two 1×1 Jordan blocks for $\lambda = 3$, one 2×2 Jordan block for $\lambda = 3$ and two 2×2 Jordan blocks for $\lambda = -1$. Thus, the Jordan form of T (up to some rearrangement of the blocks) is

$$\begin{bmatrix} 3 & 1 & & & & & & \\ 0 & 3 & & & & & & \\ & & 3 & & & & & \\ & & & 3 & & & & \\ & & & & -1 & 1 & & \\ & & & & 0 & -1 & & \\ & & & & & & -1 & 1 \\ & & & & & & 0 & -1 \end{bmatrix}$$

Thus, we have $c_T(z) = (z-3)^4(z+1)^4$ and $m_T(z) = (z-3)^2(z+1)^2$.

Note 25.1
If V is a complex finite-dimensional vector space, the Jordan form is a complete similarity invariant, i.e., if $T, S \in \mathcal{L}(V)$, then there is an isomorphism $U \in \mathcal{L}(V)$ such that $S = U^{-1} \circ T \circ U$ iff T and S have the same Jordan form. At the matrix level, $A, B \in \mathcal{M}_n(\mathbb{C})$ are similar iff A and B have the same Jordan form.

25.2 Complexification

25.2.1 Motivation

Example 25.4
Let $A = \begin{bmatrix} 0.5 & -0.6 \\ 0.75 & 1.1 \end{bmatrix}$. Moreover, let x_0 be an arbitrary point like $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$. Let's analyze the following sequence:

$$\begin{aligned} x_1 &= Ax_0 = \begin{bmatrix} 0.5 & -0.6 \\ 0.75 & 1.1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.0 \\ 1.5 \end{bmatrix} \\ x_2 &= Ax_1 = \begin{bmatrix} 0.5 & -0.6 \\ 0.75 & 1.1 \end{bmatrix} \begin{bmatrix} 1.0 \\ 1.5 \end{bmatrix} = \begin{bmatrix} -0.4 \\ 2.4 \end{bmatrix} \\ x_3 &= Ax_2 = \dots \\ x_4 &= Ax_3 = \dots \\ &\vdots \\ x_n &= Ax_{n-1} = A^n x_0 \end{aligned}$$

The sequence $x_0, x_1, \dots$ lies along an elliptical orbit. Why the rotation? What is going on here?

Definition 25.1: Real and Imaginary Components of a Vector
For any $v \in \mathbb{C}^n$, define

$$\begin{aligned} \mathrm{Re}(v) &= \begin{bmatrix} \mathrm{Re}(v_1) \\ \vdots \\ \mathrm{Re}(v_n) \end{bmatrix} \\ \mathrm{Im}(v) &= \begin{bmatrix} \mathrm{Im}(v_1) \\ \vdots \\ \mathrm{Im}(v_n) \end{bmatrix} \end{aligned}$$

So, $v = \mathrm{Re}(v) + \mathrm{Im}(v)$. We similarly define the conjugate of v as $\overline{v} = \mathrm{Re}(v) - \mathrm{Im}(v)$.

Theorem 25.2
Suppose $A \in \mathcal{M}_2(\mathbb{R})$. View A as a part of $\mathcal{M}_2(\mathbb{C})$. Suppose A has a complex eigenvalue $\lambda = a - ib$ with $b \neq 0$ and the corresponding λ-eigenvector v of A in $\mathbb{C}^2$. Then, $B = S^{-1}AS$ where

$$S = \begin{bmatrix} \mathrm{Re}(v) & \mathrm{Im}(v) \end{bmatrix}$$
$$B = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

for $a, b \in \mathbb{R}$. In other words, the real matrices A and B are similar over $\mathbb{R}$ via the real matrix S.

Proof: Look at AS:

$$\begin{aligned} A\begin{bmatrix} \mathrm{Re}(v) & \mathrm{Im}(v) \end{bmatrix} &= \begin{bmatrix} A\,\mathrm{Re}(v) & A\,\mathrm{Im}(v) \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Re}(Av) & \mathrm{Im}(Av) \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Re}(\lambda v) & \mathrm{Im}(\lambda v) \end{bmatrix} \end{aligned}$$

and compare it to SB:

$$\begin{bmatrix} \mathrm{Re}(v) & \mathrm{Im}(v) \end{bmatrix}\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} a\,\mathrm{Re}(v) + b\,\mathrm{Im}(v) & -b\,\mathrm{Re}(v) + a\,\mathrm{Im}(v) \end{bmatrix}$$

Then,

$$\begin{aligned} \lambda v &= (a - ib)(\mathrm{Re}(v) + i\,\mathrm{Im}(v)) \\ &= \underbrace{a\,\mathrm{Re}(v) + b\,\mathrm{Im}(v)}_{\mathrm{Re}(\lambda v)} + \underbrace{(-b\,\mathrm{Re}(v) + a\,\mathrm{Im}(v))\,i}_{\mathrm{Im}(\lambda v)} \end{aligned}$$

So, $AS = SB$. Since $\mathrm{Re}(v)$ and $\mathrm{Im}(v)$ can easily be shown to be linearly independent via a simple proof by contradiction oriented calculation, it follows that $B = S^{-1}AS$.

What is the point of this? If $A \in \mathcal{M}_2(\mathbb{R})$ has some strictly complex eigenvalues λ when viewed as a matrix in $\mathcal{M}_2(\mathbb{C})$, then A is similar to a matrix of the form given by B (where $\lambda = a - ib$), which is just a scaled rotation matrix. In the example above, the matrix

$$A = \begin{bmatrix} 0.5 & -0.6 \\ 0.75 & 1.1 \end{bmatrix}$$

has a strictly complex eigenvalue $\lambda = a - ib$. This explains why the trajectory of a point/vector under iterates of A will form an elliptical orbit.

If A is already the same form as B, the trajectory would be a circle. However, because the change of basis matrix S takes the standard basis to a non-orthogonal/non-rectangular basis, we uncover the elliptical distortion of the circle. In the example above, this turns out to be

$$\lambda = 0.8 - 0.6i$$
$$v = \begin{bmatrix} -2 - 4i \\ 5 \end{bmatrix}$$
$$S = \begin{bmatrix} -2 & -4 \\ 5 & 0 \end{bmatrix}$$

25.2.2 Complexification

We already worked with the concept of complexification during the proof of the real spectral theorem. We will cover it in more detail now.

Definition 25.2: Complexification
Let V be a real vector space. Then $V_{\mathbb{C}}$ is the complexification of V and has an underlying given by $V \times V = \{(u, v) \in V^2\}$. However, we denote (u, v) by the formal expression $u + iv$. Vector addition in $V_{\mathbb{C}}$ is defined by

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2)$$

and complex scalar multiplication is defined by

$$(a + ib)(u + iv) = (au - bv) + i(av + bu)$$

for $a, b \in \mathbb{R}$ and $u, v \in V$.

Theorem 25.3
Let V be a real vector space. If $v_1, \ldots, v_n$ is a basis of V (as a real vector space), then $v_1, \ldots, v_n$ is a basis of $V_{\mathbb{C}}$ (as a complex vector space). Thus, $\dim_{\mathbb{C}}(V_{\mathbb{C}}) = \dim_{\mathbb{R}}(V)$. It immediately also follows that $V = \mathbb{R}^n \implies V_{\mathbb{C}} \cong \mathbb{C}^n$ so we identify the complexification of $\mathbb{R}^n$ with $\mathbb{C}^n$.

Definition 25.3
Let V be a real vector space and $T \in \mathcal{L}(V)$. Then, $T_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}})$ is defined by $T_{\mathbb{C}}(u + iv) = Tu + iTv$ for all $u, v \in V$.

Definition 25.4
Let V be a real vector space with basis $\beta = v_1, \ldots, v_n$ and $T \in \mathcal{L}(V)$. Then, $[T_{\mathbb{C}}]_{\beta}^{\beta} = [T]_{\beta}^{\beta}$.

Theorem 25.4
Every operator has an invariant subspace of dimension 1 or 2.

Proof: Every operator on a non-zero finite-dimensional complex vector space has an eigenvalue and, thus, a T-invariant 1-dimensional subspace. Now suppose $T \in \mathcal{L}(V)$ where V is real and that $T_{\mathbb{C}}$ has an eigenvalue $\lambda = a + ib$ with eigenvector $u + iv$ (and at least one of u or v nonzero). In other words,

$$\begin{aligned} T_{\mathbb{C}}(u + iv) &= (a + ib)(u + iv) \\ Tu + iTv &= (au - bv) + i(av + bu) \end{aligned}$$

So, $Tu = au - bv$ and $Tv = av + bu$. Let $U = \operatorname{span}(u, v)$. Then, U is a T-invariant subspace of V of dimension 1 (for real eigenvalues) or 2 (for complex eigenvalues).

Theorem 25.5
Let V be a real vector space and $T \in \mathcal{L}(V)$. Then, the minimal polynomial of $T_{\mathbb{C}}$ equals the minimal polynomial of T.

Theorem 25.6
Let V be a real vector space and $T \in \mathcal{L}(V)$ with $\lambda \in \mathbb{R}$. Then, λ is an eigenvalue of $T_{\mathbb{C}}$ iff λ is an eigenvalue of T.

Theorem 25.7
Let V be a real vector space, $T \in \mathcal{L}(V)$ and $j \geq 1, u, v \in V$. Then $(T_{\mathbb{C}} - \lambda I)^j(u + iv) = 0$ iff $(T_{\mathbb{C}} - \overline{\lambda}\, I)^j(u - iv) = 0$. In other words, if w is a λ-eigenvector of $T_{\mathbb{C}}$, then $\overline{w}$ is a $\overline{\lambda}$-eigenvector of $T_{\mathbb{C}}$. So, non-real eigenvalues of $T_{\mathbb{C}}$ come in conjugate pairs.

Theorem 25.8
Let V be a real vector space, $T \in \mathcal{L}(V)$ and λ an eigenvalue of $T_{\mathbb{C}}$. Then, the algebraic multiplicity of λ as an eigenvalue of $T_{\mathbb{C}}$ equals the algebraic multiplicity of $\overline{\lambda}$ as an eigenvalue of $T_{\mathbb{C}}$.

Theorem 25.9
Let V be a real odd-dimensional vector space and $T \in \mathcal{L}(V)$. Then, T has an eigenvalue.

Proof: Complex eigenvalues of $T_{\mathbb{C}}$ come in conjugate pairs with the same algebraic multiplicities. Thus, the sum of the multiplicities of all truly complex eigenvalues is even. So if $\dim V_{\mathbb{C}}$ is odd, then $T_{\mathbb{C}}$ must have a real eigenvalue, which must be a real eigenvalue of T as well.

Theorem 25.10
Let V be a real vector space and $T \in \mathcal{L}(V)$. Then, the coefficients of the characteristic polynomial of $T_{\mathbb{C}}$ are all real.

This justifies the following definition:

Definition 25.5: Characteristic Polynomial
Let $T \in \mathcal{L}(V)$ where V is a real vector space. The characteristic polynomial $c_T(z)$ of T is defined to be the same as the characteristic polynomial $c_{T_{\mathbb{C}}}(z)$ of $T_{\mathbb{C}}$.

Theorem 25.11: Cayley-Hamilton Theorem
Let V be a real vector space and $T \in \mathcal{L}(V)$. If $c_T(z)$ is the characteristic polynomial of T, then $c_T(T) = 0$.

26 Lecture 26

26.1 Real Normal Operators

Last time, we proved the following theorem:

Theorem 26.1
Let $A \in \mathcal{M}_2(\mathbb{R})$. Suppose $A_{\mathbb{C}}$ has a truly complex eigenvalue $\lambda = a + ib$ (where $b \neq 0$) and $v \in \mathbb{C}^2$ is a λ-eigenvector of A. Then, $B = SAS^{-1}$ where

$$B = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$
$$S = \begin{bmatrix} \mathrm{Re}(v) & \mathrm{Im}(v) \end{bmatrix}$$

The proof of this theorem relied on the fact that $\mathrm{Re}(v)$ and $\mathrm{Im}(v)$ are necessarily linearly independent, given the hypothesis. We will prove that fact right now.

Proof: First, write $v = x + iy$ such that $\mathrm{Re}(v) = x$ and $\mathrm{Im}(v) = y$. Note that $y = 0$ is impossible. If it was indeed possible, then $A_{\mathbb{C}}v = Ax$ is real, but λx is not ($y = 0 \implies x \neq 0$ since x is an eigenvector of $A_{\mathbb{C}}$). Now, we claim that $x = 0$ is impossible. If it was indeed possible, then $A_{\mathbb{C}}v = A(iy) = iAy$ which is purely imaginary as A and y are real. However, $A_{\mathbb{C}}(v) = (a + ib)(iy) = -by + aiy$ but $b \neq 0$ so $A_{\mathbb{C}}v$ has a real component. Thus, the real and imaginary parts of v are both nonzero.

We now claim that $v = x + iy$ and $\overline{v} = x - iy$ are linearly independent over $\mathbb{C}$ under the given hypothesis. Suppose $c_1 v + c_2\overline{v} = 0 \implies c_1(x + iy) + c_2(x - iy) = 0 \implies (c_1 + c_2)x + i(c_1 - c_2)y = 0 \implies (c_1 + c_2)x = 0$ and $(c_1 - c_2)y = 0$. However, since $x \neq 0$ and $y \neq 0$, we get the system $c_1 + c_2 = 0$ and $c_1 - c_2 = 0$. These equations imply $c_1 = c_2 = 0$ so v and $\overline{v}$ are linearly independent over $\mathbb{C}$. However, note that

$$x = \frac{v + \overline{v}}{2}$$
$$y = \frac{v - \overline{v}}{2i}$$

Suppose there are constants b_1, b_2 such that $b_1 x + b_2 y = 0$. Then,

$$\begin{aligned} 0 &= b_1 x + b_2 y \\ &= b_1\left(\frac{v + \overline{v}}{2}\right) + b_2\left(\frac{v - \overline{v}}{2i}\right) \\ &= \left(\frac{b_1}{2} + \frac{b_2}{2i}\right)v + \left(\frac{b_1}{2} - \frac{b_2}{2i}\right)\overline{v} \end{aligned}$$

Again, this implies that $b_1 = b_2 = 0$ so $x = \mathrm{Re}(v)$ and $y = \mathrm{Im}(v)$ are linearly independent as expected.

Here is a somewhat faster version of the proof above: under the given hypothesis, v is a λ-eigenvector of $A_{\mathbb{C}}$ and $\overline{v}$ is a $\overline{\lambda}$-eigenvector of $A_{\mathbb{C}}$. As $\lambda \neq \overline{\lambda}$ (the imaginary component is non-zero), then v and $\overline{v}$ are eigenvectors of distinct eigenvalues and, thus, linearly independent over $\mathbb{C}$. The same calculation follows as before, which shows that $\mathrm{Re}(v)$ and $\mathrm{Im}(v)$ are linearly independent as well.

Theorem 26.2
Let V be a two-dimensional real inner product space and $T \in \mathcal{L}(V)$. Then, the following are equivalent:

1. T is normal but not self-adjoint
2. The matrix representation of T with respect to every orthonormal basis of V is

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

with $b \neq 0$.

3. The matrix representation of T with respect to some orthonormal basis of V is

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

with $b \neq 0$.

Proof: Recall that if V is a real inner product space with inner product $\langle \cdot, \cdot \rangle$, then

$$\langle u + iv, x + iy \rangle = \langle u, x \rangle + \langle v, y \rangle + (\langle v, x \rangle - \langle u, y \rangle)i$$

defines a complex inner product on $V_{\mathbb{C}}$. Moreover, if $T \in \mathcal{L}(V)$ is self-adjoint for $\langle \cdot, \cdot \rangle$, then $T_{\mathbb{C}}$ is self-adjoint for $\langle \cdot, \cdot \rangle_{\mathbb{C}}$. If T is normal for $\langle \cdot, \cdot \rangle$, then $T_{\mathbb{C}}$ is normal for $\langle \cdot, \cdot \rangle_{\mathbb{C}}$. Let's prove some statements now:

- statement 1 $\implies$ statement 3

  As $T_{\mathbb{C}}$ is normal but not self-adjoint, by the spectral theorem and the fact that truly complex eigenvalues of $T_{\mathbb{C}}$ come in conjugate pairs, we know that $T_{\mathbb{C}}$ is orthogonally diagonalizable over $\mathbb{C}$ as

  $$\begin{bmatrix} \lambda & \\ & \overline{\lambda} \end{bmatrix}$$

  where $\lambda = a + ib$ and $b \neq 0$ for some $a, b \in \mathbb{R}$. We also know that if V is a λ-eigenvector of $T_{\mathbb{C}}$, then $\overline{v}$ is a $\overline{\lambda}$-eigenvector of $T_{\mathbb{C}}$ and that v and $\overline{v}$ must be orthogonal since T is normal. Since $b \neq 0$, the matrix representation of T with respect to the basis $\mathrm{Im}(v), \mathrm{Re}(v)$ is

  $$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

  Now, we need to only show that $\mathrm{Im}(v)$ and $\mathrm{Re}(v)$ are orthogonal. If $v = x + iy$, then

  $$\langle v, \overline{v} \rangle = \langle x + iy, x - iy \rangle = \|x\|^2 + \|y\|^2 + 2i\,\mathrm{Re}(\langle x, y \rangle) = 0$$

  Since x, y are real, $\mathrm{Re}(\langle x, y \rangle) = \langle x, y \rangle$. The orthogonality of $v, \overline{v}$ above forces $\langle x, y \rangle = 0$. Normalizing x and y, if necessary, won't change the matrix representation of T. Thus, statement 1 implies statement 3.

- statement 3 $\implies$ statement 2

  Suppose T has a matrix representation

  $$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

  with respect to some orthonormal basis of V given by $e = e_1, e_2$. Then, any other orthonormal basis of V, given by $f = f_1, f_2$ is related to e by an isometry, i.e., via an orthogonal change of basis of matrix. So, the matrix representation of T with respect to f is

  $$B = ([S]_f^e)^T \begin{bmatrix} a & -b \\ b & a \end{bmatrix} [S]_f^e$$

  where S is a 2×2 orthogonal matrix. However, any such S represents either a rotation or a reflection and has the form

  $$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \text{ or } \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

  These commute with

  $$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

  so statement 3 implies statement 2.

- statement 2 $\implies$ statement 3
  This is trivial.

- statement 3 $\implies$ statement 1
  If
  $$[T]_e^e = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$
  with $b \neq 0$ and $e = e_1, e_2$ is an orthonormal basis, then
  $$[T^*]_e^e = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$
  Thus, $T \neq T^*$ since $b \neq 0$ so T is not self-adjoint. However, a direct calculation also shows us that
  $$[T^*T]_e^e = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} a^2+b^2 & 0 \\ 0 & a^2+b^2 \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}\begin{bmatrix} a & b \\ -b & a \end{bmatrix} = [TT^*]_e^e$$
  Hence, $T^*T = TT^*$ and T is normal.

Theorem 26.3
Let V be a real inner product space and $T \in \mathcal{L}(V)$. Then, the following are equivalent:

1. T is normal
2. There is an orthonormal basis of V with respect to which T is block diagonal, consisting of blocks that are 1×1 or 2×2 of the form
$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

Proof: The proof that we covered in class was too confusing and verbose, and I am genuinely not sure if it was even correct or complete in the first place. Refer to Axler for a better proof.

26.2 Real Canonical Form

Definition 26.1: Real Canonical Form
Let V be a real vector space and $T \in \mathcal{L}(V)$. Then there is a basis of V with respect to which the matrix of T is block diagonal with each block either a Jordan block for a real eigenvalue or a block of the form

$$\begin{bmatrix} R & I_2 & & & \\ & R & \ddots & & \\ & & \ddots & \ddots & \\ & & & R & I_2 \\ & & & & R \end{bmatrix}$$

where $R = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ with $b \neq 0$ and $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Main idea: $T_{\mathbb{C}}$ has a Jordan basis for which members of the generalized eigenspaces of real eigenvalues can be taken to be in V. If $N^k v_1, \ldots, Nv_1, v_1$ is a part of a Jordan basis satisfying the above for a complex eigenvalue $\lambda = a + ib$ block, then

$$\overline{N^k v_1} = N^k\,\overline{v_1}, \ldots, \overline{Nv_1} = N\,\overline{v_1}, \overline{v_1} = \overline{v_1}$$

can be taken as a part of the Jordan basis for a $\overline{\lambda} = a - ib$ block of the same size. Replace $N^k v_1, \ldots, Nv_1, v_1, N^k\,\overline{v_1}, \ldots, N\,\overline{v_1}, \overline{v_1}$ with $\mathrm{Im}(N^k v_1), \mathrm{Re}(N^k v_1), \ldots, \mathrm{Im}(Nv_1), \mathrm{Re}(Nv_1), \mathrm{Im}(v_1), \mathrm{Re}(v_1)$ are reorder to get the real canonical form.

For example, we might have a Jordan block

$$\begin{bmatrix} 2+i & 1 & \\ & 2+i & 1 \\ & & 2+i \end{bmatrix}$$

with respect to N^2v, Nv, v for some v, where $N = (T_{\mathbb{C}} - (2+i)I)|_{G(2+i,T_{\mathbb{C}})}$. Since $T_{\mathbb{C}} = \overline{T_{\mathbb{C}}}$ (as $T_{\mathbb{C}}$ is only a complexification of the real operator T), we have that $\overline{N} = \overline{(T_{\mathbb{C}} - (2+i)I)|_{G(2+i,T_{\mathbb{C}})}} = T_{\mathbb{C}} - (2-i)I|_{G(2-i,T_{\mathbb{C}})}$. Thus, there is a Jordan block

$$\begin{bmatrix} 2-i & 1 & \\ & 2-i & 1 \\ & & 2-i \end{bmatrix}$$

with respect to $N^2\,\overline{v}, N\,\overline{v}, \overline{v}$.

So, with respect to $\mathrm{Im}(N^2v), \mathrm{Re}(N^2v), \mathrm{Im}(Nv), \mathrm{Re}(Nv), \mathrm{Im}(v), \mathrm{Re}(v)$, we can combine the two Jordan blocks together to form

$$\begin{bmatrix} 2 & -1 & 1 & 0 & & \\ 1 & 2 & 0 & 1 & & \\ & & 2 & -1 & 1 & 0 \\ & & 1 & 2 & 0 & 1 \\ & & & & 2 & -1 \\ & & & & 1 & 2 \end{bmatrix}$$

for T's matrix representation when restricted to $G(2+i, T_{\mathbb{C}}) \oplus G(2-i, T_{\mathbb{C}})$.

Example 26.1: 9A Exercise 10

Give an example of a $T \in \mathcal{L}(\mathbb{C}^7)$ such that $T^2 + T + I$ is nilpotent.

Answer: We choose a T with minimal polynomial

$$m_T(z) = z\left(z - \left(\frac{-1+i\sqrt{3}}{2}\right)\right)\left(z - \left(\frac{-1-i\sqrt{3}}{2}\right)\right)$$

and characteristic polynomial

$$c_T(z) = z^5\left(z - \left(\frac{-1+i\sqrt{3}}{2}\right)\right)\left(z - \left(\frac{-1-i\sqrt{3}}{2}\right)\right)$$

such as

$$[T]_e^e = \begin{bmatrix} 0 & & & & & & \\ & 0 & & & & & \\ & & 0 & & & & \\ & & & 0 & & & \\ & & & & 0 & & \\ & & & & & \frac{-1+i\sqrt{3}}{2} & \\ & & & & & & \frac{-1-i\sqrt{3}}{2} \end{bmatrix}$$

Then, $T^2 + T + I$ only has an eigenvalue of 0. Thus, there is an upper-triangular matrix representation with a diagonal of 0s for some basis of $\mathbb{C}^7$. This implies that $T^2 + T + I$ must be nilpotent.

26.3 Gershgorin Circle Theorem and Perron-Frobenius Theorem

Theorem 26.4: Gershgorin Circle Theorem

Let $A \in \mathcal{M}_n(\mathbb{C})$. For each $i \in \{1, \ldots, n\}$, consider $R_i = \sum_{j \neq i} |a_{ij}|$, i.e., R_i is the sum of the absolute values of the non-diagonal entries of the ith row. Consider $D(a_{ii}, R_i) \subseteq \mathbb{C}$ to be the closed disk of radius R_i centered at a_{ii}. Then, each eigenvalue of A lies in at least one of the Gershgorin disks $D(a_{ii}, R_i)$.

Proof: Let λ be an eigenvalue of A and v a λ-eigenvector of A. Scale v so that one component v_i of v is equal to 1 and all other components have absolute value ≤ 1. Thus,

$$\sum_j a_{ij}v_j = \lambda v_i$$

$$\sum_{j\neq i} a_{ij}v_j + a_{ii} = \lambda$$

So,

$$|\lambda - a_{ii}| = \left|\sum_{j\neq i} a_{ij}v_j\right| \leq \sum_{j\neq i} |a_{ij}||v_j| \leq \sum_{j\neq i} |a_{ij}| = R_i$$

Note 26.1
The same result holds for the columns replacing the rows as well, by applying the theorem to A^T (since A and A^T have the same eigenvalues).

Theorem 26.5: Perron-Frobenius Theorem
Suppose that A is an $n \times n$ matrix with only positive entries. Consider the eigenvalues of $A_{\mathbb{C}}$. There is a unique eigenvalue λ of $A_{\mathbb{C}}$ that has the largest absolute value among the eigenvalues of $A_{\mathbb{C}}$. This λ is strictly positive and real and $E(\lambda, T) = 1$ for this eigenvalue. There is also a λ-eigenvector v with strictly positive entries.

Proof: There is no easy proof of this theorem. A famous one goes through Brouwer's fixed point theorem. Some use Gelfand's spectral radius formula:

$$\max\{|\lambda| \mid \lambda \text{ is an eigenvalue of } A_{\mathbb{C}}\} = \lim_{n\to\infty} \|A^n\|^{\frac{1}{n}}$$

which in itself is a famous result of functional analysis. However, the main point is that most (if not all) proofs of this theorem require some degree of topology and analysis that is not usually covered at an undergraduate level.

27 Lecture 27

27.1 Multilinear Algebra and Determinants

Definition 27.1: Alternating k-Form/k-Tensor
Let V be a vector space with $\dim V = n$ and $k \le n$. Then, $\phi : V^k \mapsto \mathbb{F}$ is an alternating k-form/k-tensor on V if

1. ϕ is multilinear, i.e., linear in each variable of ϕ:

$$\phi(v_1, \dots, v_j + w_j, \dots, v_k) = \phi(v_1, \dots, v_j, \dots, v_k) + \phi(v_1, \dots, w_j, \dots, v_k)$$
$$\phi(v_1, \dots, cv_j, \dots, v_k) = c\phi(v_1, \dots, v_j, \dots, v_k)$$

2. ϕ is alternating:
$$\phi(v_1, \dots, v_j, \dots, v_l, \dots, v_k) = -\phi(v_1, \dots, v_l, \dots, v_j, \dots, v_k)$$
for all $l \neq j$ (i.e., switching v_j and v_l introduces a negative sign).

Definition 27.2: k-Tensor
If $\phi : V^k \mapsto \mathbb{F}$ satisfies multilinearity but not necessarily the alternating property, then ϕ is a k-tensor on V.

Theorem 27.1
If $\dim V = n$, the space of k-tensors on V has dimension n^k.

Proof: Proof sketch: If ϕ is a j-tensor on V and ψ is a l-tensor on V, then

$$(\phi \otimes \psi)(v_1, \dots, v_j, v_{j+1}, \dots, v_{j+l}) = \phi(v_1, \dots, v_j) \cdot \phi(v_{j+1}, \dots, v_{j+l})$$

defines a $j + l$ tensor on V.
If $v_1, \dots, v_n$ is a basis for V and $\phi_1, \dots, \phi_n$ is the dual basis for V (i.e., $\phi_i(v_j) = \delta_{ij}$ for each i, j), then

$$\phi(v_{j_1}, \dots, v_{j_k}) = (\phi_{i_1} \otimes \cdots \otimes \phi_{i_k})(v_{j_1}, \dots, v_{j_k}) = \phi_{i_1}(v_{j_1}) \cdot \cdots \cdot \phi_{i_k}(v_{j_k}) = \delta_{i_1, j_1} \cdot \cdots \cdot \delta_{i_k, j_k}$$

is a k-tensor on V. The set of all $\phi_{i_1} \otimes \cdots \otimes \phi_{i_k}$, as the indices $i_1, \dots, i_k$ each vary from 1 to n (repetition allowed), give a basis for the space of all k-tensors on V. Thus, the space of k-tensors has dimension n^k.

What about alternating tensors? If ϕ is a k-tensor, define

$$\mathrm{Alt}(\phi)(v_1, \dots, v_k) = \frac{1}{k!} \sum_{\sigma \in S_k} (\operatorname{sgn} \sigma)\phi\big(v_{\sigma(1)}, \dots, v_{\sigma(k)}\big)$$

Here, S_k is the set of permutations of $\{1, 2, \dots, k\}$, i.e., the bijections between $\{1, 2, \dots, k\} \mapsto \{1, 2, \dots, k\}$. Therefore, S_k has a cardinality of $k!$ (hence the $k!$ in the denominator above). We define the sign function as

$$\operatorname{sgn} \sigma = \begin{cases} 1 & \text{if } \sigma \text{ has an even number of “inversions”} \\ -1 & \text{if } \sigma \text{ has an odd number of “inversions”} \end{cases}$$

These definitions seem a bit abstract. Let's work through a concrete example:

Example 27.1
Consider $\sigma \in S_4$ given by

$$\sigma(1) = 2$$
$$\sigma(2) = 3$$

$$\sigma(3) - 1$$
$$\sigma(4) = 4$$

There are two inversions here since $1 < 3$ but $\sigma(1) > \sigma(3)$ and $2 < 3$ but $\sigma(2) > \sigma(3)$. Thus, $\operatorname{sgn} \sigma = 1$.

Example 27.2
Now consider,

$$\sigma(1) = 3$$
$$\sigma(2) = 4$$
$$\sigma(3) = 1$$
$$\sigma(4) = 2$$

There are 4 inversions here since $1 < 3$ and $1 < 4$ but $\sigma(1) > \sigma(4) > \sigma(3)$ and $2 < 3$ and $2 < 4$ but $\sigma(2) > \sigma(4) > \sigma(3)$. Thus, $\operatorname{sgn} \sigma = 1$ this time as well.

Note 27.1
It isn't difficult to show that S_k has $\frac{k!}{2}$ even permutations and $\frac{k!}{2}$ odd permutations (see any abstract algebra textbook).

Example 27.3
If ϕ_1 and ϕ_2 are 1-tensors, then

$$\operatorname{Alt}(\phi_1 \otimes \phi_2) = \frac{1}{2}(\phi_1 \otimes \phi_2 - \phi_2 \otimes \phi_1)$$

If ϕ_1, ϕ_2 and ϕ_3 are 1-tensors, then,

$$\begin{aligned}\operatorname{Alt}(\phi_1 \otimes \phi_2 \otimes \phi_3) = \frac{1}{6}(&\phi_1 \otimes \phi_2 \otimes \phi_3 - \phi_2 \otimes \phi_1 \otimes \phi_3 - \phi_3 \otimes \phi_2 \otimes \phi_1 \\ &- \phi_1 \otimes \phi_3 \otimes \phi_2 + \phi_2 \otimes \phi_3 \otimes \phi_1 + \phi_3 \otimes \phi_1 \otimes \phi_2)\end{aligned}$$

Thus, we say $\operatorname{Alt}(\phi)$ is "totally antisymmetric," i.e., $\operatorname{Alt}(\phi_1 \otimes \phi_2) = -\operatorname{Alt}(\phi_2 \otimes \phi_1)$.

Definition 27.3: Exterior Product
If ϕ is a k-form and ψ is an l-form, we define the "wedge" or exterior product between them as

$$\phi \wedge \psi = \frac{(k+l)!}{k!l!} \operatorname{Alt}(\phi \otimes \psi)$$

In other words,

$$(\phi \wedge \psi)(v_1, \ldots, v_{k+l}) = \frac{1}{k!l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}(\sigma)\phi(v_{\sigma(1)}, \ldots, v_{\sigma(k)})\psi(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)})$$

The purpose of the binomial coefficient above is to sort of act like a normalization constant — it ensures that if $e_1, \ldots, e_n$ is the standard basis of $\mathbb{R}^n$ and $\phi_1, \ldots, \phi_n$ is the dual basis, then $(\phi_{i_1} \otimes \cdots \otimes \phi_{i_n})(e_{i_1}, \ldots, e_{i_n}) = 1$.

Theorem 27.2
If $\dim V = n$ and $v_1, \ldots, v_n$ is a basis of V with the corresponding dual basis $\phi_1, \ldots, \phi_n$, then the set of all $\phi_{i_1} \wedge \cdots \wedge \phi_{i_n}$ for $1 \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq n$ is a basis for the space of alternating k-forms on V.

The dimension of the space above is $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ for $k \leq n$. In particular, if $\dim V = n$, the space of alternating n-forms has dimension 1. This motivates the following definition:

Definition 27.4: Determinant
Let $A \in \mathcal{M}_n(\mathbb{C})$ or $\mathcal{M}_n(\mathbb{R})$. Identify A by its columns, i.e., identify $\mathcal{M}_n(\mathbb{C})$ with $\mathbb{C}^n \times \cdots \times \mathbb{C}^n = (\mathbb{C}^n)^n = \mathbb{C}^{n^2}$. Since $\mathbb{C}^n$ is n-dimensional, by the theorem above, there is a unique alternating n-form ϕ on $\mathbb{C}^n$ such that

$$\phi(e_1, \ldots, e_n) = \phi(I_n) = 1$$

This unique normalized alternating n-form $\phi : \mathcal{M}_n(\mathbb{C}) \cong (\mathbb{C}^n)^n \mapsto \mathbb{C}$ is called the determinant. That is, $\det(A) = \phi(A)$.

Alternating n-forms and the determinant play a fundamental role in defining the nature of an "orientation" on a vector space or, more generally, a differentiable manifold. Choose any basis of an n-dimensional vector space V. Any other basis of V is related to the first one via a change of basis matrix, which is invertible. The two bases are said to define the same orientation on V if this change of basis matrix has $\det > 0$. Otherwise, the bases are considered as having opposite orientations. This divides the set of all bases of V into two disjoint classes. Usually, the class including "the standard basis" (if $V = \mathbb{R}^n$ or $\mathbb{C}^n$) is taken to be the "usual or standard" orientation.

Definition 27.5: Permutation Definition of the Determinant
Let A be an $n \times n$ matrix with columns $a_1, \ldots, a_n$. Then,

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{1 \le i \le n} a_{i,\sigma(i)}$$

Compare this to the definition of an alternating n-tensor:

$$\operatorname{Alt}(\phi)(v_1, \ldots, v_n) = \frac{1}{n!} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma)\phi\big(v_{\sigma(1)}, \ldots, v_{\sigma(n)}\big)$$

Since the determinant is a normalized n-tensor, we can essentially get rid of the $n!$ term. Moreover, from the definition of tensor products, we already know that

$$\phi = \phi_1 \otimes \cdots \otimes \phi_n \implies \phi(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) = \phi_1\big(v_{\sigma(1)}\big) \cdot \cdots \cdot \phi_n\big(v_{\sigma(n)}\big) = \prod_{1 \le i \le n} \phi_i\big(v_{\sigma(i)}\big)$$

Here, it follows that $\phi_i(a_j) = a_{ij}$ where a_j is the jth column of the matrix A.

Example 27.4
The set S_2 contains only two permutations:

$$\sigma(1) = 1, \sigma(2) = 2 \text{ and } \tau(1) = 2, \tau(2) = 1$$

Then, $\operatorname{sgn}(\sigma) = 1$ and $\operatorname{sgn}(\tau) = -1$. So, for an arbitrary 2×2 matrix, we have

$$\det\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = (1)(a_{11} * a_{22}) + (-1)(a_{12} * a_{21}) = a_{11}a_{22} - a_{12}a_{21}$$

Example 27.5
Let A be a 3×3 matrix. Then, S_3 consists of $3! = 6$ permutations. Based on example 27.2, we set

$$\det(A) = a_{11}a_{22}a_{33} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$

Key observations regarding the determinant definition given above:

- When we expand out
$$\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i \le n} a_{i,\sigma(i)}$$
we see that each term in the product contains exactly one term from each row of A, namely $a_{i,\sigma(i)}$.

- Moreover, each $a_{j,k}$ is $a_{j,\sigma(j)}$ for some $\sigma \in S_n$.

From these two observations, and possibly some re-indexing, we can conclude the following statements about determinants:

1. If B results from A when a single row of it is multiplied by a scalar c, then $\det(B) = c\det(A)$.
2. If B results from A by switching two rows of A, then $\det(B) = -\det(A)$.
3. If B results from A by replacing row k of A with $\mathrm{row}(k) + c \cdot \mathrm{row}(j)$, then $\det(B) = \det(A)$.

Note 27.2
Among many other things, statements 1 and 2 imply that if A has two repeated rows, then $\det(A) = 0$. If A has a row of all zeroes, then $\det(A) = 0$.

Since $\det(B)$ is either $c \cdot \det(A)$ for some $c \neq 0$, $-\det(A)$ or $\det(A)$, if B results from A after some sequence of elementary row operations (à la Gaussian elimination), it holds that $\det(B) = 0$ iff $\det(A) = 0$.

Consider $\mathrm{RREF}(A)$. We know that $\mathrm{RREF}(A) = I_n$ iff A is invertible. Note that $\det(I_n) = 1$ since $b_{11} \cdot b_{22} \cdot \dots \cdot b_{nn} = 1$ and $\prod_{1 \le i \le n} b_{i,\sigma(i)} = 0$ for all permutations where $\sigma(i) = j \neq i$. Recall that $\mathrm{RREF}(A)$ has at least one row of all zeros iff A is not invertible. Therefore, $\det(\mathrm{RREF}(A)) = 0$ iff A is not invertible, but $\det(\mathrm{RREF}(A)) = 1$ iff A is invertible.

Combining these observations with what we have above, we have essentially proven that $\det(A) \neq 0$ iff A is invertible.

Theorem 27.3
$\det(AB) = \det(A)\det(B)$ for all $n \times n$ matrices A and B.

Proof: Note that AB is not invertible iff at least one of A and B is not invertible. In that case, $\det(AB)$ and $\det(A)\det(B)$ are both zero. Now, we need to show the case when A and B are both invertible.

Recall that an elementary matrix is the result of applying a single elementary row operation to I_n. For instance,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

is an elementary matrix that results from swapping rows 2 and 3 of I_3. If E is the elementary matrix corresponding to a row operation, then $B = EA$ is the result of applying that row operation fo A. For example,

$$\begin{bmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

implements switching rows 1 and 2 of A.

Since A is invertible iff $\mathrm{RREF}(A) = I_n$, then A is invertible iff $\mathrm{RREF}(A) = I_n = E_k \dots E_1 A$ for some elementary matrices $E_1, \dots, E_k$. However,

$$A = (E_k \cdot \dots \cdot E_1)^{-1} = (E_1)^{-1} \cdot \dots \cdot (E_k)^{-1}$$

The inverse of an elementary matrix is an elementary matrix too. Thus, A is invertible iff A is a product of elementary matrices.

Observe that $E = EI_n$ is the result of applying an elementary row operation to I_n and $\det(I_n) = 1$, so $\det(E) = c$, -1 or 1 depending on the type of row operation that E implements. Thus, if B results from A by a row operation implemented by E_1, we get that $\det(B) = \det(EA) = \det(E)\det(A)$. Furthermore, $\det(E_2(E_1A)) = \det(E_2)\det(E_1A) = \det(E_2)\det(E_1)\det(A)$, etc. Carrying out this process inductively yields that $\det(BA) = \det(B)\det(A)$ iff $B = E_j \cdot \dots \cdot E_1$ is invertible. In other words, $\det(BA) = \det(B)\det(A)$ for any $n \times n$ matrices A and B.

Theorem 27.4
If A is invertible, then $\det\left(A^{-1}\right) = \frac{1}{\det(A)}$.

Proof: $1 = \det(I_n) = \det\left(AA^{-1}\right) = \det(A)\det\left(A^{-1}\right) \implies \det\left(A^{-1}\right) = \frac{1}{\det(A)}$

Theorem 27.5
If A and B are similar, then $\det(A) = \det(B)$.

Proof: If $B = S^{-1}AS$, then $\det(B) = \frac{1}{\det(S)} \det(A) \det(S) = \det(A)$

Theorem 27.6
If A is an upper-triangular $n \times n$ matrix, then $\det(A) = a_{11} \cdot \dots \cdot a_{nn}$, i.e., the product of the terms along the diagonal.

Proof: Since A is upper-triangular, it follows that $a_{1,\sigma(1)} \cdot \dots \cdot a_{n,\sigma(n)} = 0$ unless $a_{1,\sigma(1)} = a_{11}, a_{2,\sigma(2)} = a_{22}, a_{3,\sigma(3)} = a_{33}$ and so on until the nth row. That is, for all permutations other than the identity permutation ($\sigma_I(i) = i$ for all $i \le n$), the product above will evaluate to 0. The sign of the identity permutation is 1 since there are no inversions. Thus,

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i \le n} a_{i,\sigma(i)} = (+1) \cdot a_{11} \cdot \dots \cdot a_{nn} + \sum_{\sigma \in S_n \setminus \{\sigma_I\}} \operatorname{sgn}(\sigma) \cdot 0 = a_{11} a_{22} \cdots a_{nn}$$

Recall that if $T \in \mathcal{L}(V)$ has an upper triangular matrix representation, then its eigenvalues are entries along the main diagonal. We also know that the eigenvalues of T are similarity-invariant, i.e. if $S \in \mathcal{L}(V)$ is an isometry, then T and $S^{-1}TS$ have the same eigenvalues with the same multiplicities. Thus, we could also simply define $\det(T)$ to equal the product of its eigenvalues with multiplicity (in fact, this is how Axler does it).

If V is real, then $T_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}})$ is complex and, hence, has an upper triangular matrix representation. Then, by our last theorem, we can define $\det(T)$ to be the product of the eigenvalues, with multiplicity, of $T_{\mathbb{C}}$ (if it is real) or T (if it is complex). It then becomes clear immediately that $\det(T) \neq 0$ iff T is invertible. Moreover, $\det(S \circ T) = \det(S)\det(T)$.

27.2 Geometric Interpretation of Determinants

Let T be a linear transformation and let $v_1 = T(e_1)$ and $v_2 = T(e_2)$. Then,

$$A = \begin{bmatrix} T(e_1) & T(e_2) \end{bmatrix} = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$$

is the matrix that applies the transformation T so some vector $v \in \mathbb{R}^2$. Consider the area of the parallelogram spanned by v_1, v_2 — this is equal to $\|v_1\| \|v_2\| \sin\theta$ where θ is the angle spanned between v_1 and v_2.

Note that the height of the parallelogram is given by $v_2^\perp$, the component of v_2 orthogonal to v_1. Thus, the area above can be rewritten as $\|v_1\| \|v_2^\perp\|$. We claim that

$$|\det(A)| = \|v_1\| \|v_2^\perp\|$$

Let $A = QR$ (the QR decomposition of a matrix can be achieved by performing Gram-Schmidt on its columns), where Q is orthogonal, and R is upper triangular with diagonal entries $r_{11} = \|v_1\|$ and $r_{22} = \|v_2 - \operatorname{proj}_{\operatorname{span}(v_1)} v_2\| = \|v_2^\perp\|$. However, note that $I = Q^T Q \implies \det(T) = \det(Q^T)\det(Q) = \det(Q)^2 \implies |\det(Q)| = 1$. Therefore,

$$|\det(A)| = |\det(QR)| = |\det(Q)||\det(R)| = \|v_1\| \|v_2^\perp\|$$

Thus, $|\det(A)|$ is the area of the parallelogram spanned by its columns. This generalizes to

Theorem 27.7
Let A be an $n \times n$ matrix with $v_1, \dots, v_n$ as its columns. Then, $|\det(A)| = \|v_1\| \|v_2^\perp\| \dots \|v_n^\perp\|$ where

$$v_k^\perp = v_k - \operatorname{proj}_{\operatorname{span}(v_1, \dots, v_k)} v_k$$

So, if A is a 3×3 matrix, then $|\det(A)|$ is the volume of the parallelepiped determined by v_1, v_2 and v_3. Using the linearity of a linear transformations, it also follows that

Theorem 27.8

If $T : \mathbb{R}^n \mapsto \mathbb{R}^n$, then $|\det(T)| = \frac{n\text{−volume of } T(\Omega)}{n\text{−volume of } \Omega}$ where Ω is any closed bounded region in $\mathbb{R}^n$.

Theorem 27.9: Change of Variables in Integration

Let Ω be an open subset of $\mathbb{R}^n$ and $\sigma : \Omega \mapsto \mathbb{R}^n$ is continuously differentiable at every point of Ω. If f is integrable on $\sigma(\Omega)$, then

$$\int_{\sigma(\Omega)} f(y)\,\mathrm{d}y = \int_{\Omega} f(\sigma(x))|\det \sigma'(x)|\,\mathrm{d}x$$

where $\sigma'(x)$ is the Jacobian of σ.