MATH672: Vector Spaces

Jonathan Ma
johnma@udel.edu

December 3, 2024

Contents

1 Week 1

1.1 Lecture 2. Thu Aug 29

1.1.1 Subspaces, Sums, Direct Sums

In this section, let $k = 1, \ldots, m$, and $V_k \leq V$.

Definition 1.1 (Sum of subspaces)**.** The sum $V_1 + \ldots + V_m := \{v_1 + \ldots + v_m \mid v_k \in V_k\}$.

Theorem 1.2. *The sum $W := V_1 + \ldots + V_m$ is the smallest subspace which contains $V_1, \ldots, V_m$.*

Proof. First, W is certainly a subset of V. Using the closure properties of $V_1, \ldots, V_m$, we can show W is a subspace of W. Each element of V_k is an element of W, since we can take $v_k \in V_k$ as the sum of v_k and the identity element of V, which is contained in each V_k since they are subspaces. To show W is the smallest subspace of V containing each V_k, assume W' is a subspace containing each V_k. Then $W \leq W'$, since subspaces contain all the finite sums of their elements. □

Definition 1.3 (Direct sums)**.** The sum $W := V_1 + \ldots + V_m$ is a direct sum if and only if each element $w \in W$ can be written uniquely as a sum of $v_1 + \ldots + v_m$, for $v_k \in V_k$. We write W as $W := V_1 \oplus \ldots \oplus V_m$.

Theorem 1.4 (Criteria for direct sums)**.** *A sum of subspaces $W := V_1 + \ldots + V_m$ is a direct sum if and only if the only way to write $0 \in V$ as $v_1 + \ldots + v_m$ is if each $v_k = 0$.*

Proof. For the converse implication, suppose that the only way to write 0 in W is $0 = v_1 + \ldots + v_m$, with $v_k = 0$, for all k. Take $v \in W$, with $v = u_1 + \ldots + u_m$, $u_k \in V_k$ and suppose we can also write $v = w_1 + \ldots + w_m$, $w_k \in V_k$. Then we have $0 = v - v = (u_1 - w_1) + \ldots + (u_m - v_m)$. Then we must have $u_k = w_k$, by our assumption. □

Theorem 1.5 (More criteria for direct sums)**.** *For $U, W \leq V$, $U + W$ is a direct sum if and only if $U \cap W = \{0\}$.*

Exercise 1.1. Simplify $(3 + 2i)(1 - i)$. State both associativity properties of vector spaces. For $S = [0, 1]$, what is the additive identity of $\mathbb{F}^S$?

2 Week 2

2.1 Lecture 3. Tue Sep 3

2.1.1 Span and Linear Independence

Definition 2.1 (Linear combinations)**.** Suppose $v_1, \ldots, v_m \in V$, and $a_1, \ldots, a_m \in \mathbb{F}$. Then a linear combination of $v_1, \ldots, v_m$ is $v = a_1 v_1 + \ldots + a_m v_m$.

Definition 2.2 (Span)**.** The span of $v_1, \ldots, v_m \in V$ is $\operatorname{span}(v_1, \ldots, v_m) = \{a_1 v_1 + \ldots + a_m v_m \mid a_i \in \mathbb{F}, i \in [m]\}$. We define $\operatorname{span}(\emptyset) = \{0\}$.

Theorem 2.3. *The smallest subspace of V containing $v_1, \ldots, v_m \in V$ is $\operatorname{span}(v_1, \ldots, v_m)$.*

Proof. Let $S := \operatorname{span}(v_1, \ldots, v_m)$. Now, let $u, w \in S$. Write $u = a_1v_1 + \ldots + a_mv_m$, and $w = b_1v_1 + \ldots + b_mv_m$, for $a_i, b_i \in \mathbb{F}, i \in [m]$. Then $u + w = (a_1 + b_1)v_1 + \ldots + (a_m + b_m)v_m$, and since each a_i, b_i are scalars, we know that $a_i + b_i \in \mathbb{F}$, so that $u+w$ is a linear combination of $v_1, \ldots, v_m$. A similar proof shows that S is closed under scalar multiplication. To show S contains $v_1, \ldots, v_m$, note that each v_j is a linear combination of $v_1, \ldots, v_m$ in the form of $v_j = 0v_1 + \ldots + 1v_j + \ldots + 0v_m$, for each $j \in [m]$. To show that S is the smallest subspace containing $v_1, \ldots, v_m$, let S' be a subspace of V containing $v_1, \ldots, v_m$. Then S' is closed under addition and scalar multiplication, and it follows that S is contained in S'. □

Example 2.4. The standard basis vectors $e_1 = (1, 0, 0), e_2 = (0, 1, 0), e_3 = (0, 0, 1)$ span $\mathbb{R}^3$.

Example 2.5. The polynomials $1, x, \ldots, x^m$ span $\mathcal{P}_m$.

Definition 2.6 (Finite dimensional vector spaces)**.** A vector space V is finite dimensional if it is spanned by a list. Otherwise, V is infinite dimensional.

Definition 2.7 (Linear independence)**.** A list of vectors $v_1, \ldots, v_m \in V$ is said to be linearly independent if the only way to express the zero vector 0 in V as a linear combination of $v_1, \ldots, v_m$ is if $0 = a_1v_1 + \ldots + a_mv_m$, $a_j = 0, \forall j \in [m]$. If this condition is not satisfied, then $v_1, \ldots, v_m$ is said to be linearly dependent.

Example 2.8. A single vector $v \in V$ is linearly independent if and only if $v \neq 0$.

Example 2.9. The list $v_1, v_2 \in V$ is linearly independent if and only if neither are scalar multiples of one another.

Example 2.10. The monomials $1, x, \ldots, x^m \in \mathcal{P}$ are linearly independent.

Example 2.11. Any list containing 0 is linearly dependent.

Theorem 2.12. *Let $v_1, \ldots, v_m \in V$ is a linearly dependent list of vectors. Then there exists some $j \in [m]$, such that $v_j \in S := \operatorname{span}(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_m)$, and $\operatorname{span}(v_1, \ldots, v_m) = S$.*

Proof. Because $v_1, \ldots, v_m$ are linearly dependent, there is some non-trivial representation of 0. Suppose $0 = a_1v_1 + \ldots + a_mv_m$, with some $a_j \neq 0, j \in [m]$. Solving for v_j yields

$$v_j = \frac{-a_1}{a_j}v_1 - \ldots - \frac{a_{j-1}}{a_j}v_{j-1} - \frac{a_{j+1}}{a_j}v_{j+1} - \ldots - \frac{a_m}{a_j}v_m.$$

Then $v_j \in S$. Showing $S \subseteq \operatorname{span}(v_1, \ldots, v_m)$ is trivial. Suppose that $v \in \operatorname{span}(v_1, \ldots, v_m)$. Then $v = a_1v_1 + \ldots + a_mv_m$, for $a_i \in \mathbb{F}, i \in [m]$. We have shown there is some $v_j \in S$ for $j \in [m]$, so suppose $v_j = b_1v_1 + \ldots + b_{j-1}v_{j-1} + b_{j+1}v_{j+1} + \ldots + a_mv_m$, for $b_i \in \mathbb{F}, i \in [m], i \neq j$. Then $v = (a_1 + a_jb_1)v_1 + \ldots + (a_{j-1} + a_jb_{j-1})v_{j-1} + (a_{j+1} + a_jb_{j+1})v_{j+1} + \ldots + (a_m + a_jb_m)v_m$. Therefore, $v \in S$. □

Theorem 2.13. *In a finite dimensional vector space V, the length of any linearly independent list is bounded by the length of any spanning list of V.*

Proof. Suppose $U = \{u_1, \ldots, u_m \in V\}$ are linearly independent, and $B_1 = \{w_1, \ldots, w_n\}$ spans V. Consider $u_1, w_1, \ldots, w_n$. Since B_1 spans V, we have $u_1 \in \operatorname{span}(B_1)$. Therefore, $u_1, w_1, \ldots, w_n$ is linearly dependent. Also, note that because U is linearly independent, each $u_i \neq 0$, for $i \in [m]$. Also, in $0 = b_1u_1 + a_1w_1 + \ldots + a_nw_n$, there must be some $a_j \neq 0$, for $j \in [n]$. WLOG, assume $a_1 \neq 0$. Then $\operatorname{span} B_1 = \operatorname{span}(u_1, w_2, \ldots, w_n)$. At the jth step of this process, we will have some $B_j = \{u_1, \ldots, u_{j-1}, w_j, \ldots, w_n\}$, which spans V. Therefore, $u_j \in \operatorname{span} B_j$, and u_j, B_j is a linearly dependent list. We repeat the process described earlier to yield that $\operatorname{span} B_j = \{u_1, \ldots, u_j, w_{j+1}, \ldots, w_n\}$. After the nth step of this process, we shall have $B_m = \{u_1, \ldots, u_m, w_{m+1}, \ldots, w_n\}$. □

Theorem 2.14. *Every subspace of a finite dimensional vector space is finite dimensional.*

Proof. Let $\dim V < \infty, U \leq V$. If $U = \{0\}$, then $\dim U < \infty$. Otherwise, choose $0 \neq v_1 \in U$. In the jth step of this process, let $U = \text{span}(v_1, \ldots, v_{j_1})$, then $\dim U < \infty$. Otherwise, choose $v_j \neq 0 \in U \setminus \text{span}(v_1, \ldots, v_{j-1})$. By the previous theorem, $v_1, \ldots, v_j$, $j \leq \dim V < \infty$. Therefore this process terminates after a finite number of steps. □

2.2 Lecture 4. Thu Sep 5

2.2.1 Bases

Definition 2.15. A basis is a linearly dependent and spanning set of vectors in V.

Theorem 2.16. *The set of vectors $v_1, \ldots, v_n$ is a basis of V if and only if $\forall v \in V, \exists!\, a_j$ s.t. $v = a_1v_1 + \ldots + a_nv_n$.*

Theorem 2.17. *Every spanning list of vectors in V can be reduced to a basis of V.*

Proof. Let $B_0 := v_1, \ldots, v_m$ span V. If B_0 is linearly independent, then we are done. Otherwise, we have that $0 = a_1v_1 + \ldots + a_mv_m$, for a_i not all 0, $i \in [m]$. WLOG, assume that $a_m \neq 0$, so that we may write

$$v_m = \frac{a_1}{a_m}v_1 + \ldots + \frac{a_{m-1}}{a_m}v_{m-1},$$

so that $V = \text{span}\, B = \text{span}(v_1, \ldots, v_{m-1}) := B_1$. Repeating this process guarantees we will terminate with a basis of V. □

Corollary 2.17.1. *Every finite dimensional vector space V has a basis.*

Proof. By definition of a finite dimensional vector space, V is spanned by a spanning list of vectors, which we may reduce to a basis of V. □

Theorem 2.18. *Let V be finite dimensional. Then every linearly independent list of vectors in V can be extended to a basis of V.*

Proof. Let $v_1, \ldots, v_n \in V$ be linearly independent, and let $w_1, \ldots, w_m$ be a basis of V. Then $\{v_1, \ldots, v_n, w_1, \ldots, w_m\}$ spans V. Reusing the technique shown in the proof for Theorem 2.17, it remains to note that we can retain all the v_i's, since they are linearly independent. □

Theorem 2.19. *Let $U \leq V$, for V a finite dimensional vector space. Then there exists some $W \leq V$, such that $V = U \oplus W$.*

Proof. Let $B_u := u_1, \ldots, u_m$ be a basis of U, and because B_u is linearly independent, we can extend it to a basis of V by adjoining the vectors $B_w = w_1, \ldots, w_k$. Define $W = \text{span}(w_1, \ldots, w_n)$. Then $W \leq V$. First, we shall show that $V = U + W$. Let $u \in U, w \in W$, and since $U \leq V, W \leq V$, $u + w \in V$. Also, take $v \in V$, and since $B_u \cup B_w$ is a basis of Vi we know that $v = a_1u_1 + \ldots + a_mu_m + b_1w_1 + \ldots + b_nw_n$, for some scalars $a_i, b_i \in \mathbb{F}$. It is clear then that v is the sum of an element of U, and an element of W, so that $v \in U + W$.
To show that $U \cap V = \{0\}$, let $v \in U \cap B \subseteq V$. Then $v \in U$ and $v \in W$, so that $v = a_1u_1 + \ldots + a_mu_m = b_1w_1 + \ldots + b_nw_n$, but also

$$0 = v - v = a_1u_1 + \ldots + a_mu_m - b_1w_1 - \ldots - b_nw_n,$$

and since $B_u \cup B_w$ spans V, each coefficient must be 0 so that $v = 0$.

□

2.2.2 Dimension

Theorem 2.20. *Any two bases of a finite dimensional vector space V have the same length.*

Proof. Let B_1, B_2 be bases of V. Then B_1, B_2 is linearly independent and both spans V. By Theorem 2.13, the lengths of B_1, B_2 bound each other, so that they must have equal length. $\square$

Definition 2.21. The length of basis of V is called the dimension of V

Example 2.22. Since $1, x, \dots, x^5$ is a basis of $\mathcal{P}_5$, $\dim \mathcal{P}_5 = 6$.

Theorem 2.23. *If $U \leq V$, for V a finite dimensional vector space, then $\dim U \leq \dim V$.*

Proof. A basis of U is a linearly independent set of vectors in V. Extending this basis of U to a spanning set of V, and applying Theorem 2.17 yields our result. $\square$

Theorem 2.24. *Let V be a finite dimensional vector space. Let $B := v_1, \dots, v_n \in V$. Then any two of the following shall imply the third:*

1. $\dim V = n$,
2. *B spans V,*
3. *B is linearly independent in V.*

Proof. Statement 2 and 3 are the definition of a basis, implying statement 1 immediately. For statement 1 and 3, when extending B to a basis of V, we are not including any new vectors because we already have n of them. Statement $1, 2$ imply statement 3 in analogy. $\square$

Theorem 2.25. *Let $U_1, U_2 \leq V$. Then $\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$.*

Proof. Let $B := v_1, \dots, v_n$ be a basis of $U_1 \cap U_2$. Let $n := \dim(U_1 \cap U_2)$. Then B extends to a basis of U_1, call it $v_1, \dots, v_n, u_1, \dots, u_j$, we that $\dim U_1 = n + j$. Analogously, let $v_1, \dots, v_n, w_1, \dots, w_k$ be a basis of U_2, so that $\dim U_2 = n + k$. We want to prove that $v_1, \dots, v_n, u_1, \dots, u_j, w_1, \dots, w_k$ is a basis of $U_1 + U_2$, since then $\dim(U_1 + U_2) = n + j + k = (n + j) + (n + k) - n = \dim U_1 + \dim U_2 - \dim U_1 \cap U_2$. $\square$

3 Week 3

3.1 Lecture 5. Tue Sep 10

3.1.1 Linear Maps

Definition 3.1. A linear map or transformation is a function $T : V \to W$ such that for $u, w \in V$, and $c \in \mathbb{F}$, $T(u + w) = T(u) + T(v)$, and $T(cv) = cT(v)$.

Definition 3.2. Denote by $\mathcal{L}(V, W)$ the set of all linear maps from V to W.

Theorem 3.3. *Let $v_1, \dots, v_m$ be a basis of V, and $w_1, \dots, w_m$ be a basis of W. Then there exists a unique linear map $T \in \mathcal{L}(V, W)$ such that $T(v_j) = w_j$, for $j \in [m]$.*

Proof. Let $v^1 \in V$. Then because the v_i's are a basis, there exists scalars $c_i \in \mathbb{F}$ such that $v^1 = c_1^1 v_1 + \ldots + c_m^1 v_m$. Define $T(v^1) = T(c_1^1 v_1 + \ldots + c_m^1 v_m) = c_1^1 w_1 + \ldots + c_m^1 w_m$. Since T is defined on each element of V, T is a map. To show that T is a linear map, let $v^1, v^2 \in V$, $\alpha \in \mathbb{F}$. Then

$$\begin{aligned}
T(v^1 + \alpha v^2) &= T(c_1^1 v_1 + \ldots + c_m^1 v_m + \alpha c_1^2 v_1 + \ldots + \alpha c_m^2 v_m) \\
&= T((c_1^1 + \alpha c_1^2) v_1 + \ldots + (c_m^1 + \alpha c_m^2) v_m) \\
&= (c_1^1 + \alpha c_1^2) w_1 + \ldots + (c_m^1 + \alpha c_m^2) w_m \\
&= (c_1^1 w_1 + \ldots + c_m^1 w_m) + \alpha (c_1^2 w_1 + \ldots + c_m^2 w_m) \\
&= T(v^1) + \alpha T(v^2).
\end{aligned}$$

For uniqueness, suppose T satisfies $T(v_j) = w_j$, for $j \in [m]$. Then by the unique representation of basis vectors, $T(v) = T(c_1 v_1 + \ldots + c_n w_n) = c_1 T(c_1) = \ldots + c_m T(v_m) = c_1 w_1 + \ldots + c_m w_m$, so that T is uniquely defined on $\operatorname{span}(v_1, \ldots, v_n)$. □

Theorem 3.4. *For $S, T \in \mathcal{L}(V, W)$, $\alpha \in \mathbb{F}$, define $S + T$ by $(S + T)(v) = S(v) + T(v)$, and define αS by $(\alpha S)(v) = \alpha(S(v))$. Then $\mathcal{L}(V, W)$ is a vector space.*

Definition 3.5. For $S \in \mathcal{L}(V, W)$, $T \in \mathcal{W}, \overline{\mathcal{W}}$, define TS by $(TS)(v) = T(S(v))$. Also, $TS \in \mathcal{V}, \overline{\mathcal{W}}$.

Theorem 3.6. *For $S, T, U \in \mathcal{L}(V, W)$,*

1. *$S(TU) = (ST)U$,*
2. *$I \in \mathcal{L}(V, W)$ is the multiplicative identity.*

Note. As an aside,

$$e^{iA} = \sum_{k=0}^{\infty} \frac{iA^k}{k!},$$

but

$$e^{A+B} = \sum_{k=0}^{\infty} \frac{(A+B)^k}{k!} \neq e^A e^B.$$

3.2 Lecture 6. Thu Sep 12

3.2.1 Null Space, Range

Definition 3.7. For $T \in \mathcal{L}(V, W)$, $\operatorname{null} T = \{v \in V \mid T(v) = 0\} = \ker T$.

Example 3.8. The null space of the backward shift $T(x_1, x_2, \ldots) = (x_2, \ldots)$ is $\ker T = \{(x_1, 0, \ldots) \mid x_1 \in \mathbb{F}\}$.

Example 3.9. The null space of the derivative operator D on $\mathcal{P}(\mathbb{R})$ is $\ker D = \{c \mid c \in \mathbb{F}\}$, the constant functions.

Theorem 3.10. *The null space of T is a subspace of V.*

Theorem 3.11. *Suppose $u, v \in \ker T$. Then $T(u + v) = T(u) + T(v) = 0 + 0 = 0$. Also, for $c \in \mathbb{F}$, $T(cu) = cT(u) = c0 = 0$.*

Theorem 3.12. *For $T \in \mathcal{L}(V, W)$, T is injective iff $\ker T = 0$.*

Proof. If T is injective, since $T(0) = 0$, $\ker T = 0$. If $\ker T = 0$, then let $u, v \in V$, and $T(u) = T(v)$. If $u \neq v$, then $u - v \neq 0$, so that $T(u - v) = T(u) - T(v) = 0$, contradicting $\ker T = 0$. □

Definition 3.13. Let $T : X \to Y$, for sets X, Y. Then T has $\operatorname{range} T = \{w \in W \mid \exists v \in V, T(v) = w\}$.

Theorem 3.14. *If $T \in \mathcal{L}(V, W)$, then $\operatorname{range} T \leq W$.*

Proof. Let $u, v \in \operatorname{range} T$. Then $T(u_0) = u$, and $T(v_0) = v$,for $u_0, v_0 \in V$. Then $T(u_0 + v_0) = T(u_0) + T(v_0) = u + v \Rightarrow u + v \in \operatorname{range} T$, and $T(cu_0) = cT(u_0) = cu \Rightarrow cu \in \operatorname{range} T$. □

Theorem 3.15 (Linear Maps, Fundamental Theorem of)**.** *Suppose $\dim V < \infty$, $T \in \mathcal{L}(V, W)$. Then $\operatorname{range} T$ is finite dimensional, and also $\dim V = \dim \operatorname{range} T + \dim \ker T$.*

Proof. Extend the basis $u_1, \ldots, u_m$ of $\operatorname{null} T$ to a basis $u_1, \ldots, u_m, v_1, \ldots, v_m$ of V. Then $\dim V = m + n$. We want to show that $\dim \operatorname{range} T = n < \infty$. It suffices to prove $Tv_1, \ldots, Tv_n$ is a basis of $\operatorname{range} T$. Let $v \in V$. Then $v = \sum a_j u_j + \sum b_i v_i$, so that $T(v) = \sum a_j T(u_j) + \sum b_i T(v_i) = 0 + \sum b_i T(v_i)$. Therefore, any vector $T(v) \in \operatorname{range} T$ is a linear combination of the vectors $Tv_1, \ldots, Tv_n$. Let $0 = \sum c_i T(v_i) = T(\sum c_i v_i) \Rightarrow \sum c_i v_i = \sum b_j u_j \Rightarrow \sum c_i v_i + \sum(-b_j) u_j = 0 \Rightarrow c_i$'s are all zero. □

Theorem 3.16. *Suppose $\dim W < \dim V < \infty$, and $T \in \mathcal{L}(V, W)$. Then T cannot be injective.*

Proof. We know $\dim \ker T = \dim V - \dim \operatorname{range} T \Rightarrow \dim \ker T > \dim W - \dim W = 0$. □

Theorem 3.17. *Suppose $\dim V < \dim W < \infty$, and $T \in \mathcal{L}(V, W)$. Then T cannot be surjective.*

Proof. We know $\dim \operatorname{range} T = \dim V - \dim T \Rightarrow \dim \operatorname{range} T < \dim W$. □

Theorem 3.18. *A homogeneous system of linear equations that has more variables than equations has a non-trivial solution.*

Theorem 3.19. *An inhomogeneous system of linear equations with more equations than variables has no solutions for some choice of constant coefficients.*

4 Week 4

4.1 Lecture 7. Tue Sep 17

Theorem 3.18. For $A \in \mathbb{F}^{m,n}$, write

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ \vdots & \ldots & \ldots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix}.$$

Then the rows of $Ax = c$ are

$$\begin{pmatrix} \sum_{k=1}^{n} A_{1k} x_k \\ \vdots \\ \sum_{k=1}^{n} A_{mk} x_k \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}.$$

Reinterpreting our left hand side using $T : \mathbb{F}^n \to \mathbb{F}^m$ given by

$$T(x) = \left(\sum_{k=1}^{n} A_{1k} x_k, \ldots, \sum_{k=1}^{n} A_{mk} x_k \right),$$

(and one may verify T is a linear map) we note that since $n > m$, T cannot be injective. The null space of T is the set of all x such that x is a solution to our homogeneous system. □

Theorem 3.19. Reusing the same linear map T as described above, since $m > n$, T cannot be surjective. Therefore, there is some c which does not have an x mapped to it by T. □

4.1.1 Matrices of Linear Maps

Definition 4.1. The matrix representation of $T \in \mathcal{L}(V, W)$ with respect to the basis $v_1, \ldots, v_n$ of V, and the basis $w_1, \ldots, w_n$ of W is the $m \times n$ matrix $A := \mathcal{M}(T, (v_1, \ldots, v_1), (w_1, \ldots, w_n))$, where $T(v_k) = A_{1k}w_1 + \ldots + A_{mk}w_m$.

Example 4.2. Let $D \in \mathcal{L}(\mathcal{P}_3, \mathcal{P}_2)$ be the derivative mapping with the bases $1, x, x^2, x^3$ of $\mathcal{P}_3$ and $1, x, x^2$ of $\mathcal{P}_2$. The matrix $A = \mathcal{M}(D, (1, x, x^2, x^3), (1, x, x^2))$ is written

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Theorem 4.3. *For $S, T \in \mathcal{L}(V, W)$, and $\lambda \in \mathbb{F}$, $M(S + T) = M(S) + M(T)$, and $M(\lambda S) = \lambda M(S)$.*

Theorem 4.4. *For $m, n \in \mathbb{N}$, $\mathbb{F}^{m,n}$ is a vector space of $\dim \mathbb{F}^{m,n} = mn$.*

Proposition 4.5. *Let $A \in \mathbb{F}^{m,n}$, $B \in \mathbb{F}^{n,p}$. Then*

$$\begin{aligned} (AB)_{jk} &= \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ \vdots & \ldots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ \vdots & \ldots & & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{np} \end{pmatrix} \\ &= \sum_{r=1}^{n} A_{jr}B_{rk}. \end{aligned}$$

Definition 4.6. Write $A_{j,\cdot}$ be the jth row of A. Let $B_{\cdot,k}$ be the kth column of B.

Proposition 4.7. *We have $(AB)_{\cdot,k} = AB_{\cdot,k}$.*

Proposition 4.8. *For $A \in \mathbb{F}^{m,n}, c \in \mathbb{F}^n$, $(Ac)_j = A_{j1}c_1 + \ldots + A_{jn}c_n$.*

Theorem 4.9. *The composition of two linear maps $S \in \mathcal{L}(V, W)$, $T \in \mathcal{L}(U, V)$ is given by $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$.*

Proof. Let $u_1, \ldots, u_p$ be a basis of U, $v_1, \ldots, v_n$ a basis of V, and $w_1, \ldots, w_m$ a basis of W. Write $A := \mathcal{M}(S, \{v_i\}, \{w_i\})$, and $B := \mathcal{M}(T, \{u_i\}, \{v_i\})$. Then we want to show $\mathcal{M}(ST, \{u_i\}, \{w_i\}) =$

$\mathcal{M}(S, \{v_i\}, \{w_i\})\mathcal{M}(T, \{u_i\}, \{v_i\})$. By definition,

$$\begin{aligned}
(ST)u_k &= S(Tu_k) \\
&= S\sum_{r=1}^{n} B_{rk}v_r \\
&= \sum_{r=1}^{n} B_{rk}(Sv_r) \\
&= \sum_{r=1}^{n} B_{rk}\left(\sum_{j=1}^{m} A_{jr}w_j\right) \\
&= \sum_{j=1}^{m}\left(\sum_{r=1}^{n} B_{rk}A_{jr}\right)w_j \\
&= (AB)_{jk}.
\end{aligned}$$

□

4.1.2 Discrete Dynamical Systems

Definition 4.10. Let $\{x_n\}$ be a sequence of vectors such that x_0 is defined as a starting vector, and $x_{k+1} = Ax_k$, for all $k \in \mathbb{N}$, for $A \in \mathbb{F}^{n,m}$, and $x_k \in \mathbb{F}^n$. Then $\{x_n\}$ is a discrete dynamical system, and we sometimes call $x_{k+1} = Ax_k$ the recurrence relation of the system.

Proposition 4.11. *We have for $\{x_n\}$ and a recurrence relation $x_{k+1} = Ax_k$, that $x_{k+1} = A^{k+1}x_0$.*

Example 4.12. Define $y_0 = y_1 = 1$. Define $y_{k+2} = y_{k+1} + y_k$ for all $k \in \mathbb{N}$. Let $x_0 = \begin{pmatrix} y_1 \\ y_0 \end{pmatrix}$. Define $x_{k+1} = \begin{pmatrix} y_{k+2} \\ y_{k+1} \end{pmatrix}$. In fact, we can write $x_{k+1} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} x_k$.

Now, since A is real and symmetric, let λ_1, λ_2 be eigenvalues of A corresponding to v_1, v_2 eigenvectors of A. Write $x_0 = c_1v_1 + c_2v_2$. Now,

$$\begin{aligned}
x_{k+1} = A^{k+1}x_0 &= A^{k+1}(c_1v_1 + c_2v_2) \\
&= c_1A^{k+1} + c_2A^{k+1}v_2 \\
&= c_1\lambda_1^{k+1}v_1 + c_2\lambda^{k+1}v_2.
\end{aligned}$$

4.2 Lecture 8. Thu Sep 19

4.2.1 Invertibility and Isomorphism

Definition 4.13. A linear map $T \in \mathcal{L}(V, W)$ is invertible if and only if there exists $S \in \mathcal{L}(W, V)$ such that $ST = I_V$, and $TS = I_W$. Then S is the inverse of T, written T^{-1}.

Definition 4.14. An invertible linear map is called an isomorphism. In particular, if there exists an invertible linear map from V to W, then $V \cong W$.

Theorem 4.15. *If $T \in \mathcal{L}(V, W)$ is invertible, then its inverse T^{-1} is unique.*

Proof. Let S_1, S_2 be inverses of T. Then $S_1 = S_1I = S_1TS_2 = IS_2 = S_2$. □

Theorem 4.16. *Let $T \in \mathcal{L}(V, W)$. Then T is invertible iff T is bijective.*

Proof. Let T be invertible. Then suppose $u, v \in V$, with $Tu = Tv$. Then $T^{-1}Tu = T^{-1}Tv \Rightarrow u = v$. Hence, T is injective. Let $w \in W$, then $w = TT^{-1}w \Rightarrow w \in \operatorname{range} T$. Now assume T is bijective. Then we can construct an inverse $S : W \to V$, where we define $Sw = v$, for $w \in W$, and $v \in T^{-1}(\{w\})$. However, $|T^{-1}(\{w\})| = 1$ since T is bijective, so v is unique, and S is well defined. It is easy to verify S is a linear map. For $v \in V$, $STv = Sw = v$, and $TSw = Tv = w$. Thus, $S = T^{-1}$. □

Theorem 4.17. *Two finite dimensional vector spaces are isomorphic if and only if they share the same dimension.*

Proof. Suppose V, W are isomorphic. Then let $T : V \to W$ be invertible. Then $\ker T = \{0\}$ and $\operatorname{range} T = W$. Then $\dim V = \dim T + \dim \operatorname{range} T \Rightarrow \dim V = \dim W$. In the other direction, suppose $\dim V = \dim W$. Then write $v_1, \ldots, v_n$ to be a basis of V and write $w_1, \ldots, w_n$ be a basis of W. Using the unique linear map lemme, let $T \in \mathcal{L}(V, W)$, so that $Tv_k = w_k$, $k \in [n]$. First, let $v \in \ker T$. Then $0 = T(v) = T(c_1v_1 + \ldots + c_nv_n) = c_1Tv_1 + \ldots + c_nTv_n = c_1w_1 + \ldots + c_nw_n \Rightarrow c_i$'s are zero $\Rightarrow v = 0 \Rightarrow T$ injective. Since $\dim V = \dim W$, and T is injective, T must be surjective. Hence, T is invertible, and V, W are isomorphic. □

Theorem 4.18. *If $\dim V < \infty, T \in \mathcal{L}(V)$, then T injective $\Leftrightarrow$ T surjective $\Leftrightarrow$ T invertible.*

5 Week 5

5.1 Lecture 9. Tue Sep 24

5.1.1 Polynomial Interpolation

Theorem 5.1. *Given $x_0, x_1, \ldots, x_n \in \mathbb{R}$, all distinct, and $y_0, y_1, \ldots, y_n$, there exists a unique polynomial $p(x)$ such that $p(x_j) = y_j$, for all $j \in \{0, 1, \ldots, n\}$.*

Proof. Suppose $T : \mathcal{P}_n \to \mathbb{F}^{n+1}$ so that $T(p) = \begin{pmatrix} p(x_0) \\ p(x_1) \\ \vdots \\ p(x_n) \end{pmatrix}$. It is clear that T is linear, and that $\dim \mathcal{P}_n = \dim \mathbb{F}^{n+1}$, so for T to be a bijection it is enough to show T is injective. Suppose $T(p) = 0$. Then $p(x_j) = 0$, for all $j \in \{0, 1, \ldots, n\}$. This implies that p has $n + 1$ distinct roots, but by the fundamental theorem of algebra, we know that p has at most n distinct roots. Therefore, p must be constant, in fact, p must be the 0 polynomial. Therefore, T is a bijection. □

5.1.2 Duality

Definition 5.2. A linear functional on V is a linear map from V to $\mathbb{F}$.

Definition 5.3. The dual of V, written V', is the set of all linear functionals from V to $\mathbb{F}$.

Theorem 5.4. *The dimension of $\dim V'$ is equal to $\dim V$ ($\dim V' = \dim V \dim \mathbb{F}$).*

Proposition 5.5. *The dual basis of a basis $v_1, \ldots, v_n$ of V is $\varphi_1, \ldots, \varphi_n \in V'$ so that $\varphi_k(v_j) = \delta_{kj}$.*

Example 5.6. The dual basis of $e_1, \ldots, e_n$ of $\mathbb{F}^n$ is such that $\varphi_k((x_1, \ldots, x_n)) = x_k$.

Theorem 5.7 (Representation theorem). *Let* $\varphi_1, \ldots, \varphi_n$ *be the dual basis of* $v_1, \ldots, v_k$. *Then* $\forall v \in V$, $v = \sum_{k=1}^{n} \varphi_k(v)v_k$. *Also,* $\forall f \in V'$, $f = \sum_{k=1}^{n} f(v_k)\varphi_k$.

Example 5.8 (Fourier representation of real-valued functions on $[-\pi, \pi]$). Suppose

$$f(x) = \sum A_k \underbrace{e^{inx}}_{v_k' \text{s}},$$

where

$$\underbrace{A_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-ikx}\,dx.}_{\varphi_k' \text{s}}$$

Proof. We will prove the first statement. For every $v \in V$, write $v = \sum_{i=1}^{n} c_i v_i$. Then

$$\begin{aligned}\varphi_k(v) &= \varphi\left(\sum_{i=1}^{n} c_i v_i\right)\\ &= \sum_{i=1}^{n} c_i \varphi_k(v_i)\\ &= c_k.\end{aligned}$$

□

Theorem 5.9 (Proof of Proposition 5.5). *The dual basis of a basis of* V *is a basis of* V'.

Proof. Given $v_1, \ldots, v_n$ a basis of V, suppose its dual basis is $\varphi_1, \ldots, \varphi_n$. Since $\dim V = \dim V'$, it suffices to show that $\varphi_1, \ldots, \varphi_n$ is linearly independent. Write $\sum_{k=1}^{n} a_k\varphi_k = \mathbf{0}$. Applying this to v_j, we find $0 = \mathbf{0}(v_j) = \sum_{k=1}^{n} a_k\varphi_k(v_j) = a_j$. □

Definition 5.10. The dual map of $T \in \mathcal{L}(V, W)$ is the map $T' \in \mathcal{L}(W', V')$ such that $T'(\varphi) = \varphi \circ T$, for all $\varphi \in W'$.

Theorem 5.11. *Suppose* $\dim V \leq \dim W < \infty$, *and* $T \in \mathcal{L}(V, W)$. *Then* $\mathcal{M}(T') = \mathcal{M}(T)^T$

5.1.3 Polynomials

Note. Recall that for a complex number $z = a + bi$, its conjugate $\overline{z} = a - bi$. Also, $|z| = \sqrt{\operatorname{Re} z^2 + (\operatorname{Im} z)^2}$

Proposition 5.12. *If* $z, w \in \mathbb{C}$,

1. $z + \overline{z} = 2\operatorname{Re} z$ *and* $z - \overline{z} = 2\operatorname{Im} z$,
2. $|\operatorname{Re} z| \leq |z|, |\operatorname{Im} z| \leq |z|$,
3. $|\overline{z}| = |z|$,

4. $|zq| = |z||w|$,
5. $|z + w| \le |z| + |w|$.

Theorem 5.13. *Let $p \in \mathcal{P}(\mathbb{F})$ be a polynomial of degree $m \in \mathbb{N}_0$, and $\lambda \in \mathbb{F}$. Then $p(\lambda) = 0 \iff \exists\, q \in \mathcal{P}(\mathbb{F}), \deg q = m - 1$, such that $p(z) = (z \cdot \lambda)q(z), \forall z \in \mathbb{F}$.*

Theorem 5.14. *If $p, s \in \mathcal{P}(\mathbb{F})$, with $s \neq 0$, then there exists unique $q, r \in \mathcal{P}(\mathbb{F})$, such that $p = sq + r$, with $\deg r < \deg s$.*

Theorem 5.15 (Fundamental theorem of algebra)**.** *Every nonconstant polynomial $p \in \mathcal{P}(\mathbb{C})$ has $\deg p$ roots in $\mathbb{C}$.*

Theorem 5.16. *If $p \in \mathcal{P}(\mathbb{C})$ with real coefficients satisfies $p(\lambda) = 0$, for $\lambda \in \mathbb{C}$ if and only if $p(\overline{\lambda})$.*

Proof. For $p(x) = \sum_{j=0}^{n} a_j z^j$ satisfying $p(\lambda) = 0$, we have

$$\begin{aligned} 0 = \overline{0} = \overline{p(\lambda)} &= \overline{\sum_{j=1}^{n} a_j \lambda^j} \\ &= \sum \overline{a_j}\,\overline{\lambda}^j \\ &= p(\overline{\lambda}). \end{aligned}$$

□

Theorem 5.17. *A non-constant polynomial $p \in \mathcal{P}(\mathbb{R})$ has a unique factorization into at most quadratic factors, where each quadratic factor has discriminant less than zero (irreducible over $\mathbb{R}$).*

Proof of triangle inequality over $\mathbb{C}$ (Item 5). We know for $z, w \in \mathbb{C}$,

$$\begin{aligned} |z + w|^2 &= (z + w)\overline{(z + w)} \\ &= |z|^2 + z\overline{w} + w\overline{z} + |w|^2 \\ &= |z|^2 + 2\,\mathrm{Re}(z\overline{w}) + |w|^2 \\ &\le |z|^2 + 2|\,\mathrm{Re}(z\overline{w})| + |w|^2 \\ &\le |z|^2 + 2|z\overline{w}| + |w|^2 \\ &= (|z| + |w|)^2. \end{aligned}$$

□

5.2 Lecture 10. Thu Sep 26

5.2.1 Determinants (Linear Algebra Done Wrong)

Some motivation: for n vectors in $\mathbb{R}^n$, $v_1, v_2, \ldots, v_n$, the determinant

$$D(v_1, \ldots, v_n) = \det \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}$$

represents the volume of the parallelepiped spanned by the v_i's. The idea is that the height of this parallelepiped is the distance of one vector to the subspace spanned by the other v_i's, and the base of this parallelepiped is the $(n-1)$-dim volume of the parallelepiped spanned by the remaining vectors. The sign of this determinant tells us about the orientation of these vectors. Ideally, we want this determinant to have linearity in each vector: that is

1. $D(v_1, \ldots, v_k + u_k, \ldots, v_n) = d(v_1, \ldots, v_k, \ldots v_n) + d(v_1, \ldots, u_k, \ldots, v_n)$,
2. $D(v_1, \ldots, \alpha v_k, \ldots, v_n) = \alpha D(v_1, \ldots, v_k, \ldots, v_n)$,
3. $D(v_1, \ldots, v_j + \alpha v_k, v_k, v_n) = D(v_1, \ldots, v_j, \ldots, v_n)$.

Note that 2 and 3 antisymmetry:

4. $D(v_1, \ldots, v_j, \ldots, v_j, v_n) = -D(v_1, \ldots, v_k, \ldots, v_j, \ldots, v_n)$.

To show this, we have

$$\begin{aligned} D(\ldots, v_j, \ldots, v_k, \ldots) &= D(\ldots, v_j, \ldots, v_k - v_j, \ldots) && (3) \\ &= D(\ldots, v_j + v_k - v_j, \ldots, v_k - v_j) && (3) \\ &= D(\ldots, v_k, \ldots, v_k - v_j, \ldots) && (3) \\ &= D(\ldots, v_k, \ldots, -v_j) && (2) \\ &= -D(\ldots, v_k, \ldots, v_j). && (2) \end{aligned}$$

We also want something called normalization:

5 $\det(I) = D(e_1, \ldots, e_n) = 1$.

We now want to construct a function which has these properties. We would also like to show that this function is unique. Assume we have a function D that satisfies $1, 3$, and 5.

- If A has a zero column, then $\det A = 0$ (prove with 2).
- If A has two equal columns, then $\det A = 0$ (prove with 3).
- If one column of A is a scalar multiple of another, then $\det A = 0$ (prove with 3).
- If the columns of A are linearly dependent, then $\det A = 0$ (prove with 3).

Proposition 5.18. *The function D we have described is invariant under adding to one column a linear combination of the others.*

Proposition 5.19. *The determinant of a diagonal matrix is simply the product of its diagonal entries.*

Corollary 5.19.1. *The determinant of an upper triangular matrix is the product of its diagonal entries.*

Theorem 5.20. *A matrix A has determinant 0 if and only if A is not invertible.*

Through some calculations, we conclude that

$$\det A = \sum_{j=1}^{n} A_{j,1}(-1)^{j+1} \det \Big[co(A_{j,1}) \Big],$$

where $co(A_{j,1})$ is the $(n-1) \times (n-1)$ principal submatrix obtained from A by deleting the first column and jth row of A. We can conclude the cost of calculating the determinant of an $n \times n$ matrix (using only repeated co-factorization) is $O(n!)$.

6 Week 6

Exam week. But I'll put notes about things not covered in Axler here.

6.1 Column Rank Equals Row Rank

Theorem 6.1 (Matrix multiplication as linear combinations of columns)**.** *Suppose C is an m-by-n matrix, and R is a c-by-n matrix.*

1. *If $k \in [n]$, then column k of CR is a linear combination of the columns of C, with the coefficients of this linear combination coming from column k of R.*
2. *If $j \in [m]$, then row j of CR is a linear combination of the rows of R, with the coefficients of this linear combination coming from row j of C.*

Proof. Let $k \in [n]$. Then column k of CR equals $CR_{\cdot,k}$, which equals the linear combination of the columns of C with coefficients coming from $R_{\cdot,k}$. Thus, 1 holds. The proof for 2 is similar. □

Definition 6.2. Let A be an m-by-n matrix with entries in $\mathbb{F}$.

1. The *column rank* of A is the dimension of the span of the columns of A in $\mathbb{F}^{m,1}$.
2. The *row rank* of A is the dimension of the span of the rows of A in $\mathbb{F}^{1,n}$.

Note. If A is an m-by-n matrix, then the column rank of A is at most n, and the column rank of A is also at most m (because $\dim \mathbb{F}^{m,1} = 1$). Similarly, the row rank of A is also at most $\min\{m, n\}$.

Definition 6.3. The *transpose* of the matrix A, denoted by A^t, is the matrix obtained from A by interchanging rows and columns. Specifically, if A is an m-by-n matrix, then A^t is the n-by-m matrix whose entries are given by the equation

$$(A^t)_{k,j} = A_{j,k}.$$

Proposition 6.4. *The transpose has some pretty nice algebraic properties:*

1. $(A + B)^t = A^t + B^t$,
2. $(\lambda A)^t = \lambda A^t$,
3. $(AC)^t = C^t A^t$,

for all m-by-n matrices A, B, all $\lambda \in \mathbb{F}$, and all n-by-p matrices C.

Theorem 6.5 (Column rank factorization)**.** *The following is the key towards proving that column rank is equal to row rank. Suppose A is an m-by-n matrix with entries in $\mathbb{F}$, and column rank $c \geq 1$. Then there exists an m-by-c matrix C and a c-by-n matrix R, both with entries in $\mathbb{F}$ such that $A = CR$.*

Proof. Each column of A is an m-by-1 matrix. The list $A_1, \ldots, A_n$ of columns of A can be reduced to a basis of the span of the columns of A. Let the length of this basis be c. Put the c columns of this basis into the m-by-c matrix C. If $k \in [n]$, then the column k of A is a linear combination of the columns of C. Make the coefficients of this linear combination into column k of a c-by-n matrix that we call R. Then $A = CR$. □

Theorem 6.6 (Column Rank Equals Row Rank)**.** *Suppose $A \in \mathbb{F}^{m,n}$. Then the column rank of A equals the row rank of A.*

Proof. Let c denote the column rank of A. Let $A = CR$ be the column-row factorization of A, where C is an m-by-c matrix and R is a c-by-n matrix. Then Theorem 6.1.2 tells us that every row of A is a linear combination of the rows of R. Because R has c rows, this implies that the row rank of A is less than or equal to the column rank c of A. To prove the inequality in the other direction, apply the result in the previous paragraph to A^t, getting

$$\begin{aligned}\text{column rank of } A &= \text{row rank of } A^t \\ &\leq \text{column rank of } A^t \\ &= \text{row rank of } A.\end{aligned}$$

Thus the column rank of A equals the row rank of A. □

Definition 6.7. Because the column rank of A equals the row rank of A, we can simply dispense of the terms column rank and row rank, and use the term rank of A. Define the rank of A to be the column rank of A.

6.2 Change of Basis

Theorem 6.8 (Matrix of product of linear maps)**.** *Suppose $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$. If $u_1, \ldots, u_m$ is a basis of U, $v_1, \ldots, v_n$ is a basis of V, and $w_1, \ldots, w_p$ is a basis of W. Then*

$$\mathcal{M}(ST, (u_1, \ldots, u_m), (w_1, \ldots, w_p)) = \mathcal{M}(S, (v_1, \ldots, v_n), (w_1, \ldots, w_p))\mathcal{M}(T, (u_1, \ldots, u_m), (v_1, \ldots, v_n)).$$

Theorem 6.9 (Matrix of the identity with respect to two bases)**.** *Let $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ be bases of V. Then the matrices*

$$\mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)), \text{ and } \mathcal{M}(I, (v_1, \ldots, v_n), (u_1, \ldots, u_n)),$$

are invertible, and each is the inverse of the other.

Proof. In Theorem 6.8, replace w_k with u_k and replace S and T with I, yielding

$$I = \mathcal{M}(I, (v_1, \ldots, v_n), (u_1, \ldots, u_n))\mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)).$$

Interchanging the role of the u's and the v's yields

$$I = \mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n))\mathcal{M}(I, (v_1, \ldots, v_n), (u_1, \ldots, u_n)).$$

These two equations above give the desired result. □

Theorem 6.10 (Change-of-basis formula)**.** *Suppose $T \in \mathcal{L}(V)$. Suppose $u_1, \ldots, u_n$, $v_1, \ldots, v_n$ are bases of V. Let*

$$A = \mathcal{M}(T, (u_1, \ldots, u_n)) \text{ and } B = \mathcal{M}(T, (v_1, \ldots, v_n)),$$

and $C = \mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n))$. Then $A = C^{-1}BC$.

Proof. In Theorem 6.8, replace w_k with u_k and replace S with I. Then

$$A = C^{-1}M(T, (u_1, \dots, u_n), (v_1, \dots, v_n)),$$

where we apply the result of Theorem 6.9. Again, use Theorem 6.8, this time replacing w_k with v_k, and instead, replace T with I, and replace S with T, yielding

$$M(T, (u_1, \dots, u_n), (v_1, \dots, v_n)) = BC.$$

Then $A = C^{-1}BC$. □

Theorem 6.11 (Matrix of inverse is inverse of matrix)**.** *Suppose that $v_1, \dots, v_n$ is a basis of V and $T \in \mathcal{L}(V)$ is invertible. Then $M(T^{-1}) = M(T)^{-1}$, where both matrices are with respect to the basis $v_1, \dots, v_n$.*

7 Week 7

7.1 Lecture 11. Tue Oct 8

7.1.1 Eigenvalues

Definition 7.1 (Invariant Subspace)**.** Let $T \in \mathcal{L}(V)$. Then $U \leq V$ is invariant under T iff when $u \in U$, we have $Tu \in U$.

Note. Does every bounded linear operator on a complex Banach space have a non-trivial (closed) invariant subspace?

1. Yes for $n < \dim V < \infty$.
2. No for restricted operator classes.
3. Yes for certain Banach spaces (1975).
4. Results as recent as 2023.

Definition 7.2. Let $T \in \mathcal{L}(V)$. We say λ is an eigenvalue of T if and only if there exists $v \in V$, $v \neq 0$, such that $Tv = \lambda v$, for some $\lambda \in \mathbb{F}$. We also have the following equivalences:

$$\begin{aligned} &Tv = \lambda v \\ &\Leftrightarrow \det(T - \lambda I) = 0 \\ &\Leftrightarrow T - \lambda I \text{ non-invertible} \\ &\Leftrightarrow T - \lambda I \text{ non-injective} \\ &\Leftrightarrow T - \lambda I \text{ non-surjective} . \end{aligned}$$

Theorem 7.3. *Distinct eigenvalues of a linear map T have linearly independent eigenvectors.*

Proof. Suppose $v_1, \dots, v_n$ are linearly dependent eigenvectors of T. Let j be the smallest possible index such that

$$v_j = c_1v_1 + \dots + c_{j-1}v_{j-1}.$$

Then

$$\lambda_j v_j = \lambda_j c_1 v_1 + \ldots + \lambda_j c_{j-1} v_{j-1}.$$

But we also have

$$T v_j = \lambda v_j = \lambda_1 c_1 v_1 + \ldots + \lambda_{j-1} c_{j-1} v_{j-1}.$$

Then we know

$$0 = (\lambda_j - \lambda_1) c_1 v_1 + \ldots + (\lambda_j - \lambda_{j-1}) c_{j-1} v_{j-1}.$$

Because we said v_j was the smallest index such that $v_j \in \operatorname{span}(v_1, \ldots, v_{j-1})$, we know that $(\lambda_j - \lambda_k) c_k = 0$, for all $1 \leq k \leq j-1$. Since $\lambda_j - \lambda_k$ is nonzero, this forces the c_k's to be zero, and therefore, v_j is zero, a contradiction. □

Corollary 7.3.1. *The number of distinct eigenvalues of $T \in \mathcal{L}(V) \leq \dim V$.*

7.1.2 Diagonalization

Definition 7.4. We say $T \in \mathcal{L}(V)$ is diagonalizable if and only if there exists a basis of V of eigenvectors.

Corollary 7.4.1. *For $T \in \mathcal{L}(V)$, $\dim V < \infty$, we say T is diagonalizable if the number of distinct eigenvalues of of T is equal to $\dim V$.*

Proposition 7.5. *Let $\dim V < \infty$, and let $B = \{v_1, \ldots, v_n\}$ be a basis of eigenvectors of $T \in \mathcal{L}(V)$. Then $\mathcal{M}(T, B, B)$ is a diagonal matrix whose entries are the eigenvalues λ_j corresponding to the eigenvectors v_j, $1 \leq j \leq n$.*

Theorem 7.6. *If $\mathcal{M}(T, B, B)$ is diagonal if and only if B is a basis of eigenvectors.*

Exercise 7.1. Let $\dim V < \infty$, $T \in \mathcal{L}(V, W)$, $U \leq V$ invariant under T. Let $B = \{u_1, \ldots, u_n\}$ be a basis of U, consisting of eigenvectors of T. If we extend B to a basis of V, call it B', what is $\mathcal{M}(T, B', B')$.

Example 7.7 (Jordan Normal Forms)**.** Diagonalizable linear maps make certain problems very easy: matrix exponentiation, solving for null spaces, solving general $Ax = b$ problems. If it's not possible to diagonalize a linear map, how can one get a matrix which is "as diagonal as possible?" This is the motivation for studying Jordan normal forms.

7.2 Lecture 12. Thu Oct 10

7.2.1 Existence of Eigenvalues

Definition 7.8 (Matrix Exponentiation)**.** Let $T \in \mathcal{L}(V)$, then

$$T^k = \begin{cases} T \cdot \ldots \cdot T & k \in \mathbb{N} \\ I & k = 0 \\ (T^{-1})^{-k} & -k \in \mathbb{N}. \end{cases}$$

Now, we can evaluate polynomials $p \in \mathcal{P}(\mathbb{F})$ at T by setting

$$p(T) = \sum_{k=0}^{n} a_k T^k.$$

Proposition 7.9. *Let $p, q \in \mathcal{P}(\mathbb{F})$, and let $T \in \mathcal{L}(V)$. Then $(pq)(T) = p(T)q(t)$, and $p(T)q(T) = q(T)p(T)$.*

Proof. Suppose $p(z) = \sum a_k z^k$, and $q(z) = \sum b_r z^r$. Then $(pq)(z) = \sum_k \sum_r a_k b_r z^{k+r}$. Then

$$\begin{aligned}(pq)(T) &= \sum_k \sum_r a_k b_r T^{k+r} \\ &= \Big(\sum_k a_k T^k\Big)\Big(\sum_r b_r T^r\Big) \\ &= p(T)q(T).\end{aligned}$$

Also,

$$p(T)q(T) = (pq)(T) = (qp)(T) = q(T)p(T).$$

□

Note. In general, if $T, S \in \mathcal{L}(V)$, $p(S)q(T) \neq q(T)p(S)$.

Theorem 7.10. *Let $1 \leq \dim V < \infty$, and $T \in \mathcal{L}(V)$, where $\mathbb{F} = \mathbb{C}$. Then T has at least one eigenvalue.*

Proof. Fix $v \in V$, $0 \neq v$. Then consider $v, Tv, T^2v, \ldots, T^n v$. These $(n+1)$ vectors are linearly dependent. So we can write

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

such that the a_i's are not all zero. We know that if $a_1, \ldots, a_n = 0$, then $a_0 v = 0$, forcing $v = 0$. Therefore, we can assume at least one $a_1, \ldots, a_n$ is nonzero. Consider $p(T)v := a_0 v + a_1 Tv + \ldots + a_n T^n v$. Then we can say that $p(T)$ is a nonconstant polynomial of T. Then by the Fundamental Theorem of Algebra,

$$p(z) = \sum_{k=0}^{n} a_k z^k = c \prod_{j=1}^{m} (z - \lambda_j).$$

In particular, $m = \deg p \geq 1$, so that

$$0 = p(T)v = \Big(c \prod_{j=1}^{m} (T - \lambda_j I)\Big) v.$$

Since $v \neq 0$, there is some j such that $T - \lambda_i I$ is not injective, therefore, non invertible. (A better proof is provided below.) This implies that λ_j is an eigenvalue of T. □

Definition 7.11. We call $T \in \mathcal{V}, \mathcal{W}$ nonsingular if $\ker T = \{0\}$, and singular if $\ker T \neq \{0\}$.

Proposition 7.12. *If $\dim V = \dim W < \infty$, then T is singular iff it is non-invertible.*

Proposition 7.13. *Let $\dim V < \infty$, and let $S, T \in \mathcal{L}(V)$. Then ST is singular iff S or T is singular.*

Proof. $(\Leftarrow)$ If T is singular, let $v \in \ker T$, $v \neq 0$. Then $ST(v) - S(0) = 0 \Rightarrow v \in \ker ST$. If S is singular, then S is not surjective. Therefore, ST is not surjective, therefore, it is not injective, thus, it is singular. $(\Rightarrow)$ If S, T are nonsingular, then S, T are invertible. Hence, we can show ST is invertible, thus, nonsingular. □

Proposition 7.14. *If $S_1, \dots, S_k \in \mathcal{L}(V)$, and $\dim V < \infty$, then*

$$\prod_{j=1}^{k} S_j \text{ singular } \Leftrightarrow \text{ at least one of } S_j \text{ is singular.}$$

Note. Axler takes a different approach to proving Theorem 7.10. Let $\dim V < \infty, T \in \mathcal{L}(V)$. Then if T has an upper triangular matrix representation with respect to some basis, then T is invertible if and only if all the diagonal entries of this matrix are nonzero. Furthermore, all the eigenvalues of T are on the diagonal of this matrix.

Exercise 7.2. Let $T \in \mathcal{L}(V)$, and suppose $T^n = 0$, for some $n \in \mathbb{N}$. Show that $(I - T)^{-1} = I + T + T^2 + \dots + T^{n-1}$. Explain why someone might be able to guess the formula.

Solution. Computation, and notice that the sum is just a geometric sum. □

8 Week 8

8.1 Lecture 13. Tue Oct 15

8.1.1 Upper Triangular Matrices

Theorem 8.1. *Let $\dim V < \infty$, and $T \in \mathcal{L}(V)$, $\mathbb{F} = \mathbb{C}$. Then T is has an upper triangular matrix representation with respect to some basis $\mathcal{B} = v_1, \dots, v_n$ of V.*

Proposition 8.2. *We claim that $\mathcal{M}_{\mathcal{B}}(T)$ is upper triangular iff*

$$Tv_j \in \operatorname{span}(v_1, \dots, v_j).$$

This happens iff $\operatorname{span}(v_1, \dots, v_j)$ is invariant, for all $1 \le j \le n$. Also, if T does not have an eigenvalue, then T has no upper triangular representation.

Proof of Theorem 8.1. Note that every element of $\mathbb{F}^{1,1}$ is upper triangular. Assume for all vector spaces V, $\dim V < n$, our conclusion holds. Let V be a vector space of dimension n. Let $T \in \mathcal{L}(V)$, and suppose λ_1, v_1 is an eigenvalue-eigenvector pair. Extend v_1 to a basis of V using vectors $u_2, \dots, u_n$. Then $V = \operatorname{span}(u_2, \dots, u_n) \oplus \operatorname{span}(v_1)$. Consider the projection $P_u : V \to U$, such that $P_u(e + u) = u$. Then we claim $P_u \in \mathcal{L}(V, U)$. Consider the compression $P_u T|_u = S : U \to U$, which we claim is in $\mathcal{L}(U)$. Since $\dim U = n - 1$, we know S has an upper triangular representation with respect to a basis $\overline{\mathcal{B}} = \{v_2, \dots, v_n\}$. Let $\mathcal{B} = \{v_1\} \cup \overline{\mathcal{B}}$. Then the matrix representation of T with respect to $\mathcal{B}$ is

$$\mathcal{M}_{\mathcal{B}}(T) = \left(\begin{array}{c|ccc} \lambda_1 & 0 & \dots & 0 \\ \hline 0 & & & \\ \vdots & & \mathcal{M}_{\overline{\mathcal{B}}}(S) & \\ 0 & & & \end{array}\right).$$

□

Theorem 8.3. *Let $\dim V = n < \infty, T \in \mathcal{L}(V)$, and that T has an upper triangular representation. Then T is invertible iff the diagonal entries of the upper triangular representation are all nonzero.*

Proof. Assume that A is an upper triangular matrix of T. Then suppose for some k, $1 \leq k \leq n$, that $A_{k,k} = 0$. Then $Tv_k = A_{1,k}v_1 + \ldots + A_{k-1}kv_{k-1}$. Therefore, $Tv_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$. Since A is upper triangular, $Tv_j \in \operatorname{span}(v_1, \ldots, v_{k-1})$, for all $1 \leq j \leq k-1$. This forces $Tv_1, \ldots, Tv_{k-1}, Tv_k$ to be linearly dependent. Therefore, T is not injective, hence, it is not invertible. For the converse, assume that for all k, $1 \leq k \leq n$, that $A_{k,k} \neq 0$. Hence, $Tv_1 = A_{1,1}v_1 \neq 0$. Therefore, Tv_1 is linearly independent (since it is nonzero). Assume that for $k < n$, $Tv_1, \ldots, Tv_k$ are linearly independent. Then since A is upper triangular, $Tv_j \in \operatorname{span}(v_1, \ldots, v_j)$, $1 \leq j < n$. Consider that since A is upper triangular, that $Tv_n = A_{1,n}v_1 + \ldots + A_{n-1,n}v_{n-1} + A_{n,n}v_n$. Therefore, $Tv_n \notin \operatorname{span}(v_1, \ldots, v_{n-1})$. Hence, $Tv_1, \ldots, Tv_n$ are linearly independent. This implies that $\dim \operatorname{range}(T) \geq n$, which implies T is surjective, hence, invertible. □

Example 8.4. Consider $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T \in \mathcal{L}(\mathbb{R}^2)$ with standard matrix representation

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Show that A has no eigenvalues over $\mathbb{R}$.

Example 8.5. Calculate the eigenvalues and some corresponding eigenvectors of

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

8.1.2 Eigenspaces

Definition 8.6. Let $T \in \mathcal{L}(V)$, and let λ be an eigenvalue of T. We call $E(\lambda, T) = \{v \in V \mid Av = \lambda v\}$ the eigenspace of λ under T. We call $\dim E(\lambda, T)$ the geometric multiplicity of T corresponding to λ.

Note. Note that $0 \in E(\lambda, T)$, so not all $v \in E(\lambda, T)$ are eigenvectors of T corresponding to λ.

Theorem 8.7. *Let $T \in \mathcal{L}(V)$, and $\lambda_1, \ldots, \lambda_n$ be distinct eigenvalues of T. Then*

$$E(\lambda_1, T) \oplus E(\lambda_2, T) \oplus \ldots \oplus E(\lambda_n, T) := E \leq V$$

is a direct sum, and

$$\sum_{j=1}^{n} \dim E(\lambda_j, T) \leq \dim V.$$

When there is equality, then recall that T is diagonalizable.

8.1.3 Minimal Polynomial

Theorem 8.8. *If $\dim V < \infty$, and $T \in \mathcal{L}(V)$, then there exists a unique monic polynomial $p \in \mathcal{P}(\mathbb{F})$ of smallest degree such that $p(T) = 0$. We call this the minimal polynomial of T.*

Proof. If $\dim V = 0$, then we'll set all operators on V to be the zero operator on V. Take $p = 1$, and note that $p(T) = 0$, for all $T \in \mathcal{L}(V)$. Assume the statement holds true for all V, $\dim V = k < n$. Consider V such that $\dim V = n$. Let $0 \neq v \in V$. Then let $t \in \mathcal{L}(V)$. We know that

$v, Tv, T^2v, \ldots, T^nv$ is linearly dependent. Let m be the smallest integer such that $c_0Iv + c_1Tv + c_2T^2v + \ldots + c_{m-1}T^{m-1}v + T^mv = 0$. Define

$$p(z) := c_1 + c_1z + c_2z^2 + \ldots + c_{m-1}z^{m-1} + z^m.$$

Then $p(T)v = 0$, and for all $k \in \mathbb{N}$, $p(T)T^kv = 0$, since $p(T)T^kv = T^kp(T)v = T^k0 = 0 \Rightarrow T^kv \in \ker p(T)$. But since m is the smallest integer such that our above equation held, $v, Tv, T^2v, \ldots, T^{m-1}v$ are linearly independent. Therefore $\dim\ker p(T) \geq m$, and

$$\begin{aligned} \dim \operatorname{range} p(T) &= \dim V - \dim\ker p(T) \\ &\leq n - m < n. \qquad (v \neq 0) \end{aligned}$$

We know $\operatorname{range} p(T)$ is invariant under T, since range T is. Applying the induction hypothesis to $T|_{\operatorname{range} T}$ shows us that there exists a unique monic polynomial $s \in \mathcal{P}(\mathbb{F})$ which has $\deg s \leq n - m$, and satisfies $s(T|_{\operatorname{range} p(T)}) = 0$. Also, for all $w \in V$,

$$\begin{aligned} (sp)(T)w &= s(T) \overbrace{p(T)w}^{\in \operatorname{range} p(T)} \\ &= 0. \end{aligned}$$

It remains to show that this is the unique monic polynomial of smallest degree, which is not difficult to show by contradiction. □

Theorem 8.9. *If $\dim V < \infty$, $T \in \mathcal{L}(V)$, then the zeroes of the minimal polynomial of T are the eigenvalues of T.*

8.2 Lecture 14. Thu Oct 17

8.2.1 Norms and Inner Products

Definition 8.10. Let V be a vector space over $\mathbb{R}$. A norm on V is a real valued function taking $v \mapsto \|v\|, v \in V$, which satisfies the following properties:

1. $\|v\| \geq 0, \forall v \in V, \|v\| = 0 \Leftrightarrow v = 0$,
2. $\|\alpha v\| = |\alpha|\|v\|, \forall v \in V, \forall \alpha \in \mathbb{R}$,
3. $\|u + v\| \leq \|u\| + \|v\|, \forall u, v \in V$.

Example 8.11.
1. Consider $\|x\|_1 = \sum_{i=1}^n |x_i|, \forall x \in \mathbb{R}^n$.
2. The geometric norm, the Euclidean norm, $\|x\|_2 = \left[\sum_{i=1}^n x_i^2\right]^{\frac{1}{2}}$.
3. The infinity norm, $\|x\|_\infty = \max\{|x_i| \mid 1 \leq i \leq n\}$.
4. Consider $C[a, b]$. Then we have the following norms:

- $\|f\|_{L^1} = \int_a^b |f|$,
- $\|f\|_{L^2} = \left[\int_a^b f^2\right]^{\frac{1}{2}}$,
- $\|f\|_{L^\infty} = \max\{|f(t)| \mid a \leq t \leq b\}$.

Note. We can always define a metric on V given a norm by $d(u,v) = \|u-v\|$. The most convenient, useful, interesting, elegant norms are defined by inner products.

Definition 8.12. Let V be a vector space over $\mathbb{R}$. An inner product on V is a real-valued function on $V \times V$ such that $(u,v) \mapsto \langle u, v\rangle$, which satisfies the following properties:

1. $\langle u, u\rangle \geq 0, \forall u \in V, \langle u, u\rangle = 0 \Leftrightarrow u = 0$,
2. $\langle u, v\rangle = \langle v, u\rangle, \forall u, v \in V$,
3. $\langle \alpha u + \beta v, w\rangle = \alpha\langle u, v\rangle + \beta\langle v, w\rangle, \forall u, v, w \in V, \forall \alpha, \beta \in \mathbb{R}$,

We say that inner products are positive definite symmetric bilinear forms.

Example 8.13.
1. On $\mathbb{R}^n$, the dot product is an inner product, that is,

$$x \cdot y = \sum_{j=1}^{n} x_j y_j,\ \forall x, y \in \mathbb{R}^n.$$

2. On $C[a,b]$, the L^2 inner product is an inner product:

$$\langle f, g\rangle_{L^2} = \int_a^b fg,\ \forall f, g \in C[a,b].$$

Lemma 8.14. *Let V be a vector space over $\mathbb{R}$, and let $\langle \cdot, \cdot\rangle$ be an inner product on V. Then*

$$\langle v, 0\rangle = \langle 0, v\rangle = 0,\ \forall v \in V.$$

Proof. We have

$$\langle 0, v\rangle = \langle \overbrace{0}^{\in\mathbb{R}} \cdot \overbrace{0}^{\in V}, v\rangle = 0\langle 0, v\rangle = 0.$$

□

Theorem 8.15 (Cauchy-Schwarz)**.** *Let V be an inner product space over $\mathbb{R}$. Then*

$$|\langle u, v\rangle| \leq \left[\langle u, u\rangle\right]^{\frac{1}{2}} \left[\langle v, v\rangle\right]^{\frac{1}{2}}, \forall u, v \in V,$$

with equality iff one of u, v is a multiple of the other.

Proof. First note that if $u = 0$ or $v = 0$, then the result holds. Assume $u, v \neq 0$. Now assume that $\langle u, u\rangle = \langle v, v\rangle = 1$. We must prove that $|\langle u, v\rangle| \leq 1$. We know

$$\begin{aligned} \langle u - v, u - v\rangle &\geq 0 \\ \langle u, u - v\rangle - \langle v, u - v\rangle &\geq 0 \\ \langle u, u\rangle - \langle u, v\rangle - \langle v, u\rangle + \langle v, v\rangle &\geq 0 \\ 2 - \langle u, v\rangle &\geq 0 \\ \Rightarrow \langle u, v\rangle &\leq 1. \end{aligned}$$

Also,

$$\begin{aligned}\langle u+v, u+v\rangle \geq 0 &\Rightarrow \langle u,u\rangle + 2\langle u,v\rangle + \langle v,v\rangle \geq 0\\ &\Rightarrow 2 + 2\langle u,v\rangle \geq 0\\ &\Rightarrow \langle u,v\rangle \geq -1.\end{aligned}$$

Therefore,

$$-1 \leq \langle u,v\rangle \leq 1 \Rightarrow |\langle u,v\rangle| \leq 1.$$

Finally, consider the general case that $u \neq 0, v \neq 0$. Consider

$$u' = \frac{1}{\langle u,u\rangle^{\frac{1}{2}}}u, \quad v' = \frac{1}{\langle v,v\rangle^{\frac{1}{2}}}v.$$

Then

$$\langle u', v'\rangle = \left\langle \frac{1}{\langle u,u\rangle^{\frac{1}{2}}}u, \frac{1}{\langle u,u\rangle^{\frac{1}{2}}}u \right\rangle = \frac{1}{\langle u,u\rangle}\langle u,u\rangle = 1.$$

Similarly, $\langle v', v'\rangle = 1$. Thus, $|\langle u', v'\rangle| \leq 1$, which implies

$$\begin{aligned}&\left|\left\langle \frac{1}{\langle u,u\rangle^{\frac{1}{2}}}u, \frac{1}{\langle v,v\rangle^{\frac{1}{2}}}v \right\rangle\right| \leq 1\\ &\Rightarrow \frac{1}{\langle u,u\rangle^{\frac{1}{2}}\langle v,v\rangle^{\frac{1}{2}}}|\langle u,v\rangle| \leq 1\\ &\qquad \Rightarrow |\langle u,v\rangle| \leq \langle u,u\rangle^{\frac{1}{2}}\langle v,v\rangle^{\frac{1}{2}}.\end{aligned}$$

Part two: equality iff one of u, v is a multiple of the other. First if $\langle u,u\rangle = \langle v,v\rangle = 1$, then

$$\begin{aligned}\langle u-v, u-v\rangle = 0 \Leftrightarrow u - v = 0 \Leftrightarrow u = v\\ \langle u+v, u+v\rangle = 0 \Leftrightarrow u + v = 0 \Leftrightarrow u = -v.\end{aligned}$$

So for vectors satisfying $\langle u,u\rangle = \langle v,v\rangle = 1$, we get equality iff $u = \pm v$. For general nonzero vectors u, v, we get equality iff

$$\begin{aligned}&u' = \pm v'\\ \Leftrightarrow\ &\frac{1}{\langle u,u\rangle^{\frac{1}{2}}}u = \pm\frac{1}{\langle v,v\rangle^{\frac{1}{2}}}v\\ &\Leftrightarrow u = \pm\frac{\langle u,u\rangle^{\frac{1}{2}}}{\langle v,v\rangle^{\frac{1}{2}}}v = \alpha v.\end{aligned}$$

□

Theorem 8.16. *Let V be an inner product space over $\mathbb{R}$, and define $\|v\| = \sqrt{\langle v,v\rangle}, \forall v \in V$. Then $\|\cdot\|$ is a norm on V.*

Proof. First,

$$\forall v \in V, \langle v,v\rangle \geq 0 \Rightarrow \sqrt{\langle v,v\rangle} \geq 0 \Rightarrow \|v\| \geq 0,$$

and

$$(\langle v,v\rangle = 0 \Leftrightarrow v = 0) \Rightarrow (\|v\| = \sqrt{\langle v,v\rangle} = 0 \Leftrightarrow v = 0).$$

Next,

$$\begin{aligned}
\forall v \in V, \alpha \in \mathbb{R}, \|\alpha v\| &= \sqrt{\langle \alpha v, \alpha v\rangle} \\
&= \sqrt{\alpha^2 \langle v, v\rangle} \\
&= \sqrt{\alpha^2}\sqrt{\langle v, v\rangle} \\
&= |\alpha|\sqrt{\langle v, v\rangle} \\
&= |\alpha|\|v\|.
\end{aligned}$$

Now, consider

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

□

Theorem 8.17. *Let V be a normed vector space over $\mathbb{R}$, and write the norm as $\|\cdot\|$. Then there exist an inner product $\langle \cdot, \cdot \rangle$ such that $\|v\| = \sqrt{\langle v, v\rangle}$, $\forall v \in V$, iff the parallelogram law holds, that is*

$$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2), \forall u, v \in V.$$

Proof. Suppose first that $\|v\| = \sqrt{\langle v, v\rangle}$ for some inner product. Then

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

Conversely, suppose $\|\cdot\|$ satisfies the parallelogram law. If $\|v\| = \sqrt{\langle v, v\rangle}$, for some inner product, then

$$\begin{aligned}
\|u+v\|^2 - \|u-v\|^2 &= (\langle u, u\rangle + 2\langle u, v\rangle + \langle v, v\rangle) - (\langle u, u\rangle - 2\langle u, v\rangle + \langle v, v\rangle) \\
&= 4\langle u, v\rangle \\
&\Rightarrow \langle u, v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2}{4}.
\end{aligned}$$

Define $\langle \cdot, \cdot \rangle$ by

$$\langle u, v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2}{4}.$$

It remains to prove that $\langle \cdot, \cdot \rangle$ is an inner product. (Math Stack Exchange). □

9 Week 9

9.1 Lecture 15. Tue Oct 22

9.1.1 Orthonormal Bases

Definition 9.1. Let V be an inner product space. A list of vectors $v_1, \ldots, v_n \in V$ is called orthonormal if $\langle v_i, v_j \rangle = 0$, for $1 \leq i < j \leq n$, and $\langle v_i, v_i \rangle = 1$.

Proposition 9.2. *Suppose $v_1, \ldots, v_n \in V$ are orthonormal, for V an inner product space. Then*

$$\left\| \sum_{i=1}^{n} a_i v_i \right\|^2 = \sum_{i=1}^{n} |a_i|^2.$$

Proof. Consider that

$$\begin{aligned} \left\| \sum_{i=1}^{n} a_i v_i \right\|^2 &= \sum_{i=1}^{n} \|a_i v_i\|^2 \\ &= \sum_{i=1}^{n} |a_i|^2 \|v_i\|^2 \\ &= \sum_{i=1}^{n} |a_i|^2. \end{aligned}$$

□

Proposition 9.3. *Let $v_1, \ldots, v_n \in V$ be orthonormal vectors in an inner product space V. Then $v_1, \ldots, v_n$ are linearly independent.*

Proof. Let

$$\begin{aligned} 0 &= \sum_{i=1}^{n} a_i v_i \\ \|0\|^2 &= \left\| \sum_{i=1}^{n} a_i v_i \right\|^2 \\ 0 &= \sum_{i=1}^{n} |a_i|^2 \end{aligned}$$

$\Rightarrow$ all of the a_i's are zero.

□

Theorem 9.4 (Bessel's inequality)**.** *Let $v_1, \ldots, v_n$ be orthonormal in V, an inner product space. Then*

$$\sum_{k=1}^{n} |\langle v, v_k \rangle|^2 \leq \|v\|^2, \forall v \in V.$$

Proof. Consider $v \in V$. Then

$$\begin{aligned}
v &= \underbrace{\sum_{k=1}^{n} \langle v, v_k \rangle v_k}_{:=u} + \underbrace{v - \sum_{k=1}^{n} \langle v, v_k \rangle v_k}_{:=w} \\
\Rightarrow \langle w, v_\ell \rangle &= \langle v, v_\ell \rangle - \langle u, v_\ell \rangle \\
0 &= \langle v, v_\ell \rangle - \langle \sum_{k=1}^{n} \langle v, v_k \rangle v_k, v_\ell \rangle \\
&= \langle v, v_\ell \rangle - \sum_{k=1}^{n} \langle v, v_k \rangle \langle v_k, v_\ell \rangle.
\end{aligned}$$

So $w \perp w_\ell, \forall \ell \in [n]$, and $w \perp u$. Hence

$$\begin{aligned}
\|v\|^2 = \|u + w\|^2 &= \|u\|^2 + \underbrace{\|w\|^2}_{\geq 0} \\
&\geq \|u\|^2.
\end{aligned}$$

□

Theorem 9.5. *For an inner product space V, and an orthonormal basis $B := \{v_1, \ldots, v_n\}$ of V, $\varphi_j = \langle \cdot, v_j \rangle$, for $j = 1, \ldots, n$ is a dual basis of B.*

Theorem 9.6. *Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis of an inner product space V. Let $u, v \in V$. Then*

1. $v = \sum_{i=1}^{n} \langle v, v_k \rangle v_k$,

2. $\|v\|^2 = \sum_{i=1}^{n} |\langle v, v_k \rangle|^2$,

3. $\langle u, v \rangle = \sum_{i=1}^{n} \langle u, v_k \rangle \langle v_k, v \rangle$.

9.1.2 Gram-Schmidt Procedure

Theorem 9.7. *Suppose $v_1, \ldots, v_m$ is a linearly independent list of vectors in V. Let $f_1 = v_k$. For $k = 2, \ldots, m$, define f_k inductively by*

$$f_k = v_k - \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2} f_1 - \ldots - \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2} f_{k-1}. \tag{9.7.1}$$

For each $k = 1, \ldots, m$, let $e_k = \frac{f_k}{\|f_k\|}$. Then $e_1, \ldots, e_m$ is an orthonormal list of vectors such that

$$\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(e_1, \ldots, e_k), \tag{9.7.2}$$

for each $k = 1, \ldots, m$.

Proof. Proceed by induction on k. The conclusion is obviously true for $k = 1$. Suppose for $1 \leq k \leq m$, that the conclusion holds for $k-1$. Since $v_1, \dots, v_m$ are linearly independent, we know $v_k \notin \operatorname{span}(v_1, \dots, v_{k-1}) = \operatorname{span}(e_1, \dots, e_{k-1}) = \operatorname{span}(f_1, \dots, f_{k-1})$. (Note that this also implies that $f_k \neq 0$ and that $\|f_k\| \neq 0$, so we aren't dividing by zero in Eq. (9.7.1).) It is clear that $\|f_k\| = 1$. Let $j \in \{1, \dots, k-1\}$. Then

$$\begin{aligned}
\langle e_k, e_j \rangle &= \frac{1}{\|f_k\|\|f_j\|}\langle f_k, f_j \rangle \\
&= \frac{1}{\|f_k\|\|f_j\|}\left\langle v_k - \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2}f_1 - \dots - \frac{\langle v_k, f_j \rangle}{\|f_j\|^2}f_j - \dots - \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2}f_{k-1}, f_j \right\rangle \\
&= \frac{1}{\|f_k\|\|f_j\|}(\langle v_k, f_j \rangle - \langle v_k, f_j \rangle) \\
&= 0.
\end{aligned}$$

Thus $e_1, \dots, e_k$ is an orthonormal list. Using Eq. (9.7.1), we can deduce that $v_k \in \operatorname{span}(e_1, \dots, e_k)$. Combining this information with Eq. (9.7.2), we deduce

$$\operatorname{span}(v_1, \dots, v_k) \subseteq \operatorname{span}(e_1, \dots, e_k).$$

Both lists above are linearly independent, (the v's by hypothesis, and the e's by orthonormality). Thus, both subspaces above have dimension k, and hence, they are equal. ☐

Theorem 9.8. *Every finite dimensional inner product space has an orthonormal basis.*

Theorem 9.9. *Every list of orthonormal vectors can be extended into an orthonormal basis.*

Theorem 9.10. *Suppose V is a finite dimensional inner product space, and $T \in \mathcal{L}(V)$. Then T has an upper triangular matrix representation with respect to an orthonormal basis of V iff the minimal polynomial of T factorizes into*

$$p(z) = \prod_{j=1}^{m}(z - \lambda_j), \quad \text{for } \lambda_j \in \mathbb{F}.$$

Theorem 9.11 (Schur's Theorem)**.** *Every $T \in \mathcal{L}(V)$, for V an inner product space over $\mathbb{C}$ has an upper triangular matrix representation with respect to some orthonormal basis.*

Theorem 9.12 (Riesz Representation Theorem)**.** *Let V be a finite dimensional inner product space, and let $\varphi \in V'$. Then*

$$\exists! v \in V (u \in V \Rightarrow \varphi(u) = \langle u, v \rangle).$$

Proof. Let $\{v_1, \dots, v_n\}$ be an orthonormal basis of V. Then

$$\begin{aligned}
\varphi(u) &= \varphi\left(\sum_{k=1}^{n}\langle u, v_k \rangle v_k\right) \\
&= \sum_{k=1}^{n}\langle u, v_k \rangle \varphi(v_k) \\
&= \sum_{k=1}^{n}\langle u, \overline{\varphi(v_k)}v_k \rangle \\
&= \langle u, \sum_{k=1}^{n}\overline{\varphi(v_k)}v_k \rangle.
\end{aligned}$$

Then suppose $\varphi(u) = \langle u, v_1 \rangle = \langle u, v_2 \rangle, \forall u \in U$, Then

$$\begin{aligned} 0 = \varphi(u) - \varphi(u) &= \langle u, v_1 \rangle - \langle u, v_2 \rangle \\ &= \langle u, v_1 - v_2 \rangle. \end{aligned}$$

If we choose $u = v_1 = v_2$, then

$$\langle v_1 - v_2, v_1 - v_2 \rangle = 0 \iff v_1 - v_2 = 0 \Rightarrow v_1 = v_2.$$

□

9.2 Lecture 16. Thu Oct 24

Exercise 9.1. Suppose $B = \{e_1, \ldots, e_n\} \in V$, for V an inner product space. Suppose that

$$\|\sum_{k=1}^{n} a_k e_k\|^2 = \sum_{k=1}^{n} |a_k|^2,$$

for all $a_i \in \mathbb{F}$. Show that B is an orthonormal list.

Solution. By taking $a_i \neq 0$, and $a_j = 0$, for $j \neq i$,

$$\begin{aligned} \|\sum_{k=1}^{n} a_k e_k\|^2 = \|a_i e_i\|^2 &= |a_i|^2 \|e_i\|^2 \\ &= |a_i|^2 \\ \Rightarrow \|e_i\| &= 1. \end{aligned}$$

For arbitrary $i \neq j$, let $a_i, a_j = 1$, and $a_\ell = 0$ otherwise. Then

$$\begin{aligned} \|e_i + e_j\|^2 &= \langle e_i + e_j, e_i + e_j \rangle \\ &= \|e_i\|^2 + \|e_j\|^2 + \langle e_i, e_j \rangle + \langle e_j, e_i \rangle \\ &= 2 + 2 \operatorname{Re} \langle e_i, e_j \rangle \\ &= 2 \\ \Rightarrow \operatorname{Re} \langle e_i, e_j \rangle &= 0. \end{aligned}$$

For a similar choice of a_i, a_j, $\operatorname{Im} e_i, e_j = 0$. This implies that $e_i \perp e_j, \forall i \neq j$. □

9.2.1 Orthogonal Complements and Minimization Problems

Definition 9.13. For $U \subseteq V$, and V an inner product space, the orthogonal complement

$$U^\perp = \{v \in V \mid \langle u, v \rangle = 0, \forall u \in U\}.$$

Example 9.14. Consider $H^2(\mathbb{D})$, the set of analytic functions on the disk. Suppose $\{1, z, z^2, \ldots\}$ is an orthonormal basis of $H^2(\mathbb{D})$. So

$$(\operatorname{span}(z^k))^\perp = \operatorname{span}\{z^\ell \mid \ell \in N_z \setminus \{k\}\}$$

Proposition 9.15. *The orthogonal complement of* $\{0\}$ *is* V*. Also,* $V^{\perp} = \{0\}$*.*

Theorem 9.16. *Consider* $U, W \subset V$*.*

1. *If* $U \subseteq V$*, then* $W^{\perp} \subseteq U^{\perp}$*.*
2. *Also,* $U^{\perp} \leq V$*.*
3. *Finally,* $U \cap U^{\perp} = \{0\}$*.*

Proof. Let $U \subseteq W$, and let $w \in W^{\perp}$. Then

$$\begin{aligned} &\langle v, w\rangle = 0, \forall v \in W \\ \Rightarrow &\langle v, w\rangle = 0, \forall v \in U \\ &\Rightarrow w \in U^{\perp}. \end{aligned}$$

We omit the proof for 2, since it is standard. Let $u \in U \cap U^{\perp}$. Then

$$\langle u, u\rangle = 0 \Rightarrow u = 0.$$

□

Theorem 9.17. *If* $U \leq V$*, and* $\dim U < \infty$*, then* $V = U \oplus U^{\perp}$*.*

Proof. We already know $U \cap U^{\perp} = \{0\}$. It remains to show that $V = U + U^{\perp}$. Let $v \in V$, and let $u_1, \ldots, u_n$ be an orthonormal basis of U. Then

$$v = \underbrace{\sum_{k=1}^{n} \langle v, u_k\rangle u_k}_{:=u\in U} + \underbrace{v - \sum_{k=1}^{n} \langle v, u_k\rangle u_k}_{:=u^{\perp}}$$

$$\begin{aligned} &\Rightarrow \forall k, \langle u^{\perp}, u_k\rangle = \langle v, u_k\rangle - \langle v, u_k\rangle = 0 \\ &\quad\Rightarrow u^{\perp} \perp U, u^{\perp} \in U^{\perp}. \end{aligned}$$

□

Theorem 9.18. *Suppose* $U \leq V$*,* $\dim U < \infty$*. Then* $U = (U^{\perp})^{\perp}$*.*

Proof. For any $u \in U$, we have $\langle u, v\rangle = 0, \forall v \in U^{\perp}$, hence $u \in (U^{\perp})^{\perp}$. On the other hand, let $U \in (U^{\perp})^{\perp}$. Then $u = v + w$, for $v \in U, w \in U^{\perp}$. Therefore, $u - v = w \in U^{\perp}$. However, $u - v \in (U^{\perp})^{\perp}$, by closure. Hence, $u - v = 0$. However, $0 \in U$, so that $u \in U$ (as $v \in U$). □

Corollary 9.18.1. *If also* $\dim V < \infty$*, then* $\dim U^{\perp} = \dim V - \dim U$*.*

Corollary 9.18.2. *If* $\dim U < \infty$*,* $U \leq V$*, then* $U^{\perp} = \{0\} \Leftrightarrow U = V$*.*

Theorem 9.19. *Let* $\dim U < \infty$*, and* $U \leq V$*. Then suppose* $P_u \in \mathcal{L}(V)$*, with* $P_u^2 = P_u$*, and* $\|P_u v\| \leq \|V\|$*. Furthermore, suppose* $P_u u = u, \forall u \in U$*, and* $P_u w = 0, \forall w \in U^{\perp}$*. Then*

1. *the range of* $P_u = U$*, and* $\ker P_u = U^{\perp}$*.*
2. *For any* $v \in V$*,* $v - P_u v \in U^{\perp}$*.*
3. *If* $\{u_1, \ldots, u_n\}$ *is an orthonormal basis of* U*, then* $\forall v \in V$*,* $P_u v = \sum_{k=1}^{n} \langle v, u_k\rangle u_k$*.*

Theorem 9.20. *Let* $\dim U < \infty$*, and* $U \leq V$*. Let* $v \in V, u \in U$*. Then*

$$\|v - P_u v\| \leq \|v - u\|, \forall u \in U.$$

(This is the minimization problem.)

10 Week 10

10.1 Lecture 17. Tue Oct 29

10.1.1 Riesz Representation Theorem

Note. Recall the Riesz Representation Theorem, Theorem 9.12. That is, if V is an inner product space of finite dimension, and $\varphi \in V'$, then there exists a unique $v \in V$ such that $\varphi(u) = \langle u, v \rangle$, for all $u \in V$. Also, recall that $\dim V = \dim \operatorname{range} \varphi + \dim \ker \varphi$. Here, $\dim \operatorname{range} \varphi = 1$, if $\varphi \neq 0$.

Proposition 10.1. *If $\varphi \in V'$, and $v \in V$, with $\varphi(u) = \langle u, v \rangle$, for all $u \in V$, then $v \in (\ker \varphi)^{\perp}$.*

Proof. If $\varphi \neq 0$, then $\dim (\ker \varphi)^{\perp} = 1$, so $\operatorname{span}\{v\} = (\ker \varphi)^{\perp}$. □

Theorem 10.2. *If V is a finite dimensional inner product space, for $v \in V$, define $v \mapsto \varphi_v$ by $\varphi_v(u) = \langle u, v \rangle$, for all $u \in V$. Then $v \mapsto \varphi_v : V \to V'$ is a bijection.*

Note. This mapping is linear iff $\mathbb{F} = \mathbb{R}$. If $\mathbb{F} = \mathbb{C}$, then this mapping is called antilinear.

Proof. We claim that the injectivity of this map is in analogy with Rieszz representation theorem. For the surjectivity, let $\varphi \in V'$. We want to show that there exists $v \in V$ such that $\varphi = \varphi_v$. If $\varphi = 0$, then $v = 0 \Rightarrow 0 = \varphi(0) = \langle u, 0 \rangle = 0, \forall u \in V$. Otherwise $\ker \varphi \neq V$. Let $0 \neq w \in (\ker \varphi)^{\perp}$. Define

$$0 \neq v := \frac{\overline{\varphi(w)}}{\|w\|^2} w.$$

Consider

$$\varphi(v) = \frac{|\varphi(w)|^2}{\|w\|^2} = \|v\|^2 \neq 0.$$

For $u \in V$, consider

$$\begin{aligned} u &= \overbrace{u - \frac{\varphi(u)}{\varphi(v)} v}^{\in \ker \varphi} + \overbrace{\frac{\varphi(u)}{\varphi(v)} v}^{\in (\ker \varphi)^{\perp}} \\ \varphi(u) &= 0 + \frac{\varphi(u)}{\varphi(v)} \varphi(v). \end{aligned}$$

Consider

$$\begin{aligned} \varphi_v(u) &= \langle u, v \rangle \\ &= \langle \frac{\varphi(u)}{\varphi(v)} v, v \rangle \\ &= \frac{\varphi(u)}{\|v\|^2} \langle v, v \rangle \\ &= \varphi(u). \end{aligned}$$

□

10.1.2 Projections, Pseudoinverses

Theorem 10.3. *Suppose that* $\dim U < \infty, U \leq V, V$ *possibly infinite dimensional. Let* $v \in V, u \in U$. *Then*

$$\|v - P_u(v)\| \leq \|v - u\|, \forall u \in U.$$

Sometimes, invertibility is broken by a finite dimensional subspace. Consider $T : \mathbb{F}^m \to \mathbb{F}^n$, for T injective, and $m < n < \infty$.

Theorem 10.4. *Let* $\dim V < \infty$. *Suppose that* $T \in \mathcal{L}(V, W)$. *Then* $T|_{(\ker T)^\perp}$ *is bijective onto* $\operatorname{range} T$.

Theorem 10.5 (Moore-Penrose Pseudoinverse)**.** *Let* $\dim V < \infty$, $T \in \mathcal{L}(V, W)$. *The pseudo inverse,* $T^\dagger \in \mathcal{L}(V, W)$ *of* T *is defined by*

$$T^\dagger = (T|_{(\ker T)^\perp})^{-1} P_{\operatorname{range}(T)} w, \forall w \in W.$$

Theorem 10.6. *Suppose* V *is finite dimensional, and* $T \in \mathcal{L}(V, W)$.

1. *If* T *is invertible, then* $T^\dagger = T^{-1}$.
2. $TT^\dagger = P_{\operatorname{range} T}$ *(the orthogonal projection of* W *onto* $\operatorname{range} T$*).*
3. $T^\dagger T = P_{(\ker T)^\perp}$ *(the orthogonal projection of* V *onto* $(\ker T)^\perp$*).*

Proof. Let $\tilde{w} \in (\operatorname{range} T)^\perp$. Then $P_{\mathrm{rangeT}} \tilde{w} = 0$, and $TT^\dagger \tilde{w} = 0$. Let $w \in \operatorname{range} T$. Then

$$\begin{aligned} TT^\dagger w &= T(T|_{(\ker T)^\perp}^{-1}) P_{\operatorname{range} T}(w) \\ &= w. \end{aligned}$$

□

Using the pseudoinverse. We can compute the best fit to a linear system $Ax = b$ using the pseudoinverse. If A is invertible, then $x = A^{-1}b$. If not, $x = A^\dagger b$ makes Ax as close to b is possible.

10.1.3 Self-Adjoint and Normal Operators

Note. Now, throughout this section, assume V is a finite dimensional inner product space.

Why study self-adjoint operators? We can place self adjoint operators and operators into analogy with the reals in the complex field.

Definition 10.7. The adjoint of $T \in \mathcal{L}(V, W)$ is defined to be

$$T^* : W \to V, \text{ such that } \langle Tv, w \rangle = \langle v, T^* w \rangle, \forall v \in V, \forall w \in W.$$

This is well defined, since

$$v \mapsto \langle Tv, w \rangle$$

is a linear functional which only depends on T, w. Invoke Riesz representation theorem to obtain a unique vector $T^* w$ in V.

Example 10.8. For fixed $u \in V$, $w \in W$, then define $T : V \to W$ by $v \mapsto \langle v, u\rangle_V w$. Then for all $v \in V, x \in W$,

$$\begin{aligned}\langle Tv, x\rangle_W &= \langle \langle v, u\rangle_V w, x\rangle_W \\ &= \langle v, u\rangle_V \langle w, x\rangle_W \\ &= \langle v, \underbrace{\langle x, w\rangle_W u}_{T^*x}\rangle_V\end{aligned}$$

Example 10.9. Let $T : V \to V$ with $V = \{f \in C^\infty[-1,1] \mid f(1) = f(-1)\}$, and $\langle f, g\rangle = \int_{-1}^1 f\overline{g}$. Let $Tf = if'$. Consider

$$\begin{aligned}\langle Tf, g\rangle &= \int_{-1}^1 if'(x)\overline{g(x)}\,dx \\ &= \left[if(x)\overline{g(x)}\right]_{-1}^1 + \int_{-1}^1 -if(x)\overline{g'(x)}\,dx \\ &= \langle f, ig'\rangle + i\left[f(1)\overline{g(1)} - f(-1)\overline{g(-1)}\right] \\ &= \langle f, ig'\rangle \\ \implies T^*g = ig'.\end{aligned}$$

Definition 10.10. If $T = T^*$, then T is self-adjoint.

Example 10.11. On $\mathcal{P}_2(\mathbb{R})$, consider $\langle p, q\rangle = \int_0^1 pq$. Define $T \in \mathcal{L}(\mathcal{P}_2(\mathbb{R}))$ by

$$T(ax^2 + bc + c) = bx.$$

Find T^*.

Theorem 10.12. *If $T \in \mathcal{L}(V, W)$, then $T^* \in \mathcal{L}(V, W)$.*

Proof. Consider that for all $v \in V, w_1, w_2 \in W$, and $c \in \mathbb{C}$,

$$\begin{aligned}\langle v, T^*(cw_1 + w_2)\rangle &= \langle Tv, cw_1 + w_2\rangle \\ &= \langle Tv, cw_1\rangle + \langle Tv, w_2\rangle \\ &= \langle \overline{c}v, T^*w_1\rangle + \langle v, T^*w_2\rangle \\ &= \langle v, cT^*w_1 + T^*w_2\rangle.\end{aligned}$$

□

Theorem 10.13. *If $T, S \in \mathcal{L}(V, W)$, and $\lambda \in \mathbb{F}$, then*

1. $(S + T)^* = S^* + T^*$,
2. $(\lambda T)^* = \overline{\lambda}T^*$,
3. $(T^*)^* = T$,
4. $(ST)^* = T^*S^*$,
5. $I^* = I$,
6. *T is invertible $\iff$ T^* is invertible, and in this case,* $(T^*)^{-1} = (T^{-1})^*$.

10.2 Lecture 18. Thu Oct 31

Theorem 10.14. *Let $T \in \mathcal{L}$. Then*

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

Proof. Note that $w \in \ker T^*$ if and only if

$$\begin{aligned} &T^* w = 0 \\ \iff &0 = \langle T^* w, v\rangle, \forall v \in V \\ &= \langle w, Tv\rangle \\ \iff &w \in (\operatorname{range} T)^{\perp}. \end{aligned}$$

Similarly, 2 is proven, and $1 \Leftrightarrow 4$, $2 \Leftrightarrow 3$. □

Definition 10.15. For $A \in \mathbb{F}^{m\times n}$, the conjugate transpose is

$$(A^*)_{ij} = \overline{A_{ji}}.$$

Theorem 10.16. *Let $T \in \mathcal{L}(V, W)$. Let $E = \{e_1, \dots, e_n\}$ be an orthonormal basis of V. Then there exists $F = \{f_1, \dots, f_n\}$, such that F is an orthonormal basis of W. Then $M_{F,E}(T^*) = (M_{E,F}(T))^*$.*

Proof. Consider that $M_{E,F}(T)$ contains the coefficients

$$\begin{aligned} Te_j &= \text{ some linear combination of } f_k\text{'s} \\ &= \sum_{k=1}^{n} \langle Te_j, f_k\rangle f_k \\ &= \langle Te_j, f_k\rangle. \end{aligned}$$

Consider that $(M_{F,E}(T^*))_{jk} = \overline{\langle Te_j, f_k\rangle}$. Then

$$T^* f_k = \sum_{j=1}^{n} \langle T^* f_k, e_j\rangle e_j = \sum_{j=1}^{n} \overline{\langle Te_j, f_k\rangle} e_j.$$

□

Theorem 10.17. *Let $T : V \to V$, $M_B(T)$. Then T is self-adjoint iff $M_B(T) = M_B(T)^*$.*

Theorem 10.18. *The eigenvalues of a self-adjoint operator T are real.*

Proof. Let $Tv = \lambda b, v \neq 0, \lambda \in \mathbb{F}$. WLOG, assume $\|v\| = 1$. Then

$$\lambda = \lambda\langle v, v\rangle = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, T^* v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle = \overline{\lambda}.$$

□

Theorem 10.19. *For $\mathbb{F} = \mathbb{C}$, $T \in \mathcal{L}(V)$,*

$$\langle Tv, v\rangle = 0, \forall v \in V \Leftrightarrow T = 0. \langle Tv, v\rangle \in \mathbb{R}, \forall v \in V \Leftrightarrow T \text{ is self-adjoint.}$$

Proof. The converse implication is trivial, for the forward, use the polarization identity:

$$\langle Tu, v\rangle = \frac{(\langle T(u+v), u+v\rangle - \langle T(u-v), u-v\rangle + i\langle T(u+viv), u+iv\rangle - i\langle T(u-iv), u-iv\rangle}{4}.$$

□

Note. I got a little lazy here. At this point, each of these proof which I omit are ripped from LADR, 7A. I've found Axler a better resource for this class regardless.

Theorem 10.20. *Let $T \in \mathcal{L}(V)$. Then T is self adjoint if and only if $T - T^* = 0$.*

Theorem 10.21. *Let T be self adjoint. Then*

$$\langle Tv, v\rangle = 0, \forall v \in V \iff T = 0.$$

Theorem 10.22 (7.14)**.** *Suppose V is a complex inner product space, and $T \in \mathcal{L}(V)$. Then T is self-adjoint iff $\langle Tv, v\rangle \in \mathbb{R}$, for every $v \in V$.*

Definition 10.23 (Normal operators)**.** An operator on an inner product space is called normal if it commutes with its adjoint, that is $T \in \mathcal{L}(V)$ is normal iff $TT^* = T^*T$.

Theorem 10.24. *Suppose $T \in \mathcal{L}(V)$. Then T is normal iff $\|Tv\| = \|T^*v\|$, for all $v \in V$.*

Theorem 10.25. *Suppose $T \in \mathcal{L}(V)$ is normal. Then*

1. $\text{null}(T) = \text{null}(T^*)$,
2. $\text{range}(T) = \text{range}(T^*)$,
3. $V = \text{null}(T) \oplus \text{range}(T)$,
4. $T - \lambda I$ *is normal for every* $\lambda \in \mathbb{F}$,
5. *if* $v \in V$, *and* $\lambda \in \mathbb{F}$, *then* $Tv = \lambda v$ *if and only if* $T^*v = \overline{\lambda}v$.

Theorem 10.26. *Suppose $T \in \mathcal{L}(V)$ is normal. Then eigenvectors of T corresponding to distinct eigenvalues of T are orthogonal.*

11 Week 11

11.1 Lecture 19. Thu Nov 7

11.1.1 Complex Spectral Theorem

Recall that $T \in \mathcal{L}(V)$ is normal iff $T^*T = TT^*$. Also, given a basis $\{v_1, \ldots, v_n\}$, the matrix $\mathcal{M}(T)$ is the unique matrix $A \in \mathbb{F}^{n\times n}$ such that $Tv_j = \sum_{i=1}^n A_{ij}v_i$, for $j = 1, \ldots, n$. Note that if $v \in V$, then

$$\begin{aligned} v = \sum_{j=1}^n \alpha_j v_j \Longrightarrow Tv &= T\Bigg(\sum_{j=1}^n \alpha_j v_j\Bigg) \\ &= \sum_{j=1}^n \alpha_j T v_j \\ &= \sum_{j=1}^n \left(\sum_{i=1}^n A_{ij} v_i\right) \\ &= \sum_{j=1}^n \sum_{i=1}^n (A_{ij}\alpha_j) v_i \\ &= \sum_{i=1}^n \left(\sum_{j=1}^n A_{ij}\alpha_j\right) v_i \\ &= \sum_{i=1}^n \beta_i v_i, \quad \beta = A\alpha. \end{aligned}$$

So,

$$v = \sum_{j=1}^n \alpha_j v_j, \quad Tv = \sum_{j=1}^n \beta_i v_i \Longleftrightarrow \beta = A\alpha.$$

Recall that for $S, T \in \mathcal{L}(V)$,

$$\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T),$$

and $\mathcal{M} : \mathcal{L}(V) \to \mathbb{F}^{n\times n}$ is an isomorphism. Also, if $\{v_1, \ldots, v_n\}$ is orthonormal, then $\mathcal{M}(T^*) = \mathcal{M}(T)^*$. (Where $A^* = \overline{A}^t$.) The best bases are orthonormal:

$$v \in V \Longrightarrow v = \sum_{j=1}^n \alpha_i v_i,$$

where $\alpha_j = \langle v, v_j\rangle$, for all j. The best matrices are diagonal. For example, if $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, $x \in \mathbb{F}^n$, then $(Ax)_i = \lambda_i x_i$, for all i. Thus, given $T \in \mathcal{L}(V)$, the best of all possible worlds is that there exists an orthonormal basis $\{v_1, \ldots, v_n\}$ such that $\mathcal{M}(T, \{v_1, \ldots, v_n\})$ is diagonal.

Theorem 11.1 (Spectral Theorem, Complex)**.** *Let V be a finite dimensional complex vector space, and let $T \in \mathcal{L}(V)$. Then the following are equivalent:*

1. *T is normal.*
2. *There exists an orthonormal basis for V consisting of eigenvectors for T.*

3. *There exists an orthonormal basis for V such that the matrix of T with respect to that matrix is diagonal.*

Proof. Let us show that the second condition holds if and only if the third holds. Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis for V. Suppose there exists $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ such that $Tv_j = \lambda_j v_j$, for all $j = 1, \ldots, n$. But then

$$Tv_j = \sum_{i=1}^{n} A_{ij} v_i, i = 1, \ldots, n,$$

$$A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

That is, $A_{ij} = \lambda_i$ if $i = j$, and 0 if $i \neq j$. Therefore, the third condition holds. Conversely, assume that

$$Tv_j = \sum_{i=1}^{n} A_{ij} v_i, j = 1, \ldots, n,$$

and A is diagonal. But then,

$$Tv_j = A_{jj} v_j, j = 1, \ldots, n,$$

i.e., each v_j is an eigenvector for T. (The v_j's form a basis, so they are nonzero.) Thus, the second condition holds. Now, let us show that the third and the first conditions are equivalent. Suppose now $\{v_1, \ldots, v_n\}$ is an orthonormal for V, and $A = \mathcal{M}(T)$ is diagonal. Then

$$\begin{aligned}
\mathcal{M}(T^*T) &= \mathcal{M}(T^*)\mathcal{M}(T) \\
&= \mathcal{M}(T)^*\mathcal{M}(T) \\
&= \mathcal{M}(T)\mathcal{M}(T)^* && \text{(diagonal matrices commute)} \\
&= \mathcal{M}(T)\mathcal{M}(T^*) = \mathcal{M}(TT^*).
\end{aligned}$$

But this implies that $T^*T = TT^*$. (Since $M : \mathcal{L}(V) \to \mathbb{C}^{n \times n}$ is an isomorphism.) Thus, T is normal. Conversely, suppose T is normal. Then Schur's theorem applies, that is, there exists an orthonormal basis $\{v_1, \ldots, v_n\}$ such that $A = \mathcal{M}(T, \{v_1, \ldots, v_n\}$ is upper triangular. But now

$$\begin{gathered}
T^*T = TT^* \\
\implies \mathcal{M}(T)^*\mathcal{M}(T) = \mathcal{M}(T)\mathcal{M}(T)^* \\
\implies A^*A = AA^*,
\end{gathered}$$

and since $A^*A = AA^*$, and A is upper triangular, A is diagonal. (To see this, note that

$$(A^*A)_{11} = \overline{A}_{11} A_{11} = |A_{11}|^2.$$

and

$$\begin{aligned}
(AA^*)_{11} &= A_{11}\overline{A}_{11} + A_{12}\overline{A}_{12} + \ldots + A_{1n}\overline{A}_{1n} \\
&= |A_{11}|^2 + |A_{12}|^2 + \ldots + |A_{1n}|^2.
\end{aligned}$$

So

$$(A^*A)_{11} = (AA^*)_{11} \implies |A_{12}|^2 + \ldots + |A_{1n}|^2 = 0 \implies A_{12} = \ldots = A_{1n} = 0.$$

Now assume by induction that the only nonzero entries in row $1, \ldots, k-1$ lie on the diagonal. Then we have

$$\begin{aligned} (A^*A)_{kk} &= |A_{1k}|^2 + \ldots + |A_{kk}|^2 && \text{(since } A \text{ is upper triangular)} \\ &= |A_{kk}|^2, && \text{(by induction hypothesis)} \end{aligned}$$

and

$$(AA^*)_{kk} = |A_{kk}|^2 + \ldots + |A_{kn}|^2.$$

But then

$$\begin{aligned} (A^*A)_{kk} = (AA^*)_{kk} &\Longrightarrow |A_{k,k+1}|^2 + \ldots + |A_{kn}|^2 = 0 \\ &\Longrightarrow A_{k,k+1} = \ldots = A_{kn} = 0. \end{aligned}$$

So by induction, A is diagonal.) □

Note. Where did we use the fact that V was a complex vector space? Schur's theorem requires V to be complex. Conisder that

$$\begin{gathered} \mathcal{T} = A, \quad A \text{is upper triangular} \\ Tv_1 = A_{11}v_1 \Longrightarrow v_1 \text{ is an eigenvector of } T. \end{gathered}$$

To prove Schur's theorem, we must know that T has an eigenvalue.

Proposition 11.2 (Revisited)**.** *Let V be a finite dimensional complex vector space, with $\dim V = n$. Let $T \in \mathcal{L}(V)$. Then T has an eigenvalue.*

Proof. Choose any nonzero $v \in V$, and consider

$$\{v, Tv, T^2v, \ldots, T^nv\}.$$

This is a set of $n+1$ vectors in an n-dimensional space. Hence the set is linearly dependent, so there exists $m \leq n$ by the linear dependence lemma such that

$$\{v, Tv, \ldots, T^{m-1}v\}$$

is linearly independent, and

$$\{v, Tv, \ldots, T^{m-1}v, T^mv\}$$

is linearly dependent. Thus, there exists $a_0, \ldots, a_{m-1}$ such that

$$a_0v + a_1Tv + \ldots + a_{m-1}T^{m-1} + T^mv = 0.$$

This is the same as saying

$$p(T)v = 0, p(x) = a_0 + a_1x + \ldots + a_{m-1}x^{m-1} + x^m.$$

The fact that $\{v, Tv, \ldots, T^{m-1}v\}$ is linearly independent means that for $q \in \mathcal{P}(\mathbb{C})$

$$\deg(q) < m \Longrightarrow q(T)v = 0.$$

Now, by the fundamental theorem of algebra, p has a root λ, so that

$$p(x) = (x - \lambda)q(x), \deg(q) = m - 1.$$

But then

$$p(T)v = 0 \Longrightarrow (T - \lambda I)q(T)v = 0,$$

and $q(T)v \neq 0$. Thus, $(T - \lambda I)$ is singular, so that λ is an eigenvalue of T. □

11.1.2 Real Spectral Theorem

What if V is real? In the previous proof, we are able to get as far as declaring that $\deg(q) < m \Longrightarrow q(T)v \neq 0$. Although $\mathbb{R}$ is not algebraically closed, we know that $p(x)$ can be factored into a product of irreducible (over $\mathbb{R}$) quadratics and linear factors. We want to show that in this context, there must be at least one linear factor. We have

$$p(x) = a_0 + a_1 x + \ldots + a_{m-1} x^{m-1} + x^m,$$

where $a_0, \ldots, a_{m-1} \in \mathbb{R}$, and $p(T)v = 0$. (Recall that an irreducible polynomial over $\mathbb{R}$ has the form $x^2 + bc + c, b^2 - 4c < 0$.)

Lemma 11.3. *Let V be a finite dimensional real inner product space, and let T be a self-adjoint linear operator. Suppose that $b, c \in \mathbb{R}$, and $b^2 - 4c < 0$. Then*

$$T^2 + bT + cI \text{ is nonsingular.}$$

Proof. Let $v \in V$, $v \neq 0$. We must show that $(T^2 + bT + cI)v \neq 0$. We do this by showing

$$\langle (T^2 + bT + cI)v, v \rangle > 0.$$

We have

$$\begin{aligned}
\langle (T^2 + bT + cI)v, v \rangle &= \langle T^2 v, v \rangle + b\langle Tv, v \rangle + c\langle v, v \rangle \\
&= \langle Tv, Tv \rangle + b\langle Tv, v \rangle + c\|v\|^2 \\
&\geq \|Tv\|^2 - b\|Tv\|\|v\| + c\|v\|^2 \qquad \text{(CS)} \\
&= (\|Tv\| - \frac{b}{2}\|v\|)^2 + c\|v\|^2 - \frac{b^2}{4}\|v\|^2 \\
&= (\|Tv\| - \frac{b}{2}\|v\|)^2 + \frac{4c - b^2}{4}\|v\|^2 > 0.
\end{aligned}$$

The last inequality holds since $b^2 - 4c < 0 \Longrightarrow 4c - b^2 > 0$, and $\|v\| > 0$. □

Applying our lemma, since $p(T)v = 0, v \neq 0$, $p(T)$ is singular, so $p(x)$ cannot have only irreducible quadratic factors. (Since otherwise, it would be the composition of singular operators.) Hence, $p(x)$ has a linear factor, and thus, T has an eigenvalue, assuming that T is self-adjoint.

Theorem 11.4. *Let V be a finite dimensional inner product space, and let $T \in \mathcal{L}(V)$ be self-adjoint. Then T has an eigenvalue.*

Lemma 11.5. *Let V be a real inner product space, and let $T \in \mathcal{L}(V)$ be self-adjoint. Assume that U is a subspace of V that is invariant under T. Then*

1. *$U^\perp$ is invariant under T.*
2. *The restriction $T|_U \in \mathcal{L}(U)$ is self-adjoint.*
3. *The restriction $T_{U^\perp} \in \mathcal{L}(U^\perp)$ is self-adjoint.*

Proof. For the first statement, suppose that $v \in U^\perp$. We must show that $Tv \in U^\perp$, i.e.

$$\langle Tv, u \rangle = 0, \forall u \in U.$$

But

$$\langle Tv, u\rangle = \langle v, Tu\rangle = 0,$$

since $v \in U^\perp$ and $u \in U \implies Tu \in U$. Thus, $Tv \in U^\perp$, that is, $U^\perp$ is invariant under T. For the second statement, recall that $\langle\cdot,\cdot\rangle_V : V \times V \to \mathbb{R}$, and $U \subset V \implies \langle\cdot,\cdot\rangle_U = \langle\cdot,\cdot\rangle_V|_U$. Let $u, v \in U$. Then

$$\langle (T|_U)u, v\rangle_U = \langle Tu, v\rangle_V = \langle u, Tv\rangle_V = \langle u, (T|_U)v\rangle_U.$$

Thus, $T|_U \in \mathcal{L}(U)$ is self-adjoint. Similarly, the third condition holds. □

12 Week 12

12.1 Lecture 20. Tue Nov 12

Theorem 12.1 (Real spectral theorem). *Let V be a real finite-dimensional inner product space, and let $T \in \mathcal{L}(V)$. The following are equivalent:*

1. *T is self-adjoint.*
2. *There exists an orthonormal basis $\{v_1, \ldots v_n\}$ for V consisting of eigenvectors of T.*
3. *There exists an orthonormal basis $\{v_1, \ldots, v_n\}$ such that $M(T, \{v_1, \ldots, v_n\}$ is diagonal.*

Proof. Just as in the complex case, conditions 2 and 3 are equivalent:

$$Tv_j = \lambda_j v_j \quad \forall j \iff Tv_j = \sum_{i=1}^{n} A_{ij} v_i, \quad \forall j,$$

where $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Now we prove that conditions 1 and 3 are equivalent. Suppose the third condition holds. Then

$$\begin{aligned} M(T^*) &= M(T)^T = M(T) && \text{(a diagonal matric is symmetric)} \\ &\implies T^* = T. && \text{(since } M : \mathcal{L}(V) \to \mathbb{R}^{n\times n} \text{ is bijective)} \end{aligned}$$

Thus, T is self-adjoint. Suppose now that T is self-adjoint. We argue by induction on $n = \dim V$. For $n = 1$, the result follows simply because every 1-by-1 matrix is diagonal. Now, assume that if $\dim U < n$, and if $S \in \mathcal{L}(V)$, is self-adjoint, then there exists an orthonormal basis for U such that the matrix of S respect to that orthonormal basis is diagonal. Now let $\dim V = n$, and assume $T \in \mathcal{L}(V)$ is self-adjoint. We know that T has a real eigenvalue λ_1, and corresponding eigenvector v_1. WLOG, assume $\|v_1\| = 1$. Define $U = \operatorname{span}(v_1)^\perp$ and $S = T|_U$. Note that $\operatorname{span}(v_1)$ is invariant under T, and therefore U is also invariant under T. Thus $S \in \mathcal{L}(U)$. Also, from last lecture, S is self-adjoint. Therefore, by our induction hypothesis, there exists an orthonormal basis, $\{v_2, \ldots, v_n\}$ for U such that and numbers $\lambda_2, \ldots, \lambda_n \in \mathbb{R}$ such that

$$Sv_j = \lambda_j v_j, \quad , j = 2, \ldots, n.$$

Therefore, $M(S, \{v_2, \ldots, v_n\})$ is diagonal. Since $U = \operatorname{span}(v_1)^\perp$, $\{v_1, \ldots, v_n\}$ is an orthonormal basis for V. Hence,

$$Tv_j = \lambda_j v_j, \quad j = 1, \ldots, n,$$

hence, $M(T, \{v_1, \ldots, v_n\}) = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. □

13 Week 13

13.1 Lecture 21. Tue Nov 19

13.1.1 Outer Products

Definition 13.1. Let V, W be inner product spaces, and let $v \in V, w \in W$ be fixed. We define $w \otimes v \in \mathcal{L}(V, W)$ the outer product of v and w by

$$(w \otimes v)u = \langle u, v\rangle_V w, \quad \forall u \in V.$$

For future reference, note

$$\begin{aligned}
\langle (w \otimes v)u, z\rangle_W &= \langle \langle u, v\rangle_V w, z\rangle_W \\
&= \langle u, v\rangle_V \langle w, z\rangle_W \\
&= \langle u, \overline{\langle w, z\rangle_V} v\rangle_V \\
&= \langle u, \langle z, w\rangle_W v\rangle_V \\
&= \langle u, (v \otimes w)z\rangle_V .
\end{aligned}$$

In other words,

$$(w \otimes v)^* = v \otimes w.$$

Also, note that

$$\begin{aligned}
&\mathcal{R}(w \otimes v) = \operatorname{span}(w) \\
\implies &\dim \mathcal{R}(w \otimes v) = 1 \qquad \text{(or 0 if } w = 0) \\
\implies &\operatorname{rank}(w \otimes v) = 1.
\end{aligned}$$

So we often call $w \otimes v$ a rank-one map. Let $v_1, \ldots, v_n$ be an orthonormal basis of an inner product space V. Now, consider

$$\begin{aligned}
Tv &= \sum_{j=1}^{n} \lambda_j \langle v, v_j\rangle v_j \\
&= \sum_{j=1}^{n} \lambda_j (v_j \otimes v_j) v \\
&= \left(\sum_{j=1}^{n} \lambda_j v_j \otimes v_j \right) v \\
\implies T &= \sum_{j=1}^{n} \lambda_j v_j \otimes v_j .
\end{aligned}$$

In other words, the spectral theorem allows us to write T as the sum of rank-one operators.

13.1.2 Positive Operators

Definition 13.2. Let V be an inner product space and let $T \in \mathcal{L}(V)$ be self-adjoint. We say that T is positive iff $\langle Tv, v\rangle \geq 0, \forall v \in V$. For a matrix $A \in \mathbb{R}^{n\times n}$, if $A^T = A$, and A is positive, we tend to call A positive semidefinite iff $(Ax) \cdot x \geq 0, \forall x \in \mathbb{R}^n$.

Theorem 13.3. *Let V be a finite-dimensional inner product space, and let $T \in \mathcal{L}(V)$ be self-adjoint. Then T is positive iff all of its eigenvalues are nonnegative.*

Proof. Suppose first that T is positive. If $Tv = \lambda v$, with $v \neq 0$, then $\langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle$, so $\langle Tv, v\rangle \geq 0 \implies \lambda \geq 0$, since $\langle v, v\rangle \geq 0$. Thus, all the eigenvalues of T are nonnegative. Conversely, suppose that all eigenvalues of T are nonnegative. Since T is self-adjoint, V has an orthonormal basis $v_1, \ldots, v_n$ such that

$$\begin{aligned}
T &= \sum_{j=1}^{n} \lambda_j v_j \otimes v_j, \quad \lambda_j \geq 0 \\
\implies Tv &= \sum_{j=1}^{n} \lambda_j \langle v, v_j\rangle v_j, \quad \forall v \in V \\
\implies \langle Tv, v\rangle &= \langle \sum_{j=1}^{n} \lambda_j \langle v, v_j\rangle v_j, \sum_{i=1}^{n} \langle v, v_i\rangle v_i\rangle \\
&= \sum_{j=1}^{n} \sum_{i=1}^{n} \lambda_j \langle v, v_j\rangle \overline{\langle v, v_i\rangle} \overbrace{\langle v_j, v_i\rangle}^{=\delta_{ij}} \\
&= \sum_{j=1}^{n} \lambda_j \langle v, v_j\rangle \overline{\langle v, v_j\rangle} \\
&= \sum_{j=1}^{n} \lambda_j |\langle v, v_j\rangle|^2 \geq 0, \forall v \in V.
\end{aligned}$$

Thus, T is positive. □

Definition 13.4. Let V be a vector space, and let $T \in \mathcal{L}(V)$. Then $S \in \mathcal{L}(V)$ is a square root of T iff $S^2 = T$.

Theorem 13.5. *Let V be a finite-dimensional inner product space, and let $T \in \mathcal{L}(V)$ be self-adjoint. Then T is positive iff T has a positive square root iff T has a self-adjoint square root.*

Proof. Suppose first that $S \in \mathcal{L}(V)$ is self-adjoint, and $S^2 = T$. Then

$$\begin{aligned}
\forall v \in V, \langle Tv, v\rangle &= \langle S^2 v, v\rangle \\
&= \langle Sv, Sv\rangle \\
&\geq 0.
\end{aligned}$$

Hence, T is positive. Conversely, suppose T is positive. Then V has an orthonormal basis $v_1, \ldots, v_n$ such that

$$T = \sum_{j=1}^{n} \lambda_j v_j \otimes v_j, \quad \lambda_j \geq 0, j = 1, \ldots, n.$$

Define $S \in \mathcal{L}(V)$ by (via the linear map lemma)

$$Sv_j = \sqrt{\lambda_j} v_j, \quad j = 1, \ldots, n.$$

But then

$$S^2 v_j = S(Sv_j) = S(\sqrt{\lambda_j} v_j) = \sqrt{\lambda_j} S v_j = \lambda_j v_j = T v_j,$$

for all $j = 1, \ldots n$. Hence, $S^2 = T$. Note that S is self-adjoint by the spectral theorem, and its eigenvalues are all nonnegative, so T has a positive square root. □

13.1.3 Singular Value Decomposition

Lemma 13.6. *Let V, W be inner product spaces, and let $T \in \mathcal{V}, \mathcal{W}$. Then T^*T is positive.*

Proof. The proof that T^*T is self-adjoint is easy. Consider then

$$\langle T^*Tv, v\rangle = \langle Tv, Tv\rangle \geq 0, \forall v \in V.$$

Thus, T^*T is positive. □

Theorem 13.7 (Singular value decoposition)**.** *Let V, W be finite-dimensional inner product spaces, and let $T \in \mathcal{L}(V, W)$. Then there exists orthonormal bases $\{v_1, \ldots, v_n\}$ for V and $\{u_1, \ldots, u_m\}$ for W and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k$, where k is the minimum of m, n, such that*

$$T = \sum_{j=1}^{k} \sigma_j u_j \otimes v_j.$$

Now, this theorem tell us that we can write $\mathcal{M}(T, \{v_1, \ldots, v_n\}, \{u_1, \ldots, u_m\}) = \operatorname{diag}(\sigma_1, \ldots, \sigma_k)$. So we can diagonalize any finite-dimensional linear map, but we may have to use different orthonormal bases for the domain and codomain. We call $T = \sum_{j=1}^{k} \sigma_j u_j \otimes v_j$ the singular value decomposition of T, where $\sigma_1, \ldots, \sigma_k$ are the singular values of T, and $v_1, \ldots, v_n$ are the right singular vectors of T, and $u_1, \ldots, u_m$ are the left singular vectors of T.

Proof. Assume that $\dim W = m \geq n = \dim V$, so $k = n$. We know that $T * T$ is positive, so there exists an orthonormal basis $v_1, \ldots, v_n$ for V, and nonnegative real numbers $\lambda_1, \ldots, \lambda_n$ such that

$$T^*Tv_j = \lambda_j v_j, \quad j = 1, \ldots, n.$$

□

We can re-order $v_1, \ldots, v_n$ if necessary such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$. Suppose that $\lambda_1 \geq \ldots \geq \lambda_r > \lambda_{r+1} = \ldots = \lambda_n = 0$, for $r \leq n$. Define $\sigma_j = \sqrt{\lambda_j}$, for $j = 1, \ldots, n$ and

$$u_j = \sigma_j^{-1} Tv_j, \quad j = 1, \ldots, r.$$

Then

$$\begin{aligned}
\langle u_j, u_i\rangle_W &= \langle \sigma_j^{-1} Tv_j, \sigma_i^{-1} Tv_i\rangle_W \\
&= \sigma_j^{-1}\sigma_i^{-1}\langle Tv_j, Tv_i\rangle_W \\
&= \sigma_j^{-1}\sigma_i^{-1}\langle T^*Tv_j, v_i\rangle_V \\
&= \sigma_j^{-1}\sigma_i^{-1}\langle \sigma_j^2 v_j, v_i\rangle_V \\
&= \frac{\sigma_j^2}{\sigma_j\sigma_i}\langle v_j, v_i\rangle_V \\
&= \frac{\sigma_j^2}{\sigma_j\sigma_i}\delta_{ij} \\
&= \delta_{ij}.
\end{aligned}$$

Then extend $\{u_1, \ldots, u_m\}$ to an orthonormal basis for W, and we have

$$Tv_j = \sigma_j u_j, \quad j = 1, \ldots, n.$$

If $n > m$, then the above proof applied to T^* yields orthonormal bases $\{u_1, \ldots, u_m\}$ for W and $\{v_1, \ldots, v_n\}$ for V such that

$$T^* = \sum_{j=1}^{n} \sigma_j v_j \otimes u_j.$$

But then

$$\begin{aligned} T = T^{**} &= \left(\sum_{j=1}^{m} \sigma_j v_j \otimes u_j \right)^* \\ &= \sum_{j=1}^{m} \sigma_j \left(v_j \otimes u_j \right)^* \\ &= \sum_{j=1}^{m} \sigma_j u_j \otimes v_j. \end{aligned}$$

Theorem 13.8. *Let V, W be finite-dimensional inner product spaces. Let $T \in \mathcal{L}(V, W)$, and let*

$$T = \sum_{j=1}^{n} \sigma_j u_j \otimes v_j \qquad (k = \min m, n)$$

be the SVD of T. Suppose that $\sigma_1 \geq \ldots \geq \sigma_j > \sigma_{j+1} = \ldots = \sigma_k = 0$. Then

1. *$r = \operatorname{rank} T = \dim \mathcal{R}(T)$,*
2. *$\{u_1, \ldots, u_r\}$ is an orthonormal basis for $\mathcal{R}(T)$,*
3. *$\{v_{r+1}, \ldots, v_n\}$ is an orthonormal basis for $\operatorname{null} T$,*
4. *$\{v_1, \ldots, v_r\}$ is an orthonormal basis for $\mathcal{R}(T^*)$,*
5. *$\{u_{r+1}, \ldots, r_m\}$ is an orthonormal basis for $\operatorname{null} T^*$.*

We call the four subspaces above the fundamental subspaces of T.

Sketch of proof. For all $v \in V$, we know

$$Tv = \sum_{j=1}^{r} \sigma_j \langle v, v_j \rangle_V u_j,$$

so that $Tv \in \operatorname{span}(u_1, \ldots, u_r)$. Also, for all $j = 1, \ldots, r$,

$$u_j = T(\sigma^{-1} v_j) \Longrightarrow u_j \in \mathcal{R}(T).$$

So

$$\mathcal{R}(T) \subseteq \operatorname{span}(u_1, \ldots, u_r), \quad \operatorname{span}(u_1, \ldots, u_r) \subseteq \mathcal{R}(T).$$

Then namely, $\operatorname{rank}(T) = r$. The rest follows from standard computation. □

13.2 Lecture 22. Thu Nov 21

13.2.1 Solving Least-Squares Problems Using SVD

Suppose V, W are finite-dimensional inner product spaces. Let $T \in \mathcal{L}(V, W)$, and let $w \in W$. If $Tv = w$ has no solution, i.e. suppose $w \notin \mathcal{R}(T)$. We may wish to solve the equation in least-squares sense:

$$\min_{v \in V} \|Tv - w\|_W^2.$$

This is a projection problem, since $\mathcal{R}(T)$ is a subspace of W. By the projection theorem, v solves the least-squares problem iff

$$\begin{aligned}
&\langle Tv - w, s\rangle_W = 0, \quad \forall s \in \mathcal{R}(T) \\
\iff &\langle Tv - w, Tu\rangle_W = 0, \quad \forall u \in V \\
\iff &\langle T^*(Tv - w), u\rangle_V = 0, \quad \forall u \in V \\
\iff &T^*(Tv - w) = 0 \\
\iff &T^*Tv = T^*w. && \text{(The normal equation.)}
\end{aligned}$$

Thus, $T^*Tv = T^*w$ has a solution (because the projection problem has a solution), and every solution of $T^*Tv = T^*w$. Note that $\operatorname{null}(T^*T) = \operatorname{null} T$, so if T is singular, there are infinitely many least-squares solutions. Suppose that T has singular values

$$\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_n = 0.$$

(Assume that $m \geq n$.) Let $v_1, \ldots, v_n$ be the right singular vectors and $u_1, \ldots, u_m$ the left singular vectors of T. Then

$$w = \sum_{j=1}^{n} \langle w, u_j\rangle_W u_j,$$

and

$$Tv = \left(\sum_{j=1}^{r} \sigma_j u_j \otimes v_j\right) v = \sum_{j=1}^{r} \sigma_j \langle v, v_j\rangle_V u_j.$$

Note that

$$v = \sum_{j=1}^{n} \langle v, v_j\rangle_V v_j = \sum_{j=1}^{n} \alpha_j v_j, \quad \alpha_j = \langle v, v_j\rangle_V.$$

So we can think of $\langle v, v_j\rangle_V$, $j = 1, \ldots, n$ as the unknowns. We have

$$\begin{aligned}
Tv - w &= \sum_{j=1}^{r} \sigma_j \langle v, v_j\rangle_V u_j - \sum_{j=1}^{m} \langle w, u_j\rangle_W u_j \\
&= \sum_{j=1}^{r} \Big(\sigma_j \langle v, v_j\rangle_V - \langle w, u_j\rangle_W\Big) u_j - \sum_{j=r+1}^{m} \langle w, u_j\rangle_W u_j \\
\implies \|Tv - w\|_W^2 &= \sum_{j=1}^{r} |\sigma_j \langle v, v_j\rangle_V - \langle w, u_j\rangle_W|^2 + \sum_{j=r+1}^{m} |\langle w, u_j\rangle_W|^2
\end{aligned}$$

Therefore, $\|Tv - w\|_W^2$ is minimized by taking

$$\begin{aligned} \sigma_j \langle v, v_j\rangle_V - \langle w, v_j\rangle_W &= 0, \quad j = 1, \dots, r \\ \iff \langle v, v_j\rangle_V &= \frac{\langle w, u_j\rangle_W}{\sigma_j}, \quad j = 1, \dots, r. \end{aligned}$$

(We have $\langle v, v_j\rangle_V, j = 1, 2, \dots, n$ free.) The set of all least-squares solutions is

$$\left\{ \sum_{j=1}^{r} \frac{\langle w, u_j\rangle_W}{\sigma_j} v_j + \sum_{j=r+1}^{n} \alpha_j v_j \quad \middle| \quad \alpha_{r+1}, \dots, \alpha_n \in \mathbb{F} \right\}$$

Note,

$$\left\| \sum_{j=1}^{r} \frac{\langle w, u_j\rangle_W}{\sigma_j} v_j + \sum_{j=r+1}^{n} \alpha_j v_j \right\|_V^2 = \sum_{j=1}^{r} \frac{\langle w, u_j\rangle_W}{\sigma_j^2} + \sum_{j=r+1}^{n} \alpha_j^2.$$

So

$$v = \sum_{j=1}^{r} \frac{\langle w, u_j\rangle_W}{\sigma_j} v_j$$

is a special solution: it is the minimum-norm least-squares solution. Note that

$$\sum_{j=1}^{r} \frac{\langle w, u_j\rangle_W}{\sigma_j} v_j \in \mathcal{R}(T^*) = \mathrm{null}(T)^\perp.$$

Thus,

$$\begin{aligned} T^\dagger w &= \sum_{j=1}^{r} \frac{\langle w, u_j\rangle_W}{\sigma_j} v_j \\ &= \left(\sum_{j=1}^{r} \frac{1}{\sigma_j} v_j \otimes u_j \right) w \\ &= T^\dagger = \sum_{j=1}^{r} \frac{1}{\sigma_j} v_j \otimes u_j. \end{aligned}$$

In the special case that $m = n = r$, $T^\dagger = T^{-1}$. In this case,

$$\begin{aligned} T^{-1} &= \sum_{j=1}^{n} \frac{1}{\sigma_j} (v_j \otimes u_j) \\ \implies T &= \sum_{j=1}^{n} \sigma_j (u_j \otimes v_j). \end{aligned}$$

13.2.2 Generalized Eigenvectors and Eigenspaces

Let V be an n-dimensional vector space over a field $\mathbb{F}$ (no more inner product). Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of T. Recall that

$$E(\lambda_j, T) = \{v \in V \mid Tv = \lambda_j v\} = \{v \in V \mid (T - \lambda_j I)v = 0\} = \mathrm{null}(T - \lambda_j I).$$

We know that T is diagonalizable iff

$$V = E(\lambda_1, T) \oplus \ldots \oplus E(\lambda_k, T).$$

(We know $E(\lambda_1, T) + \ldots + E(\lambda_k, T)$ is always a direct sum when $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues, but this is not always equal to V.) Since not every $T \in \mathcal{L}(V)$ is diagonalizable, we want to figure out how to choose a basis B for V such that $\mathcal{M}_B(T)$ is as closed to diagonal as possible. We can do this by achieving several goals.

Goal 1. Find $V_1, \ldots, V_k$ such that

$$V = V_1 \oplus \ldots \oplus V_k,$$

and each V_j is invariant under T. (Each eigenspace of T is invariant under T, and if we had V writable as the direct sum of its eigenspaces, we will have achieved this.) Suppose that $B_j = \{v_{j,1}, \ldots, v_{j,n_j}\}$ is a basis for V_j. Since V_j is invariant under T,

$$T_{v_{j,i}} = \sum_{\ell=1}^{n_j} B_{\ell i}^{(j)} v_{j,\ell} \Longrightarrow \mathcal{M}_{B_j}(T|_{v_j}) = B^{(j)}.$$

Write $B = B_1 \cup \ldots \cup V_k$¿ Then,

$$\begin{aligned} Tv_{j,i} &= (0v_{1,1} + \ldots + 0_{1,n_1}) + \ldots + (0v_{j-1,1} + \ldots + 0v_{j-1,n_{j-1}}) \\ &\quad + \sum_{\ell=1}^{n_j} V_{\ell i}^{(j)} v_{j,\ell} + (0v_{j+1,1} + \ldots + 0v_{j+1,n_{j+1}}) + \ldots + (0v_{k,1} + \ldots + 0v_{k,n_k}). \end{aligned}$$

Thus,

$$\mathcal{M}_B(T) = \begin{bmatrix} & & & & & \\ 0 & \ldots & B^{(j)} & 0 & \ldots & 0 \\ & & & & & \end{bmatrix}.$$

This is true for every V_j or (B_j), so

$$\mathcal{M}_B(T) = \begin{bmatrix} B^{(1)} & & & \\ & B^{(2)} & & \\ & & \ddots & \\ & & & B^{(k)} \end{bmatrix}.$$

It turns out that

$$V = V_1 \oplus \ldots \oplus V_k,$$

and each V_j is invariant under T if

$$V_j = G(\lambda_j, T) = \mathrm{null}((T - \lambda_j I)^n),$$

the *generalized eigenspace* corresponding to λ_j.

Goal 2. Choose a basis for each $G(\lambda_j, T)$ so that each block $B^{(j)}$ of $\mathcal{M}_B(T)$ is as nearly diagonal as possible.

Lemma 13.9. *Let V be an n-dimensional vector space over a field $\mathbb{F}$, and let $S \in \mathcal{L}(V)$. Then*

$$\text{null}(S^0) \subseteq \text{null}(S^1) \subseteq \text{null}(S^2) \subseteq \dots$$

and there exists $m \in \mathbb{Z}^+$, $m \leq n$ such that

$$\text{null}(S^j) \subsetneq \text{null}(S^{j+1}), \quad j = 0, \dots, m-1,$$

and

$$\text{null}(S^j) = \text{null}(S^m), \quad j \geq m.$$

Proof. Note that

$$v \in \text{null}(S^j) \implies S^j v = 0 \implies S^{j+1} v = S(S^j)v = S0 = 0 \implies v \in \text{null}(S^{j+1}).$$

Note that $\text{null}(S^j) \subsetneq \text{null}(S^{j+1})$ implies that $\dim \text{null}(S^{j+1}) \geq \dim \text{null}(S^j) + 1$. It follows that there exists $m, 0 \leq m \leq n$, such that

$$\text{null}(S^m) = \text{null}(S^{m+1}).$$

Otherwise, $\dim \text{null}(S^{n+1}) \geq n+1$, but $\text{null}(S^{n+1}) \subseteq V$. Let m be the smallest such integer. Then

$$\text{null}(S^j) \subsetneq \text{null}(S^{j+1}), \quad j = 0, 1, \dots, m-1.$$

Finally, we must show that

$$\begin{aligned} \text{null}(S^j) &= \text{null}(S^m), \quad \forall j > m \\ \iff \text{null}(S^{m+k}) &= \text{null}(S^m), \quad \forall k \geq 1. \end{aligned}$$

We argue by induction on k. The case that $k = 1$ holds by definition of m. Suppose that $k \geq 1$, and

$$\text{null}(S^{m+k}) = \text{null}(S^m).$$

Then we must prove that $\text{null}(S^{m+k+1}) \subseteq \text{null}(S^M)$. (We already know that $\text{null}(S^m) \subseteq \text{null}(S^{m+k})$.) But

$$\begin{aligned} v \in \text{null}(S^{m+k+1}) &\implies S^{m+k+1} v = 0 \\ &\implies S^{m+1}(S^k v) = 0 \\ &\implies S^k v \in \text{null}(S^{m+1}) \\ &\implies S^k v \in \text{null}(S^m) \\ &\implies S^m(S^k v) = 0 \\ &\implies S^{m+k} v = 0 \\ &\implies v \in \text{null}(S^{m+k}) = \text{null}(S^m). \end{aligned}$$

Thus,

$$\text{null}(S^{m+k}) = \text{null}(S^m), \quad \forall k \geq 1.$$

Note that for all $S \in \mathcal{L}(V)$, there is a smallest m such that $\text{null}(S^j) = \text{null}(S^m), \forall j \geq m$. But note that for this m,

$$\text{null}(S^n) = \text{null}(S^m).$$

So if we didn't want to take the trouble of identifying m, we could just refer to $\text{null}(S^n)$. □

Lemma 13.10. *Let V be an n-dimensional vector over a field $\mathbb{F}$, and let $S \in \mathcal{L}(V)$, then*

$$V = \operatorname{null}(S^n) \oplus \mathcal{R}(S^n).$$

Proof. We know that

$$\dim V = \dim \operatorname{null}(S^n) + \dim \mathcal{R}(R^n),$$

so it suffices to prove that

$$\operatorname{null}(S^n) \cap \mathcal{R}(S^n) = \{0\},$$

since

$$\dim V = \dim \operatorname{null}(S^n) + \dim \mathcal{R}(S^n) - \dim(\operatorname{null}(S^n) \cap \mathcal{R}(S^n)).$$

So suppose $x \in \operatorname{null}(S^n) \cap \mathcal{R}(S^n)$. Then, since $x \in \mathcal{R}(S^n)$, there exists $v \in V$ such that $x = S^n v$. But then

$$\begin{aligned} x \in \operatorname{null}(S^n) \Longrightarrow S^n x = 0 &\Longrightarrow S^n(S^n v) = 0 \\ &\Longrightarrow S^{2n} v = 0 \\ &\Longrightarrow S^n v = 0 && \text{(since } \operatorname{null}(S^{2n}) = \operatorname{null}(S^n)\text{)} \\ &\Longrightarrow x = S^n v = 0. \end{aligned}$$

□