MATH 245 Notes

Spencer Wilson

May 6 – July 25, 2019

Contents

Lectures

1 Rings and Fields

May 6

1.1 Definition. A *ring* is a set R equipped with the operations $+\colon R\times R\to R$ and $\cdot\colon R\times R\to R$ such that

(1) for all $a,b,c\in R$, $(a+b)+c=a+(b+c)$;

(2) for all $a,b,c\in R$, $(a\cdot b)\cdot c=a\cdot(b\cdot c)$;

(3) for all $a,b\in R$, $a+b=b+a$;

(4) there exists $0\in R$ such that $a+0=0+a=a$ for all $a\in R$;

(5) for all $a\in R$, there exists $b\in R$ such that $a+b=0$ (we denote this b by $-a$);

(6) For all $a,b,c\in R$, $a\cdot(b+c)=a\cdot b+a\cdot c$ and $(b+c)\cdot a=b\cdot a+c\cdot a$.

1.2 Notation. The above axioms allow the the following notation.

(1) $+(a,b)\equiv a+b$ and $\cdot(a,b)\equiv a\cdot b\equiv ab$.

(2) By associativity, $a+b+c$ and abc are well-defined.

(3) For $n\in\mathbb{N}$, $a^n\coloneqq\underbrace{aa\cdots a}_{n\text{ times}}$ and $na\coloneqq\underbrace{a+a+\cdots+a}_{n\text{ times}}$.

(4) $a+(-b)\equiv a-b$.

1.3 Example. The following are rings:

(a) $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$

(b) $\mathbb{Z}_n$ for $n\in\mathbb{N}$

(c) $R[x]$ and $M_n(R)$ for any ring R

(d) $R_1\oplus R_2\coloneqq\{(a,b):a\in R_1,b\in R_2\}$ for any rings R_1 and R_2.

1.4 Example. The following are *not* rings:

(a) the odd numbers, since there is no 0 element

(b) $\mathcal{C}(\mathbb{R})$ under pointwise addition and function composition, since distributivity breaks.

1.5 Definition. We also consider two special types of rings.

(1) We say a ring R is *commutative* if $ab=ba$ for all $a,b\in R$.

(2) We say a ring R is *unital* if there exists $1\in R$ such that $1a=a1=a$ for all $a\in R$. We call 1 the *unity*, or "one," of R.

1.6 Example. $2\mathbb{Z}$ is a non-unital, commutative ring. $\mathsf{M}_n(2\mathbb{Z})$ is a non-unital, non-commutative ring.

1.7 Convention. Only the trivial ring is allowed to have trivial multiplication, i.e., $ab=0$ for all $a,b\in R$. Furthermore, the trivial ring is *not* unital.

May 8

1.8 Definition. Let R be a commutative ring. We say $a\in R$ is a *zero divisor* if $a\neq 0$ and there exists $b\neq 0$ such that $ab=0$.

1.9 Example. In $\mathbb{Z}_6$, 2 and 3 are zero divisors, since $2\cdot 3=0$.

1.10 Remark. For any $a\in\mathbb{Z}_n$, a is a zero divisor if and only if $\gcd(a,n)\neq 1$ and $a\neq 0$.

1.11 Definition. A ring R is an *integral domain* if R is commutative and unital and has no zero divisors.

1.12 Example. The following are integral domains:

(a) $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$

(b) $\mathbb{Z}_p$, where p is prime

(c) $R[x]$ for any integral domain R.

1.13 Example. The following are *not* integral domains:

(a) $\mathbb{Z}_n$, where $n > 1$ is not prime (zero divisors)

(b) $M_n(\mathbb{R})$, where $n > 1$ (not commutative)

(c) $2\mathbb{Z}$ (not unital)

(d) $\mathbb{R} \oplus \mathbb{R}$ (zero divisors).

1.14 Proposition. *Let R be an integral domain. If $a, b, c \in R$ with $a \neq 0$ and $ab = ac$, then $b = c$.*

Proof. Since $ab = ac$, $ab - ac = 0$, so $a(b - c) = 0$. Since $a \neq 0$ and R is an integral domain, we must have $b - c = 0$, i.e., $b = c$.

1.15 Remark. The above is true in any commutative ring when a is not a zero divisor.

1.16 Definition. Let R be a commutative, unital ring. We say $a \in R$ is a *unit* (or is invertible) if there exists $b \in R$ such that $ab = 1$. We call b the *inverse* of a and write $b = a^{-1}$. We denote the set (group) of units of R by $R^\times$ or $\mathcal{U}(R)$.

1.17 Example. Let $1 < n \in \mathbb{Z}$.

(a) If n is prime, then $\mathcal{U}(\mathbb{Z}_p) = \mathbb{Z}_p \setminus \{0\}$.

(b) In general, $\mathbb{Z}_n^\times = \{a \in \mathbb{Z}_n : \gcd(a, n) = 1\}$.

1.18 Remark. If $a \in R^\times$, then a is *not* a zero divisor, since $a \neq 0$ and $ab = 0$ implies $a^{-1}ab = 0$, i.e., $b = 0$.

1.19 Definition. A ring F is a field if F is commutative and unital and every non-zero element is a unit.

1.20 Example. The following are fields:

(a) Z_p, where p is prime

(b) $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$

(c) $Q(\sqrt{2})$

(d) $F(x)$, the set of rational functions over F, where F is a field.

1.21 Proposition. *Every field is an integral domain.*

Proof. If F is a field, then F is commutative and unital by definition. Furthermore, since every non-zero element is a unit, F has no zero divisors by 1.18. Thus F is an integral domain.

1.22 Remark. The converse of 1.21 is not true: $\mathbb{Z}$ and $F[x]$ for any field F are integral domains but not fields.

1.23 Definition. Let R be a unital ring. We define the *characteristic* of R to be the least positive integer n such that $n = 0$ in R. That is,

$$n \coloneqq n \cdot 1 = \underbrace{1 + 1 + \cdots + 1}_{n \text{ times}} = 0.$$

If no such n exists, we say that R has characteristic 0. Our notation is $\operatorname{char}(R) = 0$ or $\operatorname{char}(R) = n$.

1.24 Example. If $R = \mathbb{Z}_4[x]$, then $\operatorname{char}(R) = 4$.

1.25 Remark. Let R be a ring with characteristic 0. Then each of $1, 2, 3, \ldots$ is distinct, and thus R is infinite.

1.26 Proposition. *If R is an integral domain, then* $\mathrm{char}(R) = 0$ *or* $\mathrm{char}(R) = p$*, where p is prime.*

Proof. If $\mathrm{char}(R) = n \neq 0$ and n is not prime, then $n = ab$ when $a, b < n$, and thus $ab = 0$ in R. But R has not zero divisors, so n must be either 0 or prime. ∎

1.27 Example. If $R = \mathbb{Z}_p(x)$, then $\mathrm{char}(R) = p$.

1.28 Definition. Let $(R, +, \cdot)$ be a ring. We say that $S \subseteq R$ is a *subring* of R if $(S, +, \cdot)$ forms a ring.

1.29 Example. $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{Q}(\sqrt{2})$ are all subrings of $\mathbb{C}$.

May 10

1.30 Proposition (Subring Test)**.** *Let R be a ring and let $\varnothing \neq S \subseteq R$. Then S is a subring of R if and only if*

(1) *for all $a, b \in S$, $a - b \in S$*

(2) *for all $a, b \in S$, $ab \in S$.*

Proof. Clearly if S is a subring of R, then conditions (1) and (2) hold. Conversely, suppose conditions (1) and (2) hold. Let $a, b \in S$. Then $0 = a - a \in S$, so $0 \in S$. Additionally, $0 - b = -b \in S$, so S contains additive inverses. Finally, $a + b = a - (-b) \in S$ and $ab \in S$, so S is closed under the operations of R. ∎

1.31 Example. We claim that $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$ is a subfield of $\mathbb{R}$. Indeed, $\mathbb{Q}(\sqrt{2})$ is a subring of $\mathbb{R}$ by the subring test, and for any $0 \neq a + b\sqrt{2} \in \mathbb{Q}(\sqrt{2})$,

$$\left(a + b\sqrt{2}\right)^{-1} = \frac{a - b\sqrt{2}}{a^2 - 2b^2}.$$

Note that $a^2 - 2b^2 \neq 0$ by the irrationality of $\sqrt{2}$.

1.32 Definition. Let R be a ring. A subring I of R is an *ideal* of R if for all $a \in I$, $r \in R$, $ar, ra \in I$.

1.33 Example. $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$.

1.34 Example. Let $R = \mathcal{C}(\mathbb{R})$. Then $I = \{f(x) \in R : f(2) = 0\}$ is an ideal of R.

1.35 Example. $\mathbb{R}$ is a subring of $\mathbb{C}$ but not an ideal.

1.36 Remark. If F is a field, then the only ideals of F are $\{0\}$ and F.

1.37 Definition. Let R be a commutative, unital ring. The ideal $\langle x \rangle \coloneqq \{rx : r \in R\}$ is called the *principal ideal* of R generated by x.

1.38 Proposition (Division Algorithm)**.** *Let F be a field. For all $f(x), g(x) \in F[x]$ with $g(x) \neq 0$, there exist unique $q(x), r(x) \in F[x]$ such that $f(x) = g(x)q(x) + r(x)$, where $r(x) = 0$ or $\deg r(x) < \deg g(x)$.*

Proof. MATH 145. ∎

1.39 Proposition. *Let F be a field. Every ideal of F is principal.*

Proof. Let I be an ideal of $F[x]$. If $I = \{0\}$, then $I = \langle 0 \rangle$. Otherwise, let $g(x) \in I$ be nonzero of minimal degree in I. We claim that $I = \langle g(x) \rangle$.

Clearly $\langle g(x) \rangle \subseteq I$. We now show that $I \subseteq \langle g(x) \rangle$. Let $f(x) \in I$. By the Division Algorithm, there exist $q(x), r(x) \in F[x]$ so that $f(x) = g(x)q(x) + r(x)$, where $r(x) = 0$ or $\deg r(x) < \deg g(x)$. But $r(x) = f(x) - g(x)q(x)$, and $f(x), g(x)q(x) \in I$, so $\deg r(x) \geq \deg g(x)$ by minimality. Thus $r(x) = 0$, so $f(x) = g(x)q(x) \in \langle g(x) \rangle$. Hence $I \subseteq \langle g(x) \rangle$, and in fact $I = \langle g(x) \rangle$. ∎

2 Polynomials of Linear Operators

May 13

2.1 Notation. Throughout this section, unless otherwise stated, F is a field and V is a finite-dimensional vector space over F.

2.2 Definition. For $A \in M_n(F)$, the *characteristic polynomial* of A is $\det(A - xI)$. For $T\colon V \to V$, the *characteristic polynomial* of T is $\det([T]_\beta - xI)$, where β is any basis for V.

2.3 Definition. Let T be a linear operator. We say a subspace $W \leq V$ is *T-invariant* if $T(W) \subseteq W$.

2.4 Remark. If W is T-invariant, then $T_W : W \to W$ is well-defined.

2.5 Example. Consider $T\colon \mathbb{R}^2 \to \mathbb{R}^2$, $T(x, y) = (x + 2y, 4y - x)$. Then $W = \{(x, x) : x \in \mathbb{R}\}$ is T-invariant.

2.6 Example. Let $T\colon V \to V$ be a linear operator, and let λ be an eigenvalue for T. If $v \in E_\lambda$, then $T(T(v)) = T(\lambda(v)) = \lambda T(v)$, so $T(v) \in E_\lambda$. Thus $E_\lambda = \{v \in V : T(v) = \lambda v\}$ is T-invariant.

2.7 Definition. Let $T\colon V \to V$ be a linear operator. Let $0 \neq x \in V$. The subspace

$$W_{T,x} \coloneqq \operatorname{Span}\left\{x, T(x), T^2(x), \dots\right\}$$

is called the *T-cyclic subspace* generated by x.

2.8 Remark. $W_{T,x}$ is the smallest T-invariant subspace of V containing x.

2.9 Proposition. *Let $T\colon V \to V$ be a linear operator. Let $W \leq V$ be T-invariant. Then the characteristic polynomial of T_W divides the characteristic polynomial of T.*

Proof. Let $\beta = \{v_1, v_2, \dots, v_m\}$ be a basis for W. Say $[T_W]_\beta = A$. Extend β to a basis

$$\gamma = (v_1, v_2, \dots, v_m, v_{m+1}, \dots, v_n)$$

for V. Say $[T]_\gamma = B$. Then

$$B = \left[\begin{array}{c|c} A & \star \\ \hline 0 & A' \end{array}\right],$$

so $\det(B - xI) = \det(A - xI)\det(A' - xI)$. Thus the characteristic polynomial of T_W divides the characteristic polynomial of T. ∎

2.10 Proposition. *Let $T\colon V \to V$ be a linear operator and $v \in V \neq 0$. Let $W = W_{T,v}$, and say $\dim W = k$.*

(1) $\left\{v, T(v), T^2(v), \dots, T^{k-1}(v)\right\}$ *is a basis for W.*

(2) *If $f(x) = x^k + a_{k-1}x^{k-1} + \cdots + a_1 x + a_0 \in F[x]$ and $f(T)(v) = 0$, then the characteristic polynomial of T_W is $(-1)^k f(x)$.*

Proof.

(1) Let $j \in \mathbb{N}$ be maximal so that $\beta = \left\{v, T(v), \dots, T^{j-1}(v)\right\}$ is linearly independent. (Note that since $v \neq 0$, j must exist.) We claim that $j = k$.

Let $U = \operatorname{Span}\beta$. We will show that $U = W$. Now, since $\left\{v, T(v), \dots, T^{j-1}(v), T^j(v)\right\}$ is linearly dependent, $T^j(v) \in U$. Thus U is T-invariant. Since $W = W_{T,v}$ is the smallest T-invariant subspace of V containing v, $W \subseteq U$. But clearly $U \subseteq W$, so $U = W$, and thus $j = k$.

(2) From (1), $\beta = \left\{v, T(v), \dots, T^{k-1}(v)\right\}$ is a basis for W. Moreover, $f(T)(v) = 0$, so

$$a_0 v + a_1 T(v) + \cdots + a_{k-1}T^{k-1}(v) + T^k(v) = 0,$$

i.e., $T^k(v) = -a_0 v - a_1 T(v) - \cdots - a_{k-1}T^{k-1}(v)$. Therefore,

$$[T_W]_\beta = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & & \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{bmatrix}.$$

By Assignment 1, the characteristic polynomial of T_W is $(-1)^k f(x)$. ∎

2.11 Theorem (Cayley-Hamilton)**.** *If $T\colon V\to V$ is a linear operator and $f(x)\in F[x]$ is its characteristic polynomial, then $f(T)=0$.* *May 15*

Proof. Let $T\colon V\to V$ be a linear operator and $f(x)\in F[x]$ be its characteristic polynomial. Since $f(T)$ is linear, $f(T)(0)=0$. Let $0\neq v\in V$. We claim that $f(T)(v)=0$.

Let $W=W_{T,v}$ and say $\dim W=k$. Since $\{v,T(v),\dots,T^{k-1}\}$ is a basis for W by 2.10, the set $\{v,T(v),\dots,T^{k-1}(v),T^k(v)\}$ is linearly dependent. Thus there exist $a_0,\dots,a_k\in F$, not all 0, such that

$$a_0v+a_1T(v)+\cdots+a_{k-1}T^{k-1}(v)+a_kT^k(v)=0.$$

We may assume without loss of generality that $a_k=1$. Let $g(x)=x^k+a_{k-1}x^{k-1}+\cdots+a_1x+a_0$, so that $g(T)(v)=0$. Since $\deg g(x)=k$, the characteristic polynomial of T_W is $h(x)=(-1)^kg(x)$ by 2.10. Since $h(x)|f(x)$ by 2.9 and $h(T)(v)=(-1)^kg(T)(v)=0$, it follows that $f(T)(v)=0$. Thus $f(T)=0$. ∎

2.12 Remark. Let $T\colon V\to V$ be a linear operator. $I=\{f(x)\in F[x]:f(T)=0\}$ is an ideal of $F[x]$ and hence a principal ideal generated by some polynomial of least degree in I. Note that if $a(x),b(x)\in F[x]$ and $\langle a(x)\rangle=\langle b(x)\rangle$, then $a(x)=cb(x)$ for some $0\neq c\in F$. Thus there is only one *monic* polynomial of least degree in I, i.e., only one monic $m(x)$ such that $I=\langle m(x)\rangle$.

2.13 Definition. We call the polynomial $m(x)$ from 2.12 the *minimal polynomial* for T.

2.14 Remark. Suppose $f(x)\in F[x]$ such that $f(T)=0$. Then $m(x)|f(x)$. In particular, $m(x)$ divides the characteristic polynomial of T by the Cayley-Hamilton Theorem.

2.15 Remark. We similarly define the minimal polynomial of $A\in M_n(F)$ to be the unique monic $m(x)$ of least degree such that $m(A)=0$.

2.16 Proposition. *Let $T\colon V\to V$ be a linear operator with minimal polynomial $m(x)$ and characteristic polynomial $f(x)$. Then $m(x)$ and $f(x)$ have the same roots in F.*

Proof. First note that since $m(x)|f(x)$, every root of $m(x)$ is a root of $f(x)$. If T has no eigenvalues, then $f(x)$ is irreducible, and thus $f(x)=(-1)^km(x)$, and obviously every root of $f(x)$ is a root of $m(x)$. Otherwise, let λ be an eigenvalue of T. We claim that $m(\lambda)=0$.

Let $0\neq v\in V$ be an eigenvector for λ. Then $m(\lambda)v=m(\lambda v)=m(T(v))=m(T)(v)=0(v)=0$. Since $v\neq 0$, it follows that $m(\lambda)=0$. Thus every root of $f(x)$ is a root of $m(x)$, and we're done. ∎

2.17 Example. Let $V=P_2(\mathbb{R})=\{f(x)\in\mathbb{R}[x]:\deg f(x)\leq 2\}$. Consider $T\colon V\to V$, $T(g(x))=g'(x)+2g(x)$. Let $\beta=\{1,x,x^2\}$ be a basis for V. Then $T(1)=2$, $T(x)=1+2x$, and $T(x^2)=2x+2x^2$, so

$$A=[T]_\beta=\begin{bmatrix}2&1&0\\0&2&2\\0&0&2\end{bmatrix}.$$

Thus the characteristic polynomial of T is $-(x-2)^3$. The minimal polynomial $m(x)$ of T must be $x-2$, $(x-2)^2$, or $(x-2)^3$. Note that $A-2I\neq 0$, and

$$(A-2I)^2=\begin{bmatrix}0&1&0\\0&0&2\\0&0&0\end{bmatrix}^2=\begin{bmatrix}0&0&2\\0&0&0\\0&0&0\end{bmatrix}\neq 0,$$

so $m(x)=(x-2)^3$.

May 17

2.18 Example. Let

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

Then the characteristic polynomial of A is

$$\det(A - xI) = \left|\begin{array}{cc|c} 3-x & -1 & 0 \\ 0 & 2-x & 0 \\ \hline 1 & -1 & 2-x \end{array}\right| = (3-x)(2-x)^2 = -(x-3)(x-2)^2.$$

So $m(x) = (x-3)(x-2)^2$ or $(x-3)(x-2)$. But

$$(A-3I)(A-2I) = \begin{bmatrix} 0 & -1 & 0 \\ 0 & -1 & 0 \\ 1 & -1 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

so $m(x) = (x-3)(x-2)$.

2.19 Definition. Let $T\colon V \to V$ be a linear operator. We say that V is T-cyclic if there exists $0 \neq v \in V$ such that $V = W_{T,v}$.

2.20 Proposition. *Let $T\colon V \to V$ be a linear operator. If $\dim V = n$ and V is T-cyclic, then the characteristic polynomial $f(x)$ and the minimal polynomial $m(x)$ for T have the same degree. In particular, $f(x) = (-1)^n m(x)$.*

Proof. Suppose $V = W_{T,v}$ for some $0 \neq v \in V$. Recall that $\{v, T(v), \dots, T^{n-1}(v)\}$ is a basis for V. Let $g(x) = a_0 + a_1 x + \cdots + a_k x^k \in F[x]$ with $a_k \neq 0$ and $k < n$. Since $\{v, T(v), \dots, T^k(v)\}$ is linearly independent, $a_0 v + a_1 T(v) + \cdots + a_k T^k(v) \neq 0$, i.e., $g(T)(v) \neq 0$. Therefore $g(T) \neq 0$, and hence $\deg m(x) \geq n$. But $m(x) | f(x)$, so $\deg m(x) = n$, and since $m(x)$ is monic, $f(x) = (-1)^n m(x)$. ∎

2.21 Theorem. *Let $T\colon V \to V$ be a linear operator. Then T is diagonalizable if and only if the minimal polynomial $m(x)$ of T is of the form $m(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_k)$, where $\lambda_1, \lambda_2, \dots, \lambda_k \in F$ are the distinct eigenvalues of T.*

Proof. ($\Rightarrow$) Suppose T is diagonalizable. Then let $\beta = \{v_1, v_2, \dots, v_n\}$ be a basis of eigenvectors of T for V. Let $p(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_k)$, where $\lambda_1, \lambda_2, \dots, \lambda_k$ are the distinct eigenvalues of T. We claim that $m(x) = p(x)$.

Since each eigenvalue is a root of the minimal polynomial of T, $p(x)|m(x)$. Choose $v_i \in \beta$. Then $T(v_i) = \lambda_j v_i$ for some $1 \leq j \leq k$. In particular, $(T - \lambda_j I)(v_i) = 0$. But then since $x - \lambda_j | p(x)$, $p(x) = q_j(x)(x-\lambda_j)$, where $q_j(x) \in F[x]$. Then $p(T)(v_i) = q_j(T)(T-\lambda_j I)(v_i) = q_j(T)(0) = 0$. Since $v_i \in \beta$ was arbitrary, $p(T) = 0$. Therefore $m(x)|p(x)$. Since $p(x)$ is monic, $m(x) = p(x)$.

($\Leftarrow$) Suppose $m(x) = (x-\lambda_1)(x-\lambda_2)(x-\lambda_k)$, where $\lambda_1, \lambda_2, \dots, \lambda_k$ are the distinct eigenvalues of T. We proceed by induction on $n = \dim V$.

If $n = 1$, then T is clearly diagonalizable. If $n > 1$, assume the result for all vector spaces over F with dimension less than n. Let $W = \text{Range}(T - \lambda_k I)$. Now $E_{\lambda_k} = \text{Null}(T - \lambda_k I) \neq \{0\}$, so $\dim W < n$ by the Rank-Nullity Theorem. Moreover, since T commutes with both itself and λI, W is T-invariant.

Consider $T_W\colon W \to W$. Since the minimal polynomial for T_W divides $m(x)$, T_W is diagonalizable by assumption. Let $\beta = \{v_1, v_2, \dots, v_m\}$ be a basis for W of eigenvectors of T_W. Let $\gamma = \{w_1, w_2, \dots, w_\ell\}$ be a basis for $\text{Null}(T - \lambda_k I) = E_{\lambda_k}$. *May 22*

By the Rank-Nullity Theorem, $\dim V = n = m + \ell$. Let $y \in W$. Then $y = (T - \lambda_k)(x)$ for some $x \in V$. Then $m(T)(x) = (T-\lambda_1 I)(T-\lambda_2 I)(T-\lambda_{k-1} I)(y) = 0$. Therefore, $(T_W - \lambda_1 I)(T_W - \lambda_2 I)\cdots(T_W - \lambda_{k-1} I) = 0$, so the minimal polynomial for T_W divides $(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_{k-1})$. Hence λ_k is not an eigenvalue of T_W. Therefore, $W \cap E_{\lambda_k} = \varnothing$, and in particular $W \cap \gamma = \varnothing$. Thus $\beta \cup \gamma$ is linearly independent, and hence $\beta \cup \gamma$ is a basis of eigenvectors of T for V. Thus T is diagonalizable. ∎

2.22 Example. Let $A \in M_n(\mathbb{C})$ such that $A^m = I$. Then $m(x)|x^m - 1$. But $x^m - 1$ has m distinct roots $1, \zeta_m, \zeta_m^2, \dots, \zeta_m^{m-1}$, where $\zeta^m = e^{2\pi i/m}$, so $m(x)$ splits over $\mathbb{C}$ and has distinct roots over $\mathbb{C}$. By 2.21, A is diagonalizable.

3 Jordan Canonical Form

3.1 Generalized Eigenvectors and Eigenspaces

3.1.1 Notation. In keeping with 2.1, throughout this section, F is a field and V is a finite-dimensional vector space over F, unless otherwise stated.

3.1.2 Definition.

(1) $A \in M_n(F)$ is a Jordan block if

$$A = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix}.$$

(2) $J \in M_n(F)$ is a Jordan matrix if

$$J = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{bmatrix},$$

where $A_1, A_2, \ldots, A_k$ are Jordan blocks.

3.1.3 Example.

$$J = \left[\begin{array}{c|cc|ccc} 2 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{array}\right]$$

is a Jordan matrix with Jordan blocks

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

Say $[T]_\beta = J$, where $T\colon V \to V$ is a linear operator and $\beta = \{v_1, v_2, \ldots, v_6\}$ is a basis for V. Then we have

i	$T(v_i)$	$(T-2I)(v_i)$	$(T-2I)^2(v_i)$	$(T-3I)(v_i)$	$(T-3I)^2(v_i)$	$(T-3I)^3(v_i)$
1	$2v_1$	0	0	$\star$	$\star$	$\star$
2	$2v_2$	0	0	$\star$	$\star$	$\star$
3	$v_2 + 2v_3$	v_2	0	$\star$	$\star$	$\star$
4	$3v_4$	$\star$	$\star$	0	0	0
5	$v_4 + 3v_5$	$\star$	$\star$	v_4	0	0
6	$v_5 + 3v_6$	$\star$	$\star$	v_5	v_4	0

May 24

3.1.4 Example. Suppose $T\colon V \to V$ is a linear operator such that for some basis $\beta = \{v_1, v_2, \ldots, v_6\}$ we have:

i	1	2	3	4	5	6
$T(v_i)$	$5v_1$	$3v_2$	$v_2 + 3v_3$	$v_3 + 3v_4$	$2v_5$	$v_5 + 2v_6$

Then

$$A = [T]_\beta = \left[\begin{array}{c|ccc|cc} 5 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{array}\right]$$

Also note that $v_1 \in E_5 = \text{Null}(A-5I)$, $v_2 \in E_3 = \text{Null}(A-3I)$, $v_3 \in \text{Null}(A-3I)^2$, $v_4 \in Null(A-3I)^3$, $v_5 \in \text{Null}(A-2I)$, $v_6 \in \text{Null}(A-2I)^2$.

3.1.5 Definition. Let $T\colon V \to V$ be a linear operator whose characteristic polynomial splits over F. Let λ be an eigenvalue of T.

(1) We say that $0 \neq v \in V$ is a *generalized eigenvector* of T is $(T-\lambda I)^p(v) = 0$ for some $p \in \mathbb{N}$.

(2) We define $K_\lambda \coloneqq \{k \in V : \exists p \in \mathbb{N},\, (T-\lambda I)^p(v) = 0\}$ to be the *generalzed eigenspace* of V.

3.1.6 Remark. Equivalently, $K_\lambda = \bigcup_{i=1}^\infty \text{Null}(T-\lambda I)^i$.

3.1.7 Proposition. *Let $T\colon V \to V$ be a linear operator whose characteristic polynomial splits over F and λ an eigenvalue for T. Then K_λ is a T-invariant subspace which contains E_λ.*

Proof. Assignment 4. ∎

3.1.8 Proposition. *Let $T\colon V \to V$ be a linear operator whose characteristic polynomial splits over F. Suppose $\lambda \neq \mu$ be eigenvalues for T. Then $T - \lambda I\colon K_\mu \to K_\mu$ is one-to-one. In particular, $K_\lambda \cap K_\mu = \{0\}$.*

Proof. Let $0 \neq x \in K_\mu$, and suppose $x \in E_\lambda$. Let $p \in \mathbb{N}$ be minimal so that $(T-\mu I)^p(x) = 0$. If $p = 1$, $x \in E_\mu$, so $x \in E_\mu \cap E_\lambda = \{0\}$, and thus $x = 0$, which is a contradiction.

If $p > 1$, consider $y = (T-\mu I)^{p-1}(x) \neq 0$. Note that since $x \in E_\lambda$ and E_λ is $T-$ and λI-invariant, $y \in E_\lambda$. Then $(T-\mu I)(y) = (T-\mu I)^p(x) = 0$, so $y \in E_\mu$. But since $y \in E_\lambda \cup E_\mu$, $y = 0$, which is a contradiction. Therefore $\text{Null}(T-\lambda I) = 0$, so $T - \lambda I\colon K_\mu \to K_\mu$ is injective. It follows then that $(T-\lambda I)^p\colon K_\mu \to K_\mu$ is injective for all $p \in \mathbb{N}$. Thus $K_\mu \cap \text{Null}(T-\lambda I)^p = \{0\}$ for all $p \in N$, i.e., $K_\mu \cap K_\lambda = 0$. ∎

3.1.9 Proposition. *Let $T\colon V \to V$ be a linear operator whose characteristic polynomial splits over F. Suppose λ is an eigenvalue for T with multiplicity m. Then $\dim K_\lambda \leq m$ and $K_\lambda = \text{Null}(T-\lambda I)^m$.*

Proof. Let $W = K_\lambda$ and consider $T_W\colon W \to W$. Let $f(x)$ be the characteristic polynomial of T and $g(x)$ the characteristic polynomial of T_W. Recall that $g(x)|f(x)$. Also, if $\mu \neq \lambda$ is an eigenvalue of T, then $(T-\mu I)\colon W \to W$ is injective, if $(T-\mu I)(v) = 0$ for some $v \in W$, then $v = 0$. Thus, the only eigenvalue of T_W is λ. Therefore $g(x) = (-1)^d(x-\lambda)^d$, where $d = \dim W$. Since $g(x)|f(x)$, $\dim W = d \leq m$.

It is clear that $W = K_\lambda \supseteq \text{Null}(T-\lambda I)^m$. By the Cayley-Hamilton Theorem, $(T_W - \lambda I)^d = 0$. Let $w \in W$. Then $(T_W - \lambda I)^m(w) = (T-\lambda I)^m(w) = (T-\lambda I)^{m-d}(T-\lambda I)^d(w) = (T-\lambda I)^{m-d}(0) = 0$. Hence $W \subseteq \text{Null}(T-\lambda I)^m$, so we're done. ∎

3.1.10 Proposition. *Let $T\colon V \to V$ be a linear operator whose characteristic polynomial splits over F. Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be the distinct eigenvalues of T. For all $x \in V$, there exist $v_1 \in K_{\lambda_1}$, $v_2 \in K_{\lambda_2}, \dots v_k \in K_{\lambda_k}$ such that $x = v_1 + v_2 + \cdots + v_k$.*

May 27

Proof. By induction on k. Suppose $k = 1$ and $\lambda = \lambda_1$. Then the characteristic polynomial for T is $(-1)^d(x-\lambda)^d$, where $d = \dim V$. By the Cayley-Hamilton Theorem, $(T-\lambda I)^d = 0$. Thus $K_\lambda = V$, and the result follows: take $x = x$.

Inductively, assume the result for operators with fewer than k eigenvalues. Let m be the multiplicity of λ_k and let $W = \text{Range}(T-\lambda_k I)^m$. Note that W is T-invariant. Recall that for $i < k$, $(T-\lambda_k)\colon K_{\lambda_i} \to K_{\lambda_i}$ is injective, so $(T-\lambda_k)^m\colon K_{\lambda_i} \to K_{\lambda_i}$ is injective. In particular, $(T-\lambda_k I)^m(K_{\lambda_i}) \subseteq K_{\lambda_i}$. But $\dim K_{\lambda_i} < \infty$, so $(T-\lambda_k I)^m\colon K_{\lambda_i} \to K_{\lambda_i}$ is also surjective by the Rank-Nullity Theorem. Thus $(T-\lambda_k I)^m(K_{\lambda_i}) = K_{\lambda_i}$, so $K_{\lambda_i} \subseteq W = \text{Range}(T-\lambda_k I)^m$. Thus $\lambda_1, \lambda_2, \dots, \lambda_{k-1}$ are eigenvalues of $T_W\colon W \to W$. By the argument in the second half of 2.21, it follows that λ_k is not an eigenvalue of T_W. Let $x \in V$. By assumption, we know that $(T-\lambda_k)^m(x) = w_1 + w_2 + \cdots + w_{k-1}$, where $w_i \in K_{\lambda_i}$. Since $(T-\lambda_k I)^m$ is onto, for every w_i there exists $v_i \in K_{\lambda_i}$ such that $(T-\lambda_k I)^m(v_i) = w_i$. Then

$$(T-\lambda_k I)^m(x) = (T-\lambda_k I)^m(v_1) + \cdots + (T-\lambda_k I)^m(v_{k-1}),$$

so

$$(T-\lambda_k I)^m(x - v_1 - v_2 - \cdots - v_{k-1}) = 0,$$

implying that $x - v_1 - v_2 - \cdots - v_{k-1} \in \text{Null}(T-\lambda_k I)^m$. Thus $x = v_1 + v_2 + \cdots + v_{k-1} + v_k$ for some $v_k \in \text{Null}(T-\lambda_k I)^m = K_{\lambda_k}$. This completes the proof. ∎

3.1.11 Theorem. *Let $T\colon V \to V$ be a linear operator whose characteristic polynomial splits over F. Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be distinct eigenvalues for T with multiplicities $m_1, m_2, \dots, m_k$ respectively. For each $1 \le i \le k$, let β_i be a basis for K_{λ_i}. Then*

(1) $\beta_i \cap \beta_j = \varnothing$ *when* $i \neq j$.

(2) $\beta \coloneqq \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$ *is a basis for* V.

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

Proof.

(1) $K_{\lambda_i} \cup K_{\lambda_j} = \{0\}$ when $i \neq j$ by 3.1.8, and thus $\beta_i \cap \beta_j = \varnothing$.

(2) By (1), β is linearly independent. Also, β spans V by 3.1.10. Thus β is a basis for β.

(3) $\dim V = |\beta| = |\beta_1| + |\beta_2| + \cdots + |\beta_k| \le m_1 + m_2 + \cdots + m_k = \dim V$, and thus $m_i = |\beta_i| = \dim K_{\lambda_i}$ for all $1 \le i \le k$.

3.2 Finding the Jordan Canonical Form of a Matrix

3.2.1 Algorithm. Let $T\colon V \to V$ be a linear operator with characteristic polynomial

$$f(x) = (-1)^n \prod_{i=1}^{k} (x - \lambda_i)^{m_1}.$$

(1) Let $A = [T]_\sigma$, where σ is the standard basis for V, and let J be a Jordan matrix which is similar to A.

(2) Fix $\lambda = \lambda_1$. Compute $d_1 \coloneqq \dim \operatorname{Null}(A - \lambda I) = \dim E_\lambda$. Say a basis for $\operatorname{Null}(A - \lambda I)$ is γ_1. Since we use γ_1 to make the first columns of the λ-Jordan blocks, d_1 is the number of λ-Jordan blocks in J.

(3) Compute $d_2 \coloneqq \dim \operatorname{Null}(A - \lambda I)^2$. We then extend γ_1 to a basis γ_2 for $\operatorname{Null}(A - \lambda I)^2$ by solving $(A - \lambda I)x = v$ for each $v \in \gamma_1$. Since we use $\gamma_2 \setminus \gamma_1$ to make our second colums, $d_2 - d_1$ is the number of λ-Jordan blocks of size at least 2×2. *May 29*

(4) Compute $d_3 = \dim \operatorname{Null}(A - \lambda I)^3$. Then $d_3 - d_2$ is the number of λ-Jordan blocks of size at least 3×3.

(5) Continue in this fashion until $d_\ell = m_1 = \dim K_\lambda$, and thus γ_ℓ is a basis for K_λ.

(6) Repeat for $\lambda_2, \dots, \lambda_k$. If β_i is a basis for each K_{λ_i}, then $\beta = \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$ is a basis for V.

(7) If the β_is are computed as above, then $[T]_\beta = J$, and $A = PJP^{-1}$, where $P = [I]_\beta^\sigma$ and J is a Jordan matrix.

3.2.2 Remark. Any J computed in this way is called "the" Jordan Canonical Form of T (or A). It is unique up to reordering of the Jordan blocks.

3.2.3 Example. Let

$$A = \begin{bmatrix} 3 & -2 \\ 8 & -5 \end{bmatrix}.$$

Then $f(x) = (x+1)^2$, so $\lambda = -1$. We have

$$\operatorname{Null}(A - \lambda I) = \operatorname{Null}(A + I) = \operatorname{Null}\left(\begin{bmatrix} 1 & -1/2 \\ 0 & 0 \end{bmatrix} \right) = \operatorname{Span}\left(\begin{bmatrix} 1/2 \\ 1 \end{bmatrix} \right).$$

Thus $d_1 = 1$ and

$$\gamma_1 = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}.$$

We now know that
$$J = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}.$$
Solving
$$(A+I)v = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$
we get
$$v = \begin{bmatrix} 1/4 \\ 0 \end{bmatrix}.$$
Thus
$$\gamma_2 = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1/4 \\ 0 \end{bmatrix} \right\} = \beta,$$
so
$$P = [I]_\beta^\sigma = \begin{bmatrix} 1 & 1/4 \\ 2 & 0 \end{bmatrix}.$$

3.2.4 Example. Let
$$A = \begin{bmatrix} 3 & 1 & -2 \\ -1 & 0 & 5 \\ -1 & -1 & 4 \end{bmatrix}.$$
Then $f(x) = -(x-3)(x-2)^2$, so $\lambda_1 = 3$, $\lambda_2 = 2$, $m_1 = 1$ and $m_2 = 2$.

$\lambda_1 = 3$: Then $1 \leq d_1 = \dim \operatorname{Null}(A - 3I) \leq 1$, so $d_1 = 1$ and our Jordan block must be $[3]$.

$\lambda_2 = 2$: Then $1 \leq d_1 = \dim \operatorname{Null}(A - 2I) \leq 2$. Note that
$$A - 2I = \begin{bmatrix} 1 & 1 & -2 \\ -1 & -2 & 5 \\ -1 & -1 & 2 \end{bmatrix},$$
so inspecting the rows of $A - 2I$ shows that $\operatorname{rank}(A - 2I) = 2$ and hence $d_1 = 1$. Thus our Jordan block must be
$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix},$$
and
$$J = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

3.2.5 Example. Let $T\colon P_2(\mathbb{R}) \to P_2(\mathbb{R})$, $T(f(x)) = 2f(x) - f'(x)$. Then $\sigma = \{x^2, x, 1\}$. Find a Jordan Canonical basis for T, i.e., a basis β such that $[T]_\beta = J$.
$$[T]_\sigma = \begin{bmatrix} 2 & 0 & 0 \\ -2 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix},$$
so $f(x) = -(x-2)^3$. Note that $(T - 2I)(f(x)) = 0 \iff 2f(x) - f'(x) - 2f(x) = 0 \iff -f'(x) = 0$, so a basis for $\operatorname{Null}(T - 2I)$ is $\{1\} = \{v_1\}$ and hence $d_1 = 1$. So we must have
$$J = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$
Similarly, $(T - 2I)(f(x)) = 1 \iff -f'(x) = 1$, so we can take $v_2 = -x$. Finally, $(T - 2I)(f(x)) = -x \iff -f'(x) = -x$, so we can take $v_3 = \frac{1}{2}x^2$. Thus $(v_1, v_2, v_3) = \left(1, -x, \frac{1}{2}x^2\right)$ is a Jordan Canonical basis for T. Then $[T]_\sigma = P[T]_\beta P^{-1}$, where
$$P = \begin{bmatrix} 0 & 0 & 1/2 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

3.2.6 Example. Let $T\colon M_2(\mathbb{R}) \to M_2(\mathbb{R})$, where *May 31*

$$T(A) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} A.$$

Then

$$\sigma = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

$$[T]_\sigma = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

so $f(x) = (x-1)^4$. Then with $\lambda = 1$, $d_1 = \dim \operatorname{Null}(A-I) = 4 - \operatorname{rank}(A-I) = 2$, so we must have 2 Jordan blocks. Also, $d_2 = \dim \operatorname{Null}(A-I)^2 = \dim \operatorname{Null}(0) = 4$, so $d_2 - d_1 = 4 - 2 = 2$, and thus both Jordan blocks must be 2×2. So,

$$J = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

3.2.7 Proposition. *Let $T\colon V \to V$ be a linear operator with minimal polynomial*

$$m(x) = \prod_{i=1}^{k} (x - \lambda_i)^{m_i},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the distinct eigenvalues of T. Then m_i is the size of the largest λ_i-Jordan block in the Jordan Canonical Form of T.

Proof. Let $[T]_\sigma = A$. Then $A = PJP^{-1}$, where

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_\ell \end{bmatrix},$$

where each J_i is a Jordan block corresponding to some eigenvalue for T. Then

$$0 = m(J) = \begin{bmatrix} m(J_1) & & & \\ & m(J_2) & & \\ & & \ddots & \\ & & & m(J_\ell) \end{bmatrix},$$

so $M(J_i) = 0$ for all i.

Fix λ_i and let J_i be a λ_i-Jordan block. For any $j \neq i$, $\det(J_i - \lambda_j I) \neq 0$, since $\lambda_i - \lambda_j \neq 0$. Thus

$$0 = m(J_i) = (J_i - \lambda_i I)^{m_i} \prod_{j \neq i} (J_i - \lambda_j)^{m_j}.$$

But since $\det(J_i - \lambda_j I) \neq 0$ for any $j \neq i$, $J_i - \lambda_j I$ is invertible for each $j \neq i$. Thus we must have $(J_i - \lambda_i I)^{m_i} = 0$. But then

$$\begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix}^{m_i} = 0.$$

Note that if $J_i - \lambda I$ is $p \times p$, then $(J_i - \lambda_i)^p = 0$. By the minimality of $m(x)$, m_i must be the size of the largest λ_i-Jordan block. ∎

4 Inner Product Spaces

4.1 Foundations

4.1.1 Convention. Throughout Section 4, we shall use F to denote either $\mathbb{R}$ or $\mathbb{C}$, and V to denote a (possibly infinite-dimensional) vector space over F.

4.1.2 Definition. An *inner product* on a vector space V is a map $\langle \cdot, \cdot \rangle \colon V \times V \to F$ such that for all $x, y, z \in V, \alpha \in F$,

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

(2) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$

(3) $\langle y, x\rangle = \overline{\langle x, y\rangle}$

(4) $\langle x, x\rangle \in \mathbb{R}$, $\langle x, x\rangle \geq 0$, and $\langle x, x\rangle = 0 \iff x = 0$.

4.1.3 Remark. The following are immediate from the definition of inner product:

(1) $\langle x, x\rangle = \overline{\langle x, x\rangle}$, so $\langle x, x\rangle \in \mathbb{R}$

(2) $\langle x, y + z\rangle = \overline{\langle y + z, x\rangle} = \overline{\langle y, x\rangle + \langle z, x\rangle} = \overline{\langle y, x\rangle} + \overline{\langle z, x\rangle} = \langle x, y\rangle + \langle x, z\rangle$

(3) $\langle x, \alpha y\rangle = \overline{\langle \alpha y, x\rangle} = \overline{\alpha\langle y, x\rangle} = \overline{\alpha}\langle x, y\rangle$

(4) $\langle x, 0\rangle = \overline{0}\langle x, 0\rangle = 0$

(5) $\langle 0, x\rangle = 0\langle 0, x\rangle = 0$.

4.1.4 Definition. If V is equipped with an inner product, we call V an *inner product space*.

4.1.5 Proposition. *Let V be an inner product space. If $y, z \in V$ and for all $x \in V$, $\langle x, y\rangle = \langle x, z\rangle$ then $y = z$. In particular, if $\langle x, y\rangle = 0$ for all $x \in V$, then $y = 0$.*

Proof. Suppose $y, z \in V$ satisfy the condition in the proposition statement. Then for all $x \in V$,

$$\langle x, y\rangle = \langle x, z\rangle \implies \langle x, y\rangle - \langle x, z\rangle = 0 \implies \langle x, y - z\rangle = 0.$$

In particular, $\langle y - z, y - z\rangle = 0$, which implies that $y - z = 0$, i.e., $y = z$.

June 3

4.1.6 Example. Let $V = F^n$. The *standard inner product*, or *dot product* is given by

$$v \cdot w \coloneqq \langle v, w\rangle = \sum_{i=1}^{n} v_i \overline{w_i},$$

for any vectors $v = (v_1, v_2, \ldots, v_n)$ and $w = (w_1, w_2, \ldots, w_n)$ in F^n. Additionally, any real scalar multiple of the dot product also forms an inner product on F^n, e.g., $\langle v, w\rangle' = 2\langle v, w\rangle$.

4.1.7 Example. Let $V = \mathcal{C}[a, b]$. An inner product on V is given by

$$\langle f, g\rangle = \frac{1}{b - a}\int_a^b f(x)g(x)\,dx.$$

4.1.8 Definition. Let $A = (a_{ij}) \in M_n(F)$. The *adjoint* (or *conjugate transpose*) of A is $A^* \in M_n(F)$ defined by $A^* = (\overline{a}_{ji})$.

4.1.9 Example. If

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

then

$$A^* = \begin{bmatrix} 1 + i & -i \\ 2 - i & 4 \end{bmatrix}.$$

4.1.10 Example. Let $V = M_n(F)$. The *Frobenius inner product* is defined by

$$\langle A, b\rangle = \operatorname{tr}(B^*A)$$

Note that if $A = (a_{ij})$ and $B = (b_{ij})$, then $(B^*A)_{jj} = \sum_{i=1}^n \overline{b_{ij}} a_{ij}$. Therefore

$$\operatorname{tr}(B^*A) = \sum_{j=1}^n \sum_{i=1}^n \overline{b_{ij}} a_{ij} = \sum_{j=1}^n \sum_{i=1}^n a_{ij}\overline{b_{ij}} = v \cdot w,$$

where

$$\begin{aligned} v &= (a_{11}, a_{12}, \ldots, a_{1n}, a_{21}, a_{22}, \ldots, a_{2n}, \ldots, a_{n1}, a_{n2}, \ldots, a_{nn}) \\ w &= (b_{11}, b_{12}, \ldots, b_{1n}, b_{21}, b_{22}, \ldots, b_{2n}, \ldots, b_{n1}, b_{n2}, \ldots, b_{nn}). \end{aligned}$$

4.1.11 Example. Let $V = \ell^2(F) \coloneqq \left\{(x_n)_{n=1}^\infty \in F^{\mathbb{N}} : \sum_{i=1}^\infty |x_i|^2 < \infty\right\}$. An inner product on V is given by

$$\langle (x_n), (y_n)\rangle = \sum_{i=1}^\infty x_i \overline{y_i}.$$

4.1.12 Definition. A *norm* on a vector space V is a map $\|\cdot\| \colon V \to \mathbb{R}$ such that for all $v, w \in V$, $\alpha \in F$

(1) $\|v\| \geq 0$ and $\|v\| = 0 \iff v = 0$

(2) $\|\alpha v\| = |\alpha| \cdot \|v\|$

(3) $\|v + w\| \leq \|v\| + \|w\|$.

If V is equipped with a norm, we call it a *normed vector space*.

4.1.13 Theorem (Cauchy-Schwarz Inequality)**.** *Let V be an inner product space, and for $x \in V$, define $\|x\| = \sqrt{\langle x, x\rangle}$. Then for all $x, y \in V$, $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$.*

Proof. Let $x, y \in V$. If $y = 0$, the result is trivial. Otherwise, $\langle y, y\rangle > 0$. Then for any $\alpha \in F$,

$$0 \leq \|x - \alpha y\|^2 = \langle x - \alpha y, x - \alpha y\rangle = \langle x, x\rangle - \overline{\alpha}\langle x, y\rangle - \alpha\langle y, x\rangle + \alpha\overline{\alpha}\langle y, y\rangle.$$

In particular, when $\alpha = \frac{\langle x,y\rangle}{\langle y,y\rangle}$,

$$0 \leq \langle x, x\rangle - \frac{\langle y, x\rangle}{\langle y, y\rangle}\langle x, y\rangle - \frac{\langle x, y\rangle}{\langle y, y\rangle}\langle y, x\rangle + \frac{\langle x, y\rangle}{\langle y, y\rangle} \cdot \frac{\langle y, x\rangle}{\langle y, y\rangle}\langle y, y\rangle = \langle x, x\rangle - \frac{\langle x, y\rangle\langle y, x\rangle}{\langle y, y\rangle}.$$

It follows that $\langle x, x\rangle\langle y, y\rangle \geq \langle x, y\rangle\overline{\langle x, y\rangle}$, i.e., $\|x\|^2\|y\|^2 \geq |\langle x, y\rangle|^2$. Therefore $\|x\|\|y\| \geq |\langle x, y\rangle|$. YEAH!

4.1.14 Proposition. *Let V be an inner product space. Then setting $\|x\| = \sqrt{\langle x, x\rangle}$ for all $x \in V$ defines a norm on V.*

Proof. We will show that this choice of norm satisfies all the necessary properties. Let $x, y \in V$, $\alpha \in F$.

(1) $\|x\| = \sqrt{\langle x, x\rangle} \geq 0$, with $\|x\| = \sqrt{\langle x, x\rangle} = 0 \iff x = 0$.

(2) $\|\alpha x\| = \sqrt{\langle \alpha x, \alpha x\rangle} = \sqrt{|\alpha|^2\langle x, x\rangle} = |\alpha|\|x\|$.

(3) This one requires a little more work and the Cauchy-Schwarz inequality:

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y\rangle \\ &= \|x\|^2 + \|y\|^2 + \langle x, y\rangle + \langle y, x\rangle \\ &= \|x\|^2 + \|y\|^2 + \langle x, y\rangle + \overline{\langle x, y\rangle} \\ &= \|x\|^2 + \|y\|^2 + 2\Re(x, y) \end{aligned}$$

$$\begin{aligned}&\le \|x\|^2 + \|y\|^2 + 2|\langle x, y\rangle| \\ &\le \|x\|^2 + \|y\|^2 + 2\|x\| \cdot \|y\| \\ &= (\|x\| + \|y\|)^2 .\end{aligned}$$

so $\|x+y\| \le \|x\| + \|y\|$.

This completes the proof. ∎

June 5

4.1.15 Example. Let $v = (-1, i, 2+i) \in \mathbb{C}^3$. Then

$$\|v\| = \sqrt{(-1, i, 2+i) \cdot (-1, i, 2+i)} = \sqrt{1 + (i)(-i) + (2+i)(2-i)} = \sqrt{2+5} = \sqrt{7}.$$

4.1.16 Example. Let

$$v = \begin{bmatrix} -1 & 3-i \\ 4 & 1 \end{bmatrix} \in M_2(\mathbb{C}).$$

Using the norm induced by the Frobenius inner product,

$$\|v\| = \sqrt{\operatorname{tr}\left(\begin{bmatrix} -1 & 4 \\ 3+i & -i \end{bmatrix} \cdot \begin{bmatrix} -1 & 3-i \\ 4 & 1 \end{bmatrix}\right)} = \sqrt{17+11} = \sqrt{28} = 2\sqrt{7}.$$

4.1.17 Example. Let $f(x) = e^x \in \mathcal{C}[0,1]$. Then

$$\|f(x)\| = \sqrt{\int_0^1 e^{2x}\, dx} = \sqrt{\frac{1}{2}(e^2 - 1)}.$$

4.2 Orthogonality and Orthonormality

4.2.1 Proposition (Parallelogram Law)**.** *Let V be an inner product space. Then for all $x, y \in V$,*

$$\|x+y\|^2 + \|x-y\|^2 = 2 \cdot \|x\|^2 + 2\|y\|^2.$$

Proof. Let $x, y \in V$. Then

$$\begin{aligned}\|x+y\|^2 + \|x-y\|^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\ &= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle + \langle x,x\rangle - \langle x,y\rangle - \langle y,x\rangle + \langle y,y\rangle \\ &= 2\langle x,x\rangle + 2\langle y,y\rangle \\ &= 2\|x\|^2 + 2\|y\|^2,\end{aligned}$$

as required. ∎

4.2.2 Remark. We now begin to translate our geometric intuition into the language of norms and inner products. The previous proposition is a generalization of the parallelogram law in Euclidean geometry, which states that the sum of the squares of the diagonals of a parallelogram is equal to the sum of the squares of its sides.

Consider the Cosine Law in classical geometry: $c^2 = a^2 + b^2 - 2ab\cos C$. In $\mathbb{R}^2$, this translates to

$$\begin{aligned}\|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\|y\cos\theta &\implies \langle x,x\rangle - \langle x,y\rangle - \langle y,x\rangle + \langle y,y\rangle = \|x\|^2 + \|y\|^2 - 2\|x\|\|y\|\cos\theta \\ &\implies -2\langle x,y\rangle = 2\|x\|y\|\cos\theta \\ &\implies \cos\theta = \frac{\langle x,y\rangle}{\|x\|\|y\|}.\end{aligned}$$

(Note that we assume $x, y \neq 0$; we want a triangle, after all.) Thus x, y are perpendicular if and only if $\cos\theta = 0$, i.e., $\langle x, y\rangle = 0$. This gives us a generalization of the notion of "perpendicular" to abstract inner product spaces.

4.2.3 Definition. Let V be an inner product space. We say $u, v \in V$ are *orthogonal* if $\langle u, v\rangle = 0$. We say a subset $S \subseteq V$ is *orthogonal* if $\langle u, v\rangle = 0$ for all $u, v \in S$. If S is orthogonal and $\|u\| = 1$ for all $u \in S$, then we say S is *orthonormal.*

4.2.4 Example. The standard basis for F^n, $\sigma = \{e_1, e_2, \ldots, e_n\}$, is orthonormal.

4.2.5 Example. When considered as a subset of $\mathcal{C}[0, 1]$, $S = \{1, x, x^2\}$ is not orthogonal. However, when S is considered as a subset of $P_2(\mathbb{R}) \cong \mathbb{R}^3$, S is orthonormal.

4.2.6 Remark. Let V be an inner product space. Suppose $S = \{v_1, v_2, v_3, \ldots,\} \subseteq V \setminus \{0\}$ is orthogonal. Then $S' = \left\{\frac{1}{\|v_1\|}v_1, \frac{1}{\|v_2\|}v_2, \frac{1}{\|v_3\|}v_3, \ldots,\right\}$ is orthonormal.

4.2.7 Example. Let H be the collection of continuous functions from $[0, 2\pi]$ to $\mathbb{C}$. Then

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)\overline{g(t)}\,dt$$

is an inner product on H. Note that if $f(x) \in H$, then $f(x) = u(x) + iv(x)$, where $u, v \in \mathcal{C}[0, 2\pi]$, and $\int f(t)\,dt := \int u(t)\,dt + i\int v(t)\,dt$.

Let $f_n(t) = e^{int} = \cos(nt) + i\sin(nt)$, and let $S = \{f_n : n \in \mathbb{Z}\}$. Then S is orthonormal.

June 7

4.2.8 Proposition. *Let V be an inner product space, and let $S = \{v_1, v_2, \ldots, v_k\} \subseteq V$ be orthogonal such that $v_i \neq 0$ for all $1 \le i \le k$. If $y \in \operatorname{Span}(S)$ such that $y = \sum_{i=1}^k c_i v_i$, where $c_1, c_2, \ldots, c_k \in F$, then for all $i \le i \le k$,*

$$c_i = \frac{\langle y, v_i\rangle}{\|v_i\|^2}.$$

Proof. Let $y = \sum_{i=1}^k c_i v_i \in \operatorname{Span}(S)$. Then for each $1 \le i \le k$,

$$\langle y, v_i\rangle = \left\langle \sum_{i=1}^k c_i v_i, v_i \right\rangle = c_i\langle v_i, v_i\rangle = c_i\|v_i\|^2,$$

and since $v_i \neq 0$, we must have

$$c_i = \frac{\langle y, v_i\rangle}{\|v_i\|^2},$$

as required. ∎

4.2.9 Remark. In the above proposition, if S is in fact orthonormal, then $c_i = \langle y, v_i\rangle$.

4.2.10 Proposition. *Let V be an inner product space, and let $S \subseteq V$ be orthogonal consisting of nonzero vectors. Then S is linearly independent.*

Proof. Let $v_1, v_2, \ldots, v_k \in S$. Suppose $\sum_{i=1}^k c_i v_i = 0$ for some $c_1, c_2, \ldots, c_k \in F$. By Proposition 4.2.8,

$$c_i = \frac{\langle 0, v_i\rangle}{\|v_i\|^2} = 0$$

for $1 \le i \le k$, so S is linearly independent. ∎

4.2.11 Proposition. *Let $A \in M_n(F)$. Suppose*

$$A = \begin{bmatrix} r_1 \\ \hline r_2 \\ \hline \vdots \\ \hline r_n \end{bmatrix},$$

where $r_1, r_2, \ldots, r_n \in F^n$ and $\{r_1, r_2, \ldots, r_n\}$ is orthogonal. Then AA^* is diagonal. If $\{r_1, r_2, \ldots, r_n\}$ is orthonormal, then $AA^* = I$.

Proof. Let $AA^* = (a_{ij})$. Since

$$AA^* = \begin{bmatrix} \underline{\quad r_1 \quad} \\ \underline{\quad r_2 \quad} \\ \vdots \\ \overline{\quad r_n, \quad} \end{bmatrix} \cdot \left[\ \overline{r_1} \mid \overline{r_2} \mid \cdots \mid \overline{r_n} \ \right],$$

we see that

$$a_{ij} = \langle r_i, r_j \rangle = \begin{cases} 0 & i \neq j \\ \|r_i\|^2 & i = j \end{cases},$$

so AA^* is diagonal. In particular, if $\{r_1, r_2, \ldots, r_n\}$ is orthogonal, $\|r_i\|^2 = 1$, so $AA^* = I$. ∎

4.2.12 Algorithm (Gram-Schmidt Procedure)**.** Let V be an inner product space. Let $\{w_1, w_2, \ldots, w_n\} \subseteq V$ be linearly independent. We wish to produce an orthogonal set $\{v_1, v_2, \ldots, v_n\} \subseteq V$ such that

$$\operatorname{Span}\{w_1, w_2, \ldots, w_n\} = \operatorname{Span}\{v_1, v_2, \ldots, v_n\}.$$

We present a procedure for $n = 3$; it is easy to see how it could be adapted for larger numbers.

(1) Take $v_i = w_i$.

(2) Note: $\operatorname{Span}\{w_1, w_2\} = \operatorname{Span}\{v_1, w_2\} = \operatorname{Span}\{v_1, w_2 - \alpha v_1\}$ for any $\alpha \in F$. Solve for α so that

$$0 = \langle w_2 - \alpha v_1, v_1 \rangle \iff 0 = \langle w_2, v_1 \rangle - \alpha \langle v_1, v_1 \rangle \iff \alpha = \frac{\langle w_2, v_1 \rangle}{\|v_1\|^2}.$$

(3) Take $v_2 = w_2 - \alpha v_1 = w_2 - \frac{\langle w_2, v_1 \rangle}{\|v_1\|^2} v_1$.

(4) Note: $\operatorname{Span}\{w_1, w_2, w_3\} = \operatorname{Span}\{v_1, v_2, w_3\} = \operatorname{Span}\{v_2, v_2, w_3 - \alpha v_1 - \beta v_2\}$. Solve for α and β:

$$\alpha = \frac{\langle w_3, v_1 \rangle}{\|v_1\|^2}, \ \beta = \frac{\langle w_3, v_2 \rangle}{\|v_2\|^2}.$$

(5) Take $v_3 = w_3 - \alpha v_1 - \beta v_2 = w_3 - \frac{\langle w_3, v_1 \rangle}{\|v_1\|^2} - \frac{\langle w_3, v_2 \rangle}{\|v_2\|^2}$.

(6) Then $\{v_1, v_2, v_3\}$ is orthogonal with $\operatorname{Span}\{w_1, w_2, w_3\} = \operatorname{Span}\{v_1, v_2, v_3\}$, and $\left\{ \frac{1}{\|v_1\|} v_1, \frac{1}{\|v_2\|} v_2, \frac{1}{\|v_3\|} v_3 \right\}$ is orthonormal.

4.2.13 Theorem (Gram-Schmidt)**.** *Let V be an inner product space. If $S = \{w_1, w_2, \ldots, w_n\} \subseteq V$ is linearly independent, then $S^1 = \{v_1, v_2, \ldots, v_n\}$ defined recursively by*

$$v_k = \begin{cases} w_1 & k = 1 \\ w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} v_j & \text{otherwise} \end{cases}$$

is an orthogonal set of nonzero vectors such that $\operatorname{Span}(S) = \operatorname{Span}(S^1)$.

Proof. Apply the Gram-Schmidt procedure. ∎

4.2.14 Corollary. *If V is a finite-dimensional inner product space, then V has an orthonormal basis.*

4.2.15 Example. Let $W = \operatorname{Span}\{w_1 = [1, 1, 0], w_2 = [0, 2, 1]\} \subseteq \mathbb{R}^3$. Take $v_1 = w_1 = [1, 1, 0]$ and

$$v_2 = w_2 - \frac{\langle w_2, v_1 \rangle}{\|v_1\|^2} v_1 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} - \frac{2}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Thus $\left\{ \frac{1}{\sqrt{2}} v_1, \frac{1}{\sqrt{3}} v_2 \right\}$ is an orthonormal basis for W.

4.2.16 Remark. Suppose $\{v_1, v_2, \dots, v_n\} \subseteq V$ is orthogonal. Then *June 10*

$$\|v_1 + v_2 + \cdots + v_n\|^2 = \langle v_1 + v_2 + \cdots + v_n, v_1 + v_2 + \cdots + v_n \rangle = \|v_1\|^2 + \|v_2\|^2 + \cdots + \|v_n\|^2.$$

4.2.17 Remark. Recall from high school linear algebra that in $\mathbb{R}^2$, the projection of $\vec{v}$ onto $\vec{w}$ is given by

$$\operatorname{proj}_{\vec{v}} \vec{w} = \frac{\vec{w} \cdot \vec{v}}{\vec{v} \cdot \vec{v}} \vec{v}.$$

Geometrically, this explains the choice of α in the Gram-Schmidt procedure: setting $\alpha = \frac{\vec{w} \cdot \vec{v}}{\vec{v} \cdot \vec{v}}$ ensures that $\vec{w} - \alpha \vec{v}$ is perpendicular to $\vec{v}$.

4.2.18 Exercise. Find an orthogonal basis for $P_2(\mathbb{R}) \subseteq \mathcal{C}[0,1]$.

Solution. Let $\beta = \{w_1 = 1, w_2 = x, w_3 = x^2\}$ be the standard basis for $P_2(\mathbb{R})$. Take $v_1 = w_1 = 1$. Then let

$$v_2 = w_2 - \frac{\langle w_2, v_1 \rangle}{\|v_1\|^2} v_1 = x - \frac{\int_0^1 x \cdot 1 \, dx}{\int_0^1 1 \cdot 1 \, dx} \cdot 1 = x - \frac{1}{2}.$$

Also, let

$$v_3 = w_3 - \frac{\langle w_3, v_2 \rangle}{\|v_2\|^2} v_2 - \frac{\langle w_3, v_1 \rangle}{\|v_1\|^2} v_1 = x^2 - \frac{\int_0^1 x^3 - \frac{1}{2} x^2 \, dx}{\int_0^1 \left(x - \frac{1}{2}\right)^2 dx} \left(x - \frac{1}{2}\right) - \frac{\int_0^1 x^2 \, dx}{\int_0^1 1 \, dx} \cdot 1 = x^2 - x + \frac{1}{6}.$$

This gives us the orthonormal basis $\left\{1, x - \frac{1}{2}, x^2 - x + \frac{1}{6}\right\}$.

4.2.19 Exercise. A throwback to high school linear algebra: Find the closest point on $x = (1,2) + (1,-1)t$ to the point $(3,3)$.

Solution. Letting $\vec{v} = (1,-1)$ and $\vec{w} = (3,3) - (1,2) = (2,1)$, we have $\operatorname{proj}_{\vec{v}} \vec{w} = \frac{1}{2}(1,-1) = \left(\frac{1}{2}, -\frac{1}{2}\right)$. This gives the closest point as $(1,2) + \left(\frac{1}{2}, -\frac{1}{2}\right) = \left(\frac{3}{2}, \frac{3}{2}\right)$.

4.2.20 Remark. We want to generalize the notion of projection to abstract subspaces, not just lines.

4.2.21 Definition. Let A, B be subspaces of a vector space V. We say that V is a *direct sum* of A and B and write $V = A \oplus B$ if

(1) $A + B \coloneqq \{a + b : a \in A, b \in B\} = V$ and

(2) $A \cap B = \{0\}$.

4.2.22 Proposition. *Suppose $V = A \oplus B$ for some $A, B \leq V$.*

(1) *Every $v \in V$ can be uniquely written as $v = a + b$, where $a \in A, b \in B$.*

(2) *If α is a basis for A and β is a basis for B, then $\alpha \cup \beta$ is a basis for V. In particular, if V is finite-dimensional, then* $\dim V = \dim A + \dim B$.

Proof.

(1) Let $v \in V$. Since $V = A \oplus B$, there exist $a \in A$, $b \in B$ such that $v = a + b$; we just need to show uniqueness. Suppose v is also equal to $\tilde{a} + \tilde{b}$, where $\tilde{a} \in A$ and $\tilde{b} \in B$. Then $a + b = \tilde{a} + \tilde{b}$, so $A \ni a - \tilde{a} = \tilde{b} - b \in B$. Since $A \cap B = \{0\}$, $a - \tilde{a} = \tilde{b} - b = 0$, i.e., $a = \tilde{a}$ and $b = \tilde{b}$.

(2) Let $\alpha = \{v_1, v_2, v_3, \dots, \}$ and $\beta = \{w_1, w_2, w_3, \dots, \}$ be bases for A and B respectively. Since $V = A + B$, $\alpha \cup \beta$ spans V. Now, suppose $\sum_{i=1}^n c_i v_i + \sum_{i=1}^n d_i w_i = 0$ for some $c_1, c_2, \dots, c_n, d_1, d_2, \dots, d_n \in F$. Then

$$A \ni \sum_{i=1}^n c_i v_i = -\sum_{i=1}^m d_i v_i \in B,$$

and since α and β are each linearly independent, we must have

$$c_1 = c_2 = \cdots = c_n = 0 = d_1 = d_2 = \cdots = d_m.$$

Therefore $\alpha \cup \beta$ is linearly independent.

This completes the proof. ∎

4.2.23 Definition. Let V be an inner product space, and let $\varnothing \neq S \subseteq V$. The *orthogonal complement* of S is defined to be
$$S^\perp := \{x \in V : \langle v, x\rangle = 0 \text{ for all } v \in S\}.$$

4.2.24 Remark. For any $\varnothing \neq S \subseteq V$, $S^\perp$ is a subspace of V.

4.2.25 Theorem. *If W is a finite-dimensional subspace of an inner product space V, then $V = W \oplus W^\perp$.*

June 12

Proof. Let V be an inner product space and let $W \leq V$ be finite-dimensional. Let $\beta = \{v_1, v_2, \ldots, v_k\}$ be an orthonormal basis for W. Let $u = \sum_{i=1}^k \langle v, v_i\rangle v_i \in W$. Let $z = v - u$. Now, for every $1 \leq j \leq k$,
$$\langle z, v_j\rangle = \langle v-u, v_j\rangle = \langle v, v_j\rangle - \langle u, v_j\rangle = \langle v, v_j\rangle - \langle v, v_j\rangle\langle v_j, v_j\rangle = \langle v, v_j\rangle - \langle v, v_j\rangle = 0.$$
It follows that $z \in W^\perp$. Since $v = u + z$ and $u \in W$, $v = u + z$. Therefore $V = W + W^\perp$. Now, if $x \in W \cap W^\perp$, then $\langle x, x\rangle = 0$, so $x = 0$. Therefore $W \oplus W^\perp = 0$, and it follows that $V = W \oplus W^\perp$. ∎

4.2.26 Definition. Let V be an inner product space and $W \leq V$ be finite-dimensional. Let $\{v_1, v_2, \ldots, v_k\}$ be an orthonormal basis for W. For $v \in V$, we call
$$u = \sum_{i=1}^k \langle v, v_i\rangle v_i \in W$$
the *orthogonal projection* of v onto W, and we write $u = \operatorname{proj}_W(v)$. Note that this vector is unique.

4.2.27 Theorem. *Let W be a finite-dimensional subspace of an inner product space V. Let $v \in V$, so that there exist unique $u \in W$ and $z \in W^\perp$ such that $v = u + z$. Then for any $x \in W$, $\|v - x\| \geq \|v - u\|$, with equality if and only if $x = u$.*

Proof. Let $x \in W$. Note that $u - x \in W$ and $z \in W^\perp$, so
$$\|v-x\|^2 = \|u+z-x\|^2 = \|u-x+z\|^2 = \langle u-x+z, u-x+z\rangle = \|u-x\|^2 + \|z\|^2 \geq \|z\|^2,$$
so $\|v - x\| \geq \|z\| = \|v-u\|$. Equality holds if and only if $\|u-x\|^2 = 0$, i.e., $x = u$, so we're done. ∎

4.2.28 Example. Let $W = \operatorname{Span}\{(i, 0, 1+i), (0, -i, 1)\} \subseteq \mathbb{C}^3$. Note that $\dim W = 2$, so $\dim W^\perp = 1$. By inspection, we see that $(1-i, 1, -i) \in W^\perp$, so $W^\perp = \operatorname{Span}\{(1-i, 1, -i)\}$.

4.2.29 Exercise. Let $V = C[0,1]$ and let $W = P_1(\mathbb{R})$. Find the closest vector in W to $f(x) = e^x \in V$.

Solution. By Theorem 4.2.27, we must find $\operatorname{proj}_W(f(x))$. First, a basis for W is clearly $\{1, x\}$. Upon applying the Gram-Schmidt procedure to this basis, we see that $\left\{1, \sqrt{12}\left(x - \frac{1}{2}\right)\right\}$ is an orthonormal basis for W. Therefore,
$$\begin{aligned}
\operatorname{proj}_W(f(x)) &= \langle e^x, 1\rangle + \left\langle e^x, \sqrt{12}\left(x - \frac{1}{2}\right)\right\rangle \left(\sqrt{12}\left(x - \frac{1}{2}\right)\right) \\
&= \int_0^1 e^x\,dx + \left(\int_0^1 e^x\left(\sqrt{12}x - \sqrt{3}\right)dx\right)\left(\sqrt{12}x - \sqrt{3}\right) \\
&= e - 1 + \left(e(\sqrt{12} - \sqrt{3}) + \sqrt{3} - \sqrt{12}(e-1)\right)\left(\sqrt{12}x - \sqrt{3}\right) \\
&= e - 1 + (3-e)\sqrt{3}\left(\sqrt{12}x - \sqrt{3}\right) \\
&= e - 1 + (3-e)(6x - 3) \\
&= (18 - 6e)x + (4e - 10).
\end{aligned}$$

4.2.30 Exercise. Find the closest symmetric matrix to *June 14*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_2(\mathbb{R}).$$

Solution. Let $W = \{X \in M_2(\mathbb{R}) : X = X^T\}$. Note that a basis for W is given by

$$\gamma = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

Furthermore, by considering the Frobenius inner product we see that this is actually any orthogonal basis. If we use

$$\beta = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\},$$

then we have an orthonormal basis. Say $\beta = \{v_1, v_2, v_3\}$. Then the closest symmetric matrix to A is

$$\begin{aligned}
&\operatorname{proj}_W(A) \\
&= \langle A, v_1\rangle v_1 + \langle A, v_2\rangle v_2 + \langle A, v_3\rangle v_3 \\
&= \operatorname{tr}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \operatorname{tr}\left(\frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)\frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \operatorname{tr}\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 & b+c \\ b+c & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix} \\
&= \begin{bmatrix} a & \frac{1}{2}(b+c) \\ \frac{1}{2}(b+c) & d \end{bmatrix},
\end{aligned}$$

which is what we might expect intuitively.

4.3 The Adjoint

4.3.1 Definition. Let V be a vector space over a field F. We say that $T\colon V \to F$ is a *linear functional* if T is linear. The *dual space* V^* of V is the vector space of linear functionals on V.

4.3.2 Theorem (Riesz Representation Theorem)**.** *Let V be a finite-dimensional inner product space. Let $T\colon V \to F$ be a linear functional. Then there exists a unique $y \in V$ such that $T(x) = \langle x, y\rangle$ for all $x \in V$.*

Proof. Assignment 5. ∎

4.3.3 Proposition. *Let V be a finite-dimensional inner product space. Let $T\colon V \to V$ be linear. Then there exists a unique linear operator $T^*\colon V \to V$ such that $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$ for all $x, y \in V$.*

Proof. For arbitrary $y \in V$, $U_y\colon V \to F$ given by $U_y(x) = \langle T(x), y\rangle$ is a linear functional. By the Riesz Representation Theorem, there exists a unique $y' \in V$ such that $U_y(x) = \langle x, y'\rangle$ for all $x \in V$. Define $T^*\colon V \to V$ by $T^*(y) = y'$.

It remains to show that T^* is linear. Let $x, y_1, y_2 \in V$, $\alpha \in F$. Then

$$\begin{aligned}
\langle x, T^*(\alpha y_1 + y_2)\rangle &= \langle T(x), \alpha y_1 + y_2\rangle \\
&= \overline{\alpha}\langle T(x), y_1\rangle + \langle T(x), y_2\rangle \\
&= \overline{\alpha}\langle x, T^*(y_1)\rangle + \langle x, T^*(y_2)\rangle \\
&= \langle x, \alpha T^*(y_1)\rangle + \langle x, T^*(y_2)\rangle \\
&= \langle x, \alpha T^*(y_1) + T^*(y_2)\rangle.
\end{aligned}$$

Since x was arbitrary, $T^*(\alpha y_1 + y_2) = \alpha T^*(y_1) + T^*(y_2)$, and thus T^* is linear, so we're done. ∎

4.3.4 Definition. We call the function T^* constructed as in the proof of Proposition 4.3.3 the *adjoint* of T.

4.3.5 Proposition. *Let V be a finite-dimsional inner product space. Let β be an orthonormal basis for V, and let $T\colon V \to V$ be linear. Then $[T^*]_\beta = [T]_\beta^*$.*

Proof. Let $\beta = \{v_1, v_2, \dots, v_n\}$. Let $A = (a_{ij}) = [T]_\beta$. Let $B = (b_{ij}) = [T^*]_\beta$. Then by Proposition 4.2.8,

$$b_{ij} = \langle T^*(v_j), v_i\rangle = \overline{\langle v_i, T^*(v_j)\rangle} = \overline{\langle T(v_i), v_j\rangle} = \overline{a_{ji}},$$

so $A^* = B$, as required. ∎

4.3.6 Remark. Let V be a finite-dimensional inner product space. Let $T\colon V \to V$ be linear. Then for all $x, y \in V$, $\langle T^*(x), y\rangle = \langle x, T(y)\rangle$.

June 17

4.3.7 Remark. Let $A \in M_n(F)$ and let σ be the standard (orthonormal) basis for F^n. Then $A = [L_A]_\sigma$, so

$$A^* = [L_A]_\sigma^* = [L_A^*]_\sigma = [L_{A^*}]_\sigma,$$

and thus $\langle Ax, y\rangle = \langle x, A^*y\rangle$.

4.1 Example. Let $T\colon P_2(\mathbb{R}) \to P_2(\mathbb{R})$ be given by $T(f(x)) = f'(x)$. Take $\sigma\{x^2, x, 1\}$ to be an orthonormal basis for $P_2(\mathbb{R})$ under the dot product. Then

$$[T]_\sigma = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

so

$$[T^*]_\sigma = [T]_\sigma^* = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Hence $T^*(x^2) = 0$, $T^*(x) = 2x^2$, and $T^*(1) = x$, so $T^*(ax^2 + bx + c) = 2bx^2 + c$.

4.3.8 Proposition. *Let V be a finite-dimensional inner product space. Let $T, U\colon V \to V$ be linear, and let $\alpha \in F$. Then*

(1) $(T + U)^* = T^* + U^*$

(2) $(\alpha T)^* = \overline{\alpha} T^*$

(3) $(T \circ U)^* = U^* \circ T^*$

(4) $(T^*)^* = T$

(5) $I^* = I$.

Proof. Let $x, y \in V$. Then

(1) $\langle (T + U)(x), y\rangle = \langle T(x), y\rangle + \langle U(x), y\rangle = \langle x, T^*(y)\rangle + \langle x, U^*(y)\rangle = \langle x, (T^* + U^*)(y)$

(2) $\langle (\alpha T)(x), y\rangle = \alpha\langle T(x), y\rangle = \alpha\langle x, T^*(y)\rangle = \langle x, (\overline{\alpha}T)(y)\rangle$

(3) $\langle (T \circ U)(x), y\rangle = \langle U(x), T^*(y)\rangle = \langle x, (U^* \circ T^*)(y)\rangle$

(4) $\langle (T^*)(x), y\rangle = \overline{\langle y, (T^*)(x)\rangle} = \overline{\langle T(y), x\rangle} = \langle x, T(y)\rangle$

(5) $\langle I(x), y\rangle = \langle x, y\rangle = \langle x, I(y)\rangle$,

and in each case the result follows by uniqueness of the adjoint. ∎

4.4 Least Squares Approximation

4.4.1 Definition. Suppose we have real data points $y_1, y_2, \ldots, y_m$ observed at times $t_1, t_2, \ldots, t_m$, and we plot each (t_i, y_i) in $\mathbb{R}^2$. Our goal is to find a line that best fits this data; i.e., to find the line so that the (vertical) distances between the points and said line is minimal. In fact, we will seek to minimize the squares of these vertical distances. Hence this *line of best fit* will also be called the *least squares line.*

4.2 Remark. We wish to find the line $y = cx + d$ that minimizes the error term minimize the error term $E = \sum_{i=1}^{n}(ct_i + d - y_i)^2$. Accordingly, we set

$$A = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix} \qquad x = \begin{bmatrix} c \\ d \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

so that $E = \|Ax - y\|^2$. Thus we must find x_0 so that $\|Ax_0 - y\|$ is minimal.

4.4.2 Remark. We extend the definition of the adjoint to include any $A \in M_{m\times n}(F)$ by defining A^* to be the conjugate transpose of A, as in the $n \times n$ case.

4.4.3 Lemma. *Let $A \in M_{m\times n}(F)$, $x \in F^n$, and $y \in F^m$. Then $\langle Ax, y\rangle = \langle x, A^*y\rangle$.*

Proof. Note that $\langle Ax, y\rangle = y^*Ax = (A^*y)^*x = \langle x, A^*y\rangle$.

4.4.4 Lemma. *Let $A \in M_{m\times n}(F)$. Then* $\operatorname{rank}(A) = \operatorname{rank}(A^*A)$. *In particular, if* $\operatorname{rank} A = n$, *then A^*A is invertible.*

Proof. We will show a stronger result, namely that $\operatorname{Null}(A) = \operatorname{Null}(A^*A)$. Clearly $\operatorname{Null}(A) \subseteq \operatorname{Null}(A^*A)$. Let $x \in \operatorname{Null}(A^*A)$. Then

$$\|Ax\|^2 = \langle Ax, Ax\rangle = \langle x, A^*Ax\rangle = \langle x, 0\rangle = 0,$$

so $Ax = 0$, and thus $\operatorname{Null}(A^*A) \subseteq \operatorname{Null}(A)$. *A fortiori*, this completes the proof.

4.3 Remark. Recall that we want to minimize $\|Ax - y\|$, where $A \in M_{m\times n}(F)$, $x \in F^n$, $y \in F^m$. Let $W = \operatorname{Range}(A)$. Let $y_0 = \operatorname{proj}_W(y) \in W$. Say $y_0 = Ax_0$ for some $x_0 \in F^n$. Then $\|Ax_0 - y\|$ is minimal. *June 19*
Now,

$$\begin{aligned} y - y_0 \in W^\perp &\implies y - Ax_0 \in W^\perp \\ &\implies \langle Ax, y - Ax_0\rangle = 0\ \forall x \in F^n \\ &\implies \langle x, A^*(y - Ax_0)\rangle = 0\ \forall x \in F^n \\ &\implies A^*(y - Ax_0) = 0 \\ &\implies A^*y = A^*Ax_0. \end{aligned}$$

If $\operatorname{rank} A = n$ (which it always is for our real world applications), $x_0 = (A^*A)^{-1}A^*y$.

4.4 Example. In the last four Spring terms at Waterloo, the MATH 245 final exam averages have been 75, 82, 60, and 70.[1] To find the line of best fit for this data, we plot the points $(1, 75)$, $(2, 82)$, $(3, 60)$, and $(4, 70)$ and use the matrices

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \qquad y = \begin{bmatrix} 75 \\ 82 \\ 60 \\ 70 \end{bmatrix}.$$

This gives us

$$A^* = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \qquad A^*A = \begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix} \qquad (A^*A)^{-1} = \frac{1}{20}\begin{bmatrix} 4 & -10 \\ -10 & 30 \end{bmatrix} = \begin{bmatrix} 1/5 & -1/2 \\ -1/2 & 3/2 \end{bmatrix}.$$

[1] Data fabricated for the purpose of this example.

Therefore

$$x_0 = (A^*A)^{-1}A^*y = \begin{bmatrix} 1/5 & -1/2 \\ -1/2 & 3/2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 75 \\ 82 \\ 60 \\ 70 \end{bmatrix} = \begin{bmatrix} 1/5 & -1/2 \\ -1/2 & 3/2 \end{bmatrix} \begin{bmatrix} 699 \\ 287 \end{bmatrix} = \begin{bmatrix} -3.7 \\ 81 \end{bmatrix},$$

which gives us the line of best fit $y = -3.7x + 81$.

4.5 Remark. We similarly can find the polynomial of best fit, $y = a_n x^n + a_{n-1}x^{n-1} + a_1 x + a_0$, by using the matrices

$$A = \begin{bmatrix} t_1^n & t_1^{n-1} & \cdots & t_1 & 1 \\ t_2^n & t_2^{n-1} & \cdots & t_2 & 1 \\ \vdots & \vdots & \cdots & \vdots & \\ t_m^n & t_m^{n-1} & \cdots & t_m & 1 \end{bmatrix} \qquad x = \begin{bmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_0 \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

4.5 Normal, Hermitian, and Unitary Operators

4.5.1 Remark. Note that for $A \in M_n(F)$, if the columns of A form an orthonormal basis for F^n, then $A^*A = I$, and thus $A^{-1} = A^*$.

4.5.2 Lemma. *Let V be a finite-dimensional inner product space. Let $T\colon V \to V$ be linear. If T has an eigenvector, then T^* has an eigenvector.*

Proof. Suppose there exists $0 \neq v \in V$ such that $T(v) = \lambda v$ for some $\lambda \in F$. Then

$$\begin{aligned} (T - \lambda I)(v) = 0 &\implies \langle (T-\lambda I)(v), x\rangle = 0 \ \forall x \in V \\ &\implies \langle v, (T-\lambda I)^*(x)\rangle = 0 \ \forall x \in V \\ &\implies \langle v, (T^* - \overline{\lambda} I)(x)\rangle = 0 \ \forall x \in V. \end{aligned}$$

Hence $0 \neq v \in \operatorname{Range}(T^* - \overline{\lambda} I)^\perp$, so $\operatorname{Range}(T^* - \overline{\lambda} I) \neq V$. In particular, $\operatorname{Null}(T^* - \overline{\lambda} I) \neq \{0\}$, so T^* has a $\overline{\lambda}$ eigenvector. ∎

4.5.3 Theorem (Schur)**.** *Let V be a finite-dimensional inner product space. Let $T\colon V \to V$ be linear such that the characteristic polynomial of T splits over F. Then there exists an orthonormal basis for V such that $[T]_\beta$ is upper triangular.*

June 21

Proof. By induction on $n = \dim V$. If $n = 1$, we're clearly done. Inductively, assume the result for all inner product spaces with dimension less than n. Suppose $\dim V = n$. Since the characteristic polynomial of T splits, T must have an eigenvector. By Lemma 4.5.2, so does T^*. Let an eigenvector for T^* be $0 \neq v$; then $T^*(v) = \lambda v$ for some $\lambda \in F$. Without loss of generality, we may assume that $\|v\| = 1$. Take $W = \operatorname{Span} v$. Then $V = W \oplus W^\perp$.

We claim that $W^\perp$ is T-invariant. Accordingly, let $y \in W^\perp$. Then $\langle T(y), v\rangle = \langle y, T^*(v)\rangle = \langle y, \lambda v\rangle = \overline{\lambda}\langle y, v\rangle = 0$, because $y \in W^\perp$. Since $W = \operatorname{Span} v$, this shows that $T(y) \in W^\perp$; hence, $W^\perp$ is T-invariant.

Now, $\dim W^\perp = n - 1$, and the characteristic polynomial of T_W splits, since it divides the characteristic polynomial of T. Therefore there exists an orthonormal basis γ or $W^\perp$ such that $[T_{W^\perp}]_\gamma$ is upper triangular. Then $\beta = \gamma \cup \{v\}$ is an orthonormal basis for V, and

$$[T]_\beta = \left[\begin{array}{ccc|c} & & & * \\ & [T_{W^\perp}]_\gamma & & * \\ & & & \vdots \\ \hline 0 & \cdots & 0 & * \end{array}\right],$$

which is upper triangular. This completes the proof. ∎

4.5.4 Corollary. [2] *Let $A \in M_n(F)$ such that the characteristic polynomial of A splits. Then there exist $U, B \in M_n(F)$ such that $U^{-1} = U^*$, B is upper triangular, and $A = UBU^*$.*

Proof. By Schur's theorem, choose an ordered basis β such that $[L_A]_\beta$ is upper triangular, and set $B = [L_A]_\beta$. Then take $U = [I]_\beta^\sigma$. The columns of U are orthonormal, so $[I]_\sigma^\beta = U^{-1} = U^*$. Then

$$A = [L_A]_\sigma = [I]_\beta^\sigma [L_A]_\beta [I]_\sigma^\beta = UBU^*,$$

as required. ∎

4.5.5 Definition. Let V be a finite-dimensional inner product space. Let $T : V \to V$ be linear. We say that T is *normal* if $TT^* = T^*T$. Similarly, we say that $A \in M_n(F)$ is normal if $AA^* = A^*A$.

4.5.6 Proposition. *Let V be a finite-dimensional inner product space. Let $T : V \to V$ be normal.*

(1) *For all $x \in V$, $\|T(x)\| = \|T^*(x)\|$.*

(2) *Every λ-eigenvector of T is a $\overline{\lambda}$-eigenvector of T^*.*

(3) *If x is a λ-eigenvector of T and y is a μ-eigenvector of T, where $\lambda \neq \mu$, then x and y are orthogonal.*

Proof.

(1) $\|T(x)\|^2 = \langle T(x), T(x)\rangle = \langle x, T^*T(x)\rangle = \langle x, TT^*(x)\rangle = \langle T*(x), T^*(x)\rangle = \|T^*(x)\|^2$.

(2) Suppose $T(v) = \lambda v$ for some $0 \neq v \in V$ and $\lambda \in F$. consider $U = T - \lambda I$. Then $UU^* = U^*U$, so $\|U^*(v)\| = \|U(v)\| = 0$, which means that $U^*(v) = 0$. Therefore $T^*(v) = \overline{\lambda} v$.

(3) Suppose $T(x) = \lambda x$ and $T(y) = \mu y$ for some $0 \neq x, y \in V$, $\lambda, \mu \in F$ with $\lambda \neq \mu$. Then

$$\lambda\langle x, y\rangle = \langle T(x), y\rangle = \langle x, T^*(y)\rangle = \langle x, \overline{\mu} y\rangle = \mu\langle x, y\rangle.$$

Since $\lambda \neq \mu$, we must have $\langle x, y\rangle = 0$, so x and y are orthogonal.

This completes the proof. ∎

4.5.7 Theorem. *Let V be a finite-dimensional inner product space over $\mathbb{C}$. Let $T : V \to V$ be linear. Then T is normal if and only if there exists an orthonormal basis β for V composed of eigenvectors of T.*

June 24

Proof. ($\Rightarrow$) Assume T is normal. By Schur's Theorem, there exists an orthonormal basis $\beta = \{v_1, v_2, \dots, v_n\}$ such that $[T]_\beta$ is upper triangular. Say $A = [T]_\beta = (a_{ij})$. Now $T(v_1) = a_{11}v_1$, so v_1 is an eigenvector of T. Inductvely, suppose $v_1, v_2, \dots, v_{k-1}$ are eigenvectors of T for some $k \geq 2$. Say $T(v_i) = \lambda_i b_k$ for $i \in \{1, 2, \dots, k-1\}$. We claim that v_i is an eigenvector of T.

Note that since A is upper triangular,

$$[T(v_k)]_\beta = [T]_\beta [v_k]_\beta = A[v_k]_\beta = (a_{1k}, a_{2k}, \dots, a_{kk}, 0, \dots, 0),$$

and therefore $T(v_k) = a_{1k}v_1 + a_{2k}v_2 + \cdots + a_{kk}v_k$. By Proposition 4.2.8,

$$a_{ik} = \langle T(v_k), v_i\rangle = \langle v_k, T^*(v_i)\rangle = \langle v_k, \overline{\lambda_i} v_i\rangle = \lambda\langle v_k, v_i\rangle = 0$$

for $1 \leq i \leq k$, so in fact $T(v_k) = a_{kk}v_k$. By induction, β is an orthonormal basis for V composed of eigenvectors of T.

($\Leftarrow$) Assume there exists an orthonormal basis β for V composed of eigenvectors of T. Then $[T]_\beta$ is diagonal. Since β is orthonormal, $[T^*]_\beta = [T]_\beta^*$, which must be diagonal. Therefore

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

Hence $T^*T = T T*$, so T is normal. This completes the proof. ∎

[2] Corollary 4.5.4 was presented on June 17, before the proof of Schur's Theorem.

4.5.8 Corollary. [3] *Let $A \in M_n(\mathbb{C})$. Then A is normal if and only if there exists $U, D \in M_n(\mathbb{C})$ such that $U^{-1} = U^*$, D is diagonal, and $A = UDU^*$.*

4.5.9 Example. A word of warning:

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \in M_2(\mathbb{R})$$

satisfies $AA^* = A^*A = I$, but its characteristic polynomial is $x^2 + 1$, so it is not diagonalizable.

4.5.10 Definition. Let T be a linear operator on a finite-dimensional inner product space V. We say that T is *Hermitian* if $T = T^*$. Similarly, we say that $A \in M_n(F)$ is *Hermitian* if $A = A^*$.

4.5.11 Proposition. *Let V be a finite-dimensional inner product space. Let $T\colon V \to V$ be a Hermitian operator. Then*

(1) *every eigenvalue of T is real;*

(2) *the characteristic polynomial of T splits over F.*

Proof.

(1) Since T is Hermitian, T is normal. Let $\lambda \in F$ be an eigenvalue of T with eigenvector $0 \neq x$. Then $\lambda x = T(x) = T^*(x) = \overline{\lambda} x$. Since $x \neq 0$, $\lambda = \overline{\lambda}$, so $\lambda \in \mathbb{R}$.

(2) We know that the characteristic polynomial of T must split over $\mathbb{C}$. Since every eigenvalue of T is real, the characteristic polynomial of T has no complex roots, so it must also split over $\mathbb{R}$.

This completes the proof.

4.5.12 Theorem. *Let V be a finite-dimensional inner product space over $\mathbb{R}$. Let $T\colon V \to V$ be linear. Then T is Hermitian if and only if there exists an orthonormal basis β for V composed of eigenvectors of T.*

Proof. ($\Rightarrow$) Assume T is Hermitian. By Proposition 4.5.11, its characteristic polynomial splits over $\mathbb{R}$. By Schur's Theorem, there exists an orthonormal basis β such that $[T]_\beta$ is upper triangular. Furthermore $[T]_\beta^* = [T^*]_\beta = [T]_\beta$, so $[T]_\beta$ is symmetric and hence diagonal.

($\Leftarrow$) Assume there exists an orthonormal basis β for V composed of eigenvectors of T. Then $[T^*]_\beta = [T]_\beta^*$. Since $[T]_\beta$ is diagonal, $[T]_\beta^* = [T]_\beta$, so in fact $[T^*]_\beta = [T]_\beta$. Therefore $T = T^*$, so we're done.

4.5.13 Corollary. *Let $A \in M_n(\mathbb{R})$. Then A is Hermitian if and only if there exist $U, D \in M_n(\mathbb{R})$ such that $U^T = U^{-1}$, D is diagonal, and $A = UDU^T$.*

June 26

4.5.14 Example. The matrix

$$A = \begin{bmatrix} i & i \\ i & 1 \end{bmatrix} \in M_2(\mathbb{C})$$

is symmetric but not normal, since

$$A^*A = \begin{bmatrix} -i & -i \\ -i & 1 \end{bmatrix} \begin{bmatrix} i & i \\ i & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1-i \\ 1+i & 2 \end{bmatrix} \neq \begin{bmatrix} 2 & 1+i \\ 1-i & 2 \end{bmatrix} = \begin{bmatrix} i & i \\ i & 1 \end{bmatrix} \begin{bmatrix} -i & -i \\ -i & 1 \end{bmatrix} = AA^*.$$

4.5.15 Definition. Let V be a finite-dimensional inner product space. Let $T\colon V \to V$ be linear. If $T^{-1} = T^*$, then we say T is

(1) orthogonal if $F = \mathbb{R}$

(2) unitary if $F = \mathbb{C}$.

Similarly, we say that $A \in M_n(\mathbb{R})$ ($M_n(\mathbb{C})$) is orthogonal (unitary) if $A^{-1} = A^*$.

4.5.16 Remark. $A \in M_n(F)$ is unitary/orthogonal if and only if L_A is unitary/orthogonal.

[3] Corollary 4.5.8 and Example 4.5.9 were presented on June 19, before the proof of Theorem 4.5.7.

4.5.17 Proposition. *Let V be a finite-dimensional inner product space. Let $T\colon V \to V$ be linear. The following are equivalent:*

(1) *T is unitary/orthogonal.*

(2) *$\langle T(x), T(y)\rangle = \langle x, y\rangle$ for all $x, y \in V$.*

(3) *If β is an orthonormal basis for V, then $T(\beta)$ is an orthonormal basis for V.*

(4) *There exists an orthonormal basis β for V such that $T(\beta)$ is an orthonormal basis for V.*

(5) *For all $x \in V$, $\|T(x)\| = \|x\|$.*

Proof. (1) $\Rightarrow$ (2) Assume (1). Then for all $x, y \in V$,

$$\langle T(x), T(y)\rangle = \langle x, T^*T(y)\rangle = \langle x, T^{-1}T(y)\rangle = \langle x, y\rangle.$$

(2) $\Rightarrow$ (3) Assume (2). Let $\beta = \{v_1, v_2, \dots, v_n\}$ be an orthonormal basis for V. We first show that T is injective. Let $x \in \mathrm{Null}(T)$. Then $0 = \|T(x)\| = \langle T(x), T(x)\rangle = \langle x, x\rangle$, so $x = 0$. Therefore $\mathrm{Null}(T) = \{0\}$, so T is injective, and in particular $T(\beta)$ is a basis for B. Finally, for any $1 \leq i, j \leq n$ with $i \neq j$, $\langle T(v_i), T(v_j)\rangle = \langle v_i, v_j\rangle = 0$ and $\|T(v_i)\|^2 = \langle T(v_i), T(v_i)\rangle = \langle v_i, v_i\rangle = 1$, so $T(\beta)$ is orthonormal.

(3) $\Rightarrow$ (4) Trivial.

(4) $\Rightarrow$ (5) Assume (4). Let β be an orthonormal basis for V such that $T(\beta)$ is an orthonormal basis for V. Say $\beta = \{v_1, v_2, \dots, v_n\}$. Let $x \in V$. Say $x = a_1v_1 + a_2v_2 + \cdots + a_nv_n$, where $a_1, a_2, \dots, a_n \in F$. Then

$$\|x\|^2 = \langle a_1v_1+a_2v_2+\cdots+a_nv_n, a_1v_1+a_2v_2+\cdots+a_nv_n\rangle = a_1\overline{a_1}+a_2\overline{a_2}+\cdots+a_n\overline{a_n} = \|a_1\|^2+\|a_2\|^2+\cdots+\|a_n\|^2$$

and similarly

$$\|T(x)\|^2 = \langle a_1T(v_1)+a_2T(v_2)+\cdots+a_nT(v_n), a_1T(v_1)+a_2T(v_2)+\cdots+a_nT(v_n)\rangle = \|a_1\|^2+\|a_2\|^2+\cdots+\|a_n\|^2,$$

so $\|x\| = \|T(x)\|$.

(5) $\Rightarrow$ (1) Assume $\|T(x)\| = \|x\|$ for all $x \in V$. Then for all $x \in V$, $\langle T(x), T(x)\rangle = \langle x, x\rangle$, so $\langle x, T^*T(x)\rangle = \langle x, x\rangle$, which implies that $\langle x, (T^*T - I)(x)\rangle = 0$. Setting $U = T^*T - I$, we note that $U = U^*$. Therefore there exists an orthonormal basis for V composed of eigenvectors of U. Accordingly, let $0 \neq v \in V$ such that $U(v) = \lambda v$ for some $\lambda \in F$. Then $0 = \langle v, U(v)\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle$, and since $v \neq 0$ we must have $\lambda = 0$. Therefore all the eigenvalues of U are 0, and since U is diagonalizable, $U = 0$. Therefore $T^*T = I$.

This completes the proof. ∎

4.5.18 Definition. Let $A, B \in M_n(F)$. We say that A and B are *orthogonally/unitarily equivalent* if there exists an orthogonal/unitary matrix U such that $A = UBU^*$. If B is diagonal, we say that A is *orthogonally/unitarily diagonalizable*.

4.5.19 Corollary. *Let V be a finite-dimensional inner product space. Let $T\colon V \to V$ be orthogonal/unitary. Then every eigenvalue of T has absolute value 1.*

Proof. If $T(x) = \lambda x$ for some $0 \neq x \in V$, $\lambda \in F$, then $|\lambda|\|x\| = \|\lambda x\| = \|T(x)\| = \|x\|$, so $|\lambda| = 1$. ∎

June 28

4.5.20 Corollary. *Let V be a finite-dimensional inner product space over $\mathbb{R}$. Let $T\colon V \to V$ be linear. Then T is orthogonal and Hermitian if and only if there exists an orthonormal basis for V composed of ± 1 eigenvectors of T.*

4.5.21 Corollary. *Let V be a finite-dimensional inner product space over $\mathbb{C}$. Let $T\colon V \to V$ be linear. Then T is unitary if and only if there exists an orthonormal basis for V composed of eigenvectors of T where each eigenvector corresponds to an eigenvalue of modulus 1.*

4.5.22 Example. Let
$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$
Then A is normal, but not Hermitian, so A is unitarily diagonalizable but not orthogonally diagonalizable. The characteristic polynomial of A is $f(x) = x^2 + 1 = (x+i)(x-i)$.

We see by inspection that $(i, 1)$ and $(1, i)$ are eigenvectors for A corresponding to eigenvalues $\lambda_1 = i$ and $\lambda_2 = -i$ respectively. Therefore $E_{\lambda_1} = \operatorname{Span}\{(i, 1)\}$ and $E_{\lambda_2} = \operatorname{Span}\{(1, i)\}$. Since A is normal, these eigenvectors must be orthogonal, so $\beta = \left(\frac{1}{\sqrt{2}}(i, 1), \frac{1}{\sqrt{2}}(1, i)\right)$ is an orthonormal basis for $\mathbb{C}^2$. Therefore $A = UDU^*$, where
$$U = \begin{bmatrix} i/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & i/\sqrt{2} \end{bmatrix} \qquad D = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

4.5.23 Example. Let
$$A = \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix}.$$
Note that A is symmetric and thus orthogonally diagonalizable. The characteristic polynomial of A is
$$\begin{aligned} f(x) &= \begin{vmatrix} -x & 2 & x \\ 2 & -x & 2 \\ 2 & 2 & -x \end{vmatrix} \\ &= -x \begin{vmatrix} -x & 2 \\ 2 & -x \end{vmatrix} - 2 \begin{vmatrix} 2 & 2 \\ 2 & -x \end{vmatrix} + 2 \begin{vmatrix} 2 & 2 \\ -x & 2 \end{vmatrix} \\ &= -x(x^2-4) - 2(-2x-4) + 2(4+2x) \\ &= -x^3 + 4x + 4x + 8 + 8 + 4x \\ &= -x^3 + 12x + 16 \\ &= (x-4)(-x^2-4x-4) \\ &= -(x-4)(x+2)^2. \end{aligned}$$
Setting $\lambda_1 = 4$, we note that $(1, 1, 1)$ is a λ_1 eigenvector, so $E_{\lambda_1} = \operatorname{Span}\{(1, 1, 1)\}$. Setting $\lambda_2 = -2$, we see that $(1, -1, 0)$ and $(-1, 0, 1)$ are λ_2 eigenvectors, so $E_{\lambda_2} = \operatorname{Span}\{(1, -1, 0), (-1, 0, 1)\}$. Applying the Gram-Schmidt procedure produces orthonormal bases $\left\{\frac{1}{\sqrt{3}}(1, 1, 1)\right\}$ and $\left\{\frac{1}{\sqrt{2}}(1, -1, 0), \frac{1}{\sqrt{6}}(-1, -1, 2)\right\}$ for E_{λ_1} and E_{λ_2} respectively. Since A is normal, an orthonormal basis for V is given by
$$\left\{\frac{1}{\sqrt{3}}(1, 1, 1), \frac{1}{\sqrt{2}}(1, -1, 0), \frac{1}{\sqrt{6}}(-1, -1, 2)\right\}$$
. Therefore $A = UDU^T$, where
$$U = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & -1\sqrt{2} & -1\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{bmatrix} \qquad D = \begin{bmatrix} 4 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

4.6 Rigid Motions

4.6.1 Definition. Let V be a finite-dimensional inner product space over $\mathbb{R}$. We say that $f \colon V \to V$ is a *rigid motion* if $\|f(x) - f(y)\| = \|x - y\|$ for all $x, y \in V$.

4.6.2 Example. In $\mathbb{R}^2$, rotation, translation, and reflection are all rigid motions. As we will show, these are the only rigid motions in $\mathbb{R}^2$.

4.6.3 Definition. Let V be a vector space. A *translation* is a function $f\colon V \to V$ given by $f(x) = x + v$ for some fixed $v \in V$.

4.6.4 Proposition. *Let V be a finite-dimensional inner product space over $\mathbb{R}$. Let $f\colon V \to V$ be a rigid motion. Then there exists a unique orthogonal operator $T\colon V \to V$ and a unique translation $g\colon V \to V$ such that $f = g \circ T$.*

Proof. Define $T\colon V \to V$ by $T(x) = f(x) - f(0)$. We claim that T is linear and orthogonal. First note the following:

(1) For all $x, y \in V$, $\|T(x) - T(y)\| = \|f(x) - f(0) - f(y) + f(0)\| = \|f(x) - f(y)\| = \|x - y\|$.

(2) For all $x \in V$, $\|T(x)\|^2 = \|f(x) - f(0)\|^2 = \|x - 0\|^2 = \|x\|^2$, so $\|T(x)\| = \|x\|$.

(3) For $x, y \in V$, $\|T(x) - T(y)\|^2 = \|T(x)\|^2 + \|T(y)\|^2 - 2\langle T(x), T(y)\rangle = \|x\|^2 + \|y\|^2 - 2\langle T(x), T(y)\rangle$ and $\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y\rangle$. By (1), $\|T(x) - T(y)\| = \|x - y\|$, so $\langle x, y\rangle = \langle T(x), T(y)\rangle$.

Let $x, y \in V$, $\alpha \in \mathbb{R}$. Then

$$\begin{aligned}
\|T(x + \alpha y) - T(x) - \alpha T(y)\|^2 &= \|T(x + \alpha y) - T(x)\|^2 + \|\alpha T(y)\|^2 - 2\langle T(x + \alpha y) - T(x), \alpha T(y)\rangle \\
&= \|x - \alpha y - x\|^2 + \|\alpha y\|^2 - 2\langle x + \alpha y - x, \alpha y\rangle \\
&= 2\alpha^2\|y\|^2 - 2\alpha^2\langle y, y\rangle \\
&= 0,
\end{aligned}$$

so $T(x + \alpha y) = T(x) + \alpha T(y)$. Therefore T is linear. By (2), T is orthogonal.

It remains to show uniqueness. Suppose T and U are orthogonal and $a, b \in V$ such that $f(x) = T(x) + a = U(x) + b$. Then $f(a) = T(0) = a = U(0) = b$, so $a = b$, and therefore $U = T$. This completes the proof. ∎

4.6.5 Example. Consider $T_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ corresponding to counter-clockwise rotation by θ. Then for all $x \in \mathbb{R}^2$, $\|T_\theta(x)\| = \|x\|$, so T_θ is orthogonal. Moreover, $T_\theta(e_1) = (\cos\theta, \sin\theta)$ and $T_\theta(e_2) = (-\sin\theta, \cos\theta)$. Therefore

$$[T_\theta]_\sigma = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

4.6.6 Example. Consider $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ corresponding to reflection over the line $y = mx$. Let α be the (positive) angle between the x-axis and $y = mx$. Take $v_1 = (\cos\alpha, \sin\alpha)$ and $v_2 = (-\sin\alpha, \cos\alpha)$; note that $\|v_1\| = 1 = \|v_2\|$ and $v_1 \cdot v_2 = 0$, so $\beta = \{v_1, v_2\}$ is an orthonormal basis for bbR^2. Therefore

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

This gives us $[T]_\sigma = U[T]_\beta U^T$, where

$$U = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}.$$

Therefore

$$\begin{aligned}
[T]_\sigma &= \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \\
&= \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{bmatrix} \\
&= \begin{bmatrix} \cos^2\alpha - \sin^2\alpha & 2\cos\alpha\sin\alpha \\ 2\cos\alpha\sin\alpha & \sin^2\alpha - \cos^2\alpha \end{bmatrix} \\
&= \begin{bmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{bmatrix}.
\end{aligned}$$

4.6.7 Proposition. *Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be an orthogonal operator. Then T is a rotation or a reflection over a line through the origin. In particular, if $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is a rigid motion, then f is either a rotation or a reflection followed by a translation.*

July 3

Proof. Let $\sigma = \{e_1, e_2\}$ be the standard basis for $\mathbb{R}^2$. Since T is orthogonal, $T(\sigma) = \{T(e_1), T(e_2)\}$ is an orthonormal basis for $\mathbb{R}^2$. Then $\|T(e_1)\| = 1$, so $T(e_1) = (\cos\theta, \sin\theta)$ for some $\theta \in [0, 2\pi)$. Moreover, $\langle T(e_1), T(e_2)\rangle = 0$ and $\|T(e_2)\| = 1$. It follows that either $T(e_2) = \left(\cos\left(\theta + \frac{\pi}{2}\right), \sin\left(\theta + \frac{\pi}{2}\right)\right) = (-\sin\theta, \cos\theta)$ or $T(e_2) = \left(\cos\left(\theta - \frac{\pi}{2}\right), \sin\left(\theta - \frac{\pi}{2}\right)\right) = (\sin\theta, -\cos\theta)$. Therefore $[T]_\sigma$ is one of

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad\qquad \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

and therefore T is either a rotation or a reflection, as required. ∎

4.7 Spectral Decomposition

4.7.1 Definition. Let V be a finite-dimensional vector space. Let $W_1, W_2 \leq V$ such that $V = W_1 \oplus W_2$. Recall that for every $v \in V$ there exist unique $x_v \in W_1$, $y_v \in W_2$ such that $v = x_v + y_v$. The linear map $T\colon V \to V$ given by $T(v) = x_v$ is called the *projection on* W_1 *along* W_2. If $W_2 = W_1^\perp$, then $T(v) = \operatorname{proj}_W(v)$, and we say that T is an *orthogonal projection.*

4.6 Remark. If T is as in the above definition, $\operatorname{Range}(T) = W_1$ and $\operatorname{Null}(T) = W_2$.

4.7.2 Proposition. *Let V be a finite dimensional vector space. Then a linear operator $T\colon V \to V$ is a projection if and only if $T = T^2$.*

Proof. ($\Rightarrow$) Assume T is a projection. Thus T is the projection on $\operatorname{Range}(T)$ along $\operatorname{Null}(T)$. Let $v \in V$. Then $v = T(x) + z$, for some $T(x) \in \operatorname{Range}(T)$, $z \in \operatorname{Null}(T)$. Then $T(v) = T^2(x) + T(z) = T^2(x) = T(x)$, so $T^2(v) = T^2(x) = T(x) = T(v)$. Therefore $T = T^2$.

($\Leftarrow$) Assume $T = T^2$. We claim that $V = \operatorname{Range}(T) \oplus \operatorname{Null}(T)$. Indeed, if $x \in \operatorname{Range}(T) \cap \operatorname{Null}(T)$, then $T(x) = 0$ and there is some $y \in V$ such that $x = T(y)$. But then $0 = T(x) = T^2(y) = T(y) = x$, so $\operatorname{Range}(T) \cap \operatorname{Null}(T) = \{0\}$. But $\dim(\operatorname{Range}(T) + \operatorname{Null}(T)) = \dim V$ by the Rank-Nullity Theorem, so $V = \operatorname{Range}(T) \oplus \operatorname{Null}(T)$. Moveover, if $v = T(x) + z$, where $T(x) \in \operatorname{Range}(T)$ and $z \in Null(T)$, then $T(v) = T^2(x) = T(x)$, so T is the projection on $\operatorname{Range}(T)$ along $\operatorname{Null}(T)$. ∎

4.7.3 Proposition. *Let V be a finite-dimensional inner product space. Let $T\colon V \to V$ be a linear operator. Then T is an orthogonal projection if and only if $T = T^2 = T^*$.*

Proof. ($\Rightarrow$) Assume T is an orthogonal projection. By Proposition 4.7.2, we know that $T = T^2$, so it suffices to show that T is Hermitian. Let $x, y \in V$. Then $x = T(v_1) + z_1$ and $y = T(v_2) + z_2$ for some $T(v), T(v_2) \in \operatorname{Range}(T)$, $z_1, z_2 \in \operatorname{Null}(T)$. Then

$$\langle T(x), y\rangle = \langle T^2(v_1), T(v_2) + z_2\rangle = \langle T(v_1), T(v_2) + z_2\rangle = \langle T(v_1), T(v_2)\rangle + \langle T(v_1), z_2\rangle.$$

Since T is an orthogonal projection, $\langle T(v_1), z_2\rangle = 0$. Therefore $\langle T(x), y\rangle = \langle T(v_1), T(v_2)\rangle$. Similarly,

$$\langle x, T(y)\rangle = \langle T(v_1) + z_1, T^2(v_2)\rangle = \langle T(v_1) + z_1, T(v_2)\rangle = \langle T(v_1), T(v_2)\rangle + \langle z_1, T(v_2)\rangle = \langle T(v_1), T(v_1)\rangle.$$

Therefore $\langle x, T(y)\rangle = \langle y, T(x)\rangle$, and since x and y were arbitrary, $T = T^*$.

($\Leftarrow$) Assume $T = T^2 = T^*$. By Proposition 4.7.2, we know that T is a projection, so it suffices to show that T is an orthogonal projection, i.e., that $\operatorname{Null}(T) = \operatorname{Range}(T)^\perp$. Accordingly, let $T(x) \in \operatorname{Range}(T)$, $y \in \operatorname{Null}(T)$. Then

$$\langle T(x), y\rangle = \langle x, T^*(y)\rangle = \langle x, T(y)\rangle = \langle x, 0\rangle = 0,$$

so $\operatorname{Null}(T) \subseteq \operatorname{Range}(T)^\perp$. But $\dim \operatorname{Null}(T) = \dim V - \operatorname{Range}(T) = \operatorname{Range}(T)^\perp$, so $\operatorname{Null}(T) = \operatorname{Range}(T)^\perp$. Therefore T is an orthogonal projection. ∎

4.7.4 Definition. For $i, j \in \mathbb{Z}$, we define

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

4.7.5 Theorem (Spectral Theorem)**.** *Let V be a finite-dimensional inner product space. Let $T: V \to V$ be linear. Let the distinct eigenvalues of T be $\lambda_1, \lambda_2, \ldots, \lambda_k$. If $F = \mathbb{C}$, assume T is normal; if $F = \mathbb{R}$, assume T is Hermitian. For $1 \leq i \leq k$, let $W_i = E_{\lambda_i}$ and $T_i(x) = \operatorname{proj}_{W_i}(x)$. Then*

(1) $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$

(2) $W_i^{\perp} = W_1 \oplus W_2 \oplus \cdots \oplus W_{i-1} \oplus W_{i+1} \oplus \cdots \oplus W_k =: W_i'$

(3) $T_i \circ T_j = \delta_{ij} T_i$

(4) $I = T_1 + T_2 + \cdots + T_k$

(5) $T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$.

July 5

Proof. Fix $1 \leq i, j \leq k$.

(1) Since T is normal/Hermitian, its distinct eigenspaces intersect trivially; therefore $W_1 \oplus W_2 \oplus \cdots \oplus W_k \leq V$. Also, T is diagonalizable, so $\dim W_1 + \dim W_2 + \cdots + \dim W_k = \dim V$. It follows that $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$.

(2) Let $x = x_1 + x_2 + \cdots + x_{i-1} + x_{i+1} + \cdots + x_k \in W_i'$, $y \in W_i$. Then

$$\langle x, y\rangle = \langle x_1, y\rangle + \langle x_2, y\rangle + \cdots + \langle x_{i-1}, y\rangle + \langle x_{i+1}, y\rangle + \cdots + \langle x_k, y\rangle = 0 + 0 + \cdots + 0 + 0 + \cdots + 0 = 0,$$

since T is normal. Therefore $W_i' \subseteq W_i^{\perp}$, and since $\dim W_i' = \dim V - \dim W_i = \dim W_i^{\perp}$, $W_i' = W_i^{\perp}$.

(3) First note that $T_i \circ T_i = T_i^2 = T_i$, since T_i is a projection. If $i \neq j$, then $T_i \circ T_j = 0$, since T is normal/Hermitian and therefore E_{λ_i} and E_{λ_j} intersect trivially.

(4) Let $x = x_1 + x_2 + \cdots + x_k$, where each $x_i \in W_i$. Then

$$(T_1 + T_2 + \cdots + T_k)(x) = T_1(x) + T_2(x) + \cdots + T_k(x) = x_1 = x_2 + \cdots + x_k,$$

so $T_1 + T_2 + \cdots + T_k = I$.

(5) Again,let $x = x_1 + x_2 + \cdots + x_k$, where each $x_i \in W_i$. Then

$$(\lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k)(x) = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k = T(x_1) + T(x_2) + \cdots + T(x_k) = T(x),$$

so $(\lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k) = T$.

This completes the proof. ■

4.7.6 Definition. Let V be a vector space. Let $T: V \to V$ be linear. The set of eigenvalues of T is called the *spectrum* of T and denoted $\sigma(T)$. The expression

$$T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$$

as in the spectral theorem is called the *spectral decomposition* of T.

4.7.7 Remark (Lagrange Interpolation). Let $c_0, c_1, \ldots, c_n \in F$ be distinct. Define

$$f_i(x) = \frac{\prod_{j \neq i}(x - c_j)}{\prod_{j \neq i}(c_i - c_j)},$$

and note that $f_i(c_j) = \delta_{ij}$ for all $1 \leq i, j \leq n$.

We claim that $\{f_0, f_1, \ldots, f_n\}$ is a basis for $P_n(F)$. Indeed, if

$$a_0 f_0 + a_1 f_1 + \cdots + a_n f_n = 0$$

for some $a_0, a_1, \ldots, a_n \in F$, then for every $1 \le i \le n$

$$0 = (a_0 f_0 + a_1 f_1 + \cdots + a_n f_n)(c_i) = a_i,$$

so $\{f_0, f_1, \ldots, f_n\}$ is linearly independent and hence forms a basis for $P_n(F)$.

4.7.8 Remark. With V and T as in the Spectral Theorem, $T^\ell = \lambda_1^\ell T_1 + \lambda_2^\ell T_2 + \cdots + \lambda_k^\ell T_k$ for $\ell \in \mathbb{N}$. It follows that for $f(x) \in F[x]$, $f(T) = f(\lambda_1)T_1 + f(\lambda_2)T_2 + \cdots + f(\lambda_k)T_k$.

4.7.9 Corollary. *Let V be a finite-dimensional inner product space over $\mathbb{C}$. Let $T\colon V \to V$ be a linear opreator. Then T is normal if and only if $T^* = f(T)$ for some $f(x) \in F[x]$.*

Proof. ($\Rightarrow$) Suppose T is normal. Let $T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$ be the spectral decomposition of T. Using Lagrange interpoation, let $f(x) \in F[x]$ such that $f(\lambda_i) = \overline{\lambda_i}$ for $1 \le i \le k$. Then

$$T^* = \sum_{i=1}^{k} \overline{\lambda_i} T_i^* = \sum_{i=1}^{k} \overline{\lambda_i} T_i = \sum_{i=1}^{k} f(\lambda_i) T_i = f(T),$$

since each T_i is an orthogonal projection and hence is Hermitian.

($\Leftarrow$) For any $f(x) \in F[x]$, $Tf(T) = f(T)T$, so $TT^* = T^*T$, and we're done.

4.7.10 Corollary. *Let V and T be as in the Spectral Theorem. Then for each $1 \le i \le k$, there exists $g_i \in F[x]$ such that $g_i(T) = T_i$.*

Proof. For each $1 \le i \le k$, choose $g_i \in F[x]$ such that $g_i(\lambda_j) = \delta_{ij}$ for $1 \le j \le k$.

July 8

4.7.11 Corollary. *Let V be a finite-dimensional inner product space over $\mathbb{C}$. Let $T\colon V \to V$ be linear. Then T is Hermitian if and only if T is normal and $\sigma(T) \subseteq \mathbb{R}$.*

Proof. ($\Rightarrow$) See part (1) of Proposition 4.5.11.

($\Leftarrow$) Let $\lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$ be the spectral decomposition of T, where $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$. Then

$$T^* = \overline{\lambda_1} T_1^* + \overline{\lambda_2} T_2^* + \cdots + \overline{\lambda_k} T_k^* = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k = T,$$

so T is Hermitian, and we're done.

4.7.12 Corollary. *Let T be a finite-dimensional inner product space over $\mathbb{C}$. Let $T\colon V \to V$ be linear. Then T is unitary if and only if T is normal and $|\lambda| = 1$ for all $\lambda \in \sigma(T)$.*

Proof. ($\Rightarrow$) See Corollary 4.5.19.

($\Leftarrow$) Suppose T is normal and $|\lambda| = 1$ for all $\lambda \in \sigma(T)$. Say the spectral decomposition of T is $T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$. Then $T^* = \overline{\lambda_1} T_1 + \overline{\lambda_2} T_2 + \cdots + \overline{\lambda_k} T_k$, so

$$TT^* = \lambda_1 \overline{\lambda_1} T_1 + \lambda_2 \overline{\lambda_2} T_2 + \cdots + \lambda_k \overline{\lambda_k} T_k = T_1 + T_2 + \cdots + T_k = I,$$

so T is unitary, and we're done.

4.8 Singular Value Decomposition

4.8.1 Definition. Let $T: V \to W$ be linear, where V and W are finite- dimensional inner product spaces over the same field F. Let $T: V \to W$ be linear. A function $T^*: W \to V$ is called an *adjoint* of T if $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$ for all $x \in V$ and $y \in W$.

4.8.2 Proposition. *Let $T: V \to W$ be linear, where V and W are finite- dimensional inner product spaces over the same field F. Let $T: V \to W$ be linear. Then:*

(1) *T^* exists, is unique and is linear.*

(2) *If β and γ are orthonormal bases for V and W respectively, then $[T^*]_\gamma^\beta = ([T]_\beta^\gamma)^*$.*

Proof.

(1) For arbitrary $y \in W$, $U_y: V \to F$ defined by $U_y(x) = \langle T(x), y\rangle$ is a linear functional. By the Riesz Representation Theorem, there exists a unique $y' \in V$ such that $U_y(x) = \langle x, y'\rangle$ for all $x \in V$. Define $T^*: W \to V$ by $T*(y) = y'$.

It remains to show that T^* is linear. Let $x \in V$, $y_1, y_2 \in W$, $\alpha \in F$. Then

$$\begin{aligned}\langle x, T^*(\alpha y_1 + y_2)\rangle &= \langle T(x), \alpha y_1 + y_2\rangle \\ &= \overline{\alpha}\langle T(x), y_1\rangle + \langle T(x), y_2\rangle \\ &= \overline{\alpha}\langle x, T^*(y_1)\rangle + \langle x, T^*(y_2)\rangle \\ &= \langle x, \alpha T^*(y_1)\rangle + \langle x, T^*(y_2)\rangle \\ &= \langle x, \alpha T^*(y_1) + T^*(y_2)\rangle.\end{aligned}$$

Since $x \in V$ was arbitrary, it follows that $T^($\alpha y_1 + y_2) = \alpha T^*(y_1) + T^*(y_2)$, so T^* is linear.

(2) Say $\beta = \{v_1, v_2, \ldots, v_n\}$ and $\gamma = \{w_1, w_2, \ldots, w_n\}$. Let $A = (a_{ij}) = [T]_\beta^\gamma$. Let $B = (b_{ij}) = [T^*]_\gamma^\beta$. Then by Proposition 4.2.8,

$$b_{ij} = \langle T^*(w_j), v_i\rangle = \overline{\langle v_i, T^*(w_j)\rangle} = \overline{\langle T(v_i), w_j\rangle} = \overline{a_{ji}},$$

so $B = A^*$, as required.

This completes the proof. ∎

4.8.3 Definition. Let V be a finite-dimensional inner product space. Let $T: V \to V$ be linear. We say that T is *positive semidefinite* if T is Hermitian and $\langle T(x), x\rangle \geq 0$ for all $x \in V$.

4.8.4 Proposition. *Let V be a finite-dimensional inner product space. Let $T: V \to V$ be a linear operator. Then*

(1) *T is positive semidefinite if and only if $T = T^*$ and $\sigma(T) \subseteq [0, \infty)$*

(2) *T is positive semidefinite if and only if $T = U^*U$ for some linear operator $U: V \to V$.*

Proof.

(1) ($\Rightarrow$) Suppose T is positive semidefinite. Then $T = T^*$ by definition; it follows that $\sigma(T) \subseteq \mathbb{R}$. If T has a negative eigenvalue, i.e., there exists $0 > \lambda \in \mathbb{R}$, $0 \neq v \in V$ such that $T(v) = \lambda v$, then $\langle T(v), v\rangle = \langle -\lambda v, v\rangle = -\lambda\langle v, v\rangle < 0$, since $\langle v, v\rangle \in \mathbb{R}^+$. But since T is positive semidefinite, this cannot be, so T has no negative eigenvalues. Hence $\sigma(T) \subseteq [0, \infty)$.

($\Leftarrow$) Suppose $T = T^*$ and $\sigma(T) \subseteq [0, \infty)$. Let $\beta = \{v_1, v_2, \ldots, v_n\}$ be an orthonormal basis for V composed of eigenvectors of T, where $n = \dim V$. Say $T(v_i) = \lambda_i v_i$ for $1 \leq i \leq n$. Let $x \in V$. Then $x = a_1v_1 + a_2v_2 + \cdots + a_nv_n$ for some $a_1, a_2, \ldots, a_n \in F$. Then

$$\langle T(x), x\rangle = \langle a_1T(v_1) + a_1T(v_2) + \cdots + a_nT(v_n), a_1v_2 + a_2v_2 + \cdots + a_nv_n\rangle$$

$$\begin{aligned}
&= \langle a_1\lambda_1 v_1 + a_1\lambda_2 v_2 + \cdots + a_n\lambda_n v_n, a_1 v_2 + a_2 v_2 + \cdots + a_n v_n\rangle \\
&= a_1\lambda_1\overline{a_1}\langle v_1, v_1\rangle + a_2\lambda_2\overline{a_2}\langle v_2, v_2\rangle + a_n\lambda_n\overline{a_n}\langle v_n, v_n\rangle \\
&= |a_1|^2\lambda_1 + |a_2|^2\lambda_2 + \cdots + |a_n|^2\lambda_n.
\end{aligned}$$

Since $\lambda_1, \lambda_2, \dots, \lambda_n \geq 0$, $\langle T(x), x\rangle \geq 0$, so T is positive semidefinite.

(2) ($\Leftarrow$) Suppose $T = U^*U$ for some linear operator $U\colon V \to V$. Let $x \in V$. Then

$$\langle T(x), x\rangle = \langle U^*U(x), x\rangle = \langle U(x), U(x)\rangle \in [0, \infty),$$

so T is positive definite.

($\Rightarrow$) Suppose T is positive semidefinite. Since T is Hermitian by definition, there exists an orthonormal basis $\beta = \{v_1, v_2, \dots, v_n\}$ for V consisting of eigenvectors of T. Furthermore, each v_i corresponds to some eigenvalue $\lambda_i \in [0, \infty)$. Define $U(v_i) = \sqrt{\lambda_i}v_i$ for each v_i and extend via linearity. Then β is an orthonormal basis for V composed of eigenvectors of U, so U is normal. Therefore each v_i is a $\overline{\sqrt{\lambda_i}} = \sqrt{\lambda_i}$ eigenvector for U^*. Then for each $v_i \in \beta$, $U^*U(v_i) = \lambda_i v_i = T(v_i)$, so $T = U^*U$.

This completes the proof. ∎

4.8.5 Proposition. *Let V, W be finite-dimensional inner product spaces over F. Let $T\colon V \to W$ be linear. Then T^*T and TT^* are positive semidefinite with* $\operatorname{rank}(T^*T) = \operatorname{rank}(T)$ *and* $\operatorname{rank}(T^*) = \operatorname{rank}(TT^*)$.

Proof. Let $x \in V$, $y \in W$. We see that

$$\langle T^*T(x), x\rangle = \langle T(x), T(x)\rangle \in [0, \infty) \qquad \langle TT^*(y), y\rangle = \langle T^*(y), T^*(y)\rangle \in [0, \infty),$$

so T^*T and TT^* are positive semidefinite.

We claim that $\operatorname{Null}(T) = \operatorname{Null}(T^*T)$. Clearly $\operatorname{Null}(T) \subseteq \operatorname{Null}(T^*T)$. Let $x \in \operatorname{Null}(T^*T)$. Then

$$\langle T(x), T(x)\rangle = \langle x, T^*T(x)\rangle = \langle x, 0\rangle = 0$$

so $T(x) = 0$ and in fact $\operatorname{Null}(T^*T) = \operatorname{Null}(T)$. Similarly, $\operatorname{Null}(TT^*) = \operatorname{Null}(T^*)$. Therefore we must have $\operatorname{rank}(T) = \operatorname{rank}(T^*T)$ and $\operatorname{rank}(T^*) = \operatorname{rank}(TT^*)$, so we're done. ∎

4.8.6 Theorem (Singular Value Decomposition)**.** *Let V, W be finite-dimensional inner product spaces over the same field F. Let $T\colon V \to W$ be linear. Let $\operatorname{rank}(T) = r$. Then there exist orthonormal bases $\{v_1, v_2, \dots, v_n\}$ and $\{u_1, u_2, \dots, u_m\}$ for V and W respectively and real scalars $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ such that $T(v_i) = \sigma_i u_i$ for $i \leq r$ and $T(v_i) = 0$ for $i > r$. (For $r < i < n$, we define $\sigma_i = 0$.) Conversely, if the above conclusion holds, then each v_i is a σ_i^2 eigenvector of T^*T. In particular, the σ_is are uniquely determined.*

Proof. Consider $T^*T\colon V \to V$. By Proposition 4.8.5, T^*T is positive semidefinite, and $\operatorname{rank}(T^*T) = \operatorname{rank}(T) = r$. Since T^*T is Hermitian, there exists an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ for V consisting of eigenvectors of T^*T. Say $T^*T(v_i) = \lambda_i v_i$ for $1 \leq i \leq n$, where $\lambda_i \in [0, \infty)$. For $1 \leq i \leq n$, let $\sigma_i = \sqrt{\lambda_i}$. Without loss of generality, assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ and $\lambda_i = 0$ for $r < i \leq n$. Then $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_i = 0$ for $r < i \leq n$. For $i \leq r$, let $u_i = \frac{1}{\sigma_i}T(v_i)$. Then

$$\langle u_i, u_j\rangle = \left\langle \frac{1}{\sigma_i}T(v_i), \frac{1}{\sigma_j}T(v_j)\right\rangle = \frac{1}{\sigma_i\sigma_j}\langle T^*T(v_i), v_j\rangle = \frac{1}{\sigma_i\sigma_j}\langle \lambda_i v_i, v_j\rangle = \frac{\lambda_i}{\sigma_i\sigma_j}\langle v_i, v_j\rangle = \delta_{ij}.$$

Therefore $\{u_1, u_2, \dots, u_r\}$ is an orthonormal set. By the Gram-Schmidt procedure, we may extend this to an orthonormal basis $\{u_1, u_2, \dots, u_r, \dots, u_m\}$ for W. Then for $i \leq r$, $T(v_i) = \sigma_i u_i$ and for $i > r$, $T^*T(v_i) = 0$, so by the proof of Proposition 4.8.5, $T(v_i) = 0$. *July 10*

It remains to show that the σ_is are uniquely determined. Suppose we have u_is, v_is, and σ_is as in the theorem statement. Then

$$T^*(u_i) = \sum_{j=1}^{n}\langle T^*(u_i), v_j\rangle v_j = \sum_{j=1}^{n}\langle u_i, T(v_j)\rangle v_j = \begin{cases}\sigma_i v_i & 1 \leq i \leq r \\ 0 & r < i \leq n\end{cases}.$$

For $i \le r$, $T^*T(v_i) = T^*(\sigma_i u_i) = \sigma_i T^*(u_i) = \sigma_i^2 v_i$ and for $i > r$ $T^*T(v_i) = T^*(0) = \sigma_i v_i$. Therefore the σ_is are the square roots of the eigenvalues of T^*T, so they are uniquely determined, and we're done. YEAH!

4.8.7 Definition. [4] In the Singular Value Theorem, the σ_is are called the *singular values* of T. If $r < m, n$, then we also call $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_k = 0$, where $k = \min(m,n)$, singular values of T.

4.8.8 Example. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ be given by $T(x,y) = (x, x+y, x-y)$. Let β_i be the standard (orthonormal) basis for $\mathbb{R}^i$. [5] Then

$$[T]_{\beta_2}^{\beta_3} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} = A.$$

Then

$$[T^*T]_{\beta_2}^{\beta_2} = [T^*]_{\beta_3}^{\beta_2}[T]_{\beta_2}^{\beta_3} = \left([T]_{\beta_2}^{\beta_3}\right)^*[T]_{\beta_2}^{\beta_3} = A^*A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}.$$

Therefore $\beta_2 = \{v_1 = e_1, v_2 = e_2\}$ is an orthonormal basis for $\mathbb{R}^2$ such that $T^*T(v_1) = 3v_1$ and $T^*T(v_2) = 2v_2$. Setting $\lambda_1 = 3 > 2 = \lambda_2$, we obtain $\sigma_1 = \sqrt{3}$ and $\sigma_2 = \sqrt{2}$. Let $u_1 = \frac{1}{\sigma_1}T(v_1) = \frac{1}{\sqrt{3}}(1,1,1)$ and $u_2 = \frac{1}{\sigma_2}T(v_2) = \frac{1}{\sqrt{2}}(0,1,-1)$. Conveniently, $\{u_1, u_2, e_3\}$ is a basis for $\mathbb{R}^3$. Applying the Gram-Schmidt procedure, let

$$u_3 = e_3 - \frac{\langle e_3, u_1\rangle}{\|u_1\|^2}u_1 - \frac{\langle e_3, u_2\rangle}{\|u_2\|^2}u_2 = e_3 - \frac{1}{\sqrt{3}}u_1 + \frac{1}{\sqrt{2}}u_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

Then setting $u_3 = \frac{1}{\sqrt{6}}(-2,1,1)$ gives us an orthonormal basis $\gamma = \{u_1, u_2, u_3\}$ for $\mathbb{R}^3$. Let $\beta = \{v_1, v_2\}$. Then

$$[T]_{\beta_2}^{\beta_3} = [I]_{\gamma}^{\beta_3}[T]_{\beta}^{\gamma}[I]_{\beta_2}^{\beta} = [I]_{\gamma}^{\beta_3}[T]_{\beta}^{\gamma}\left([I]_{\beta}^{\beta_2}\right)^* = UDV^*,$$

where

$$U = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix} \qquad D = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \qquad V = \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

4.8.9 Definition. Let $A \in M_{m\times n}(F)$. The *singular values* of A are the singular values of $L_A\colon F^n \to F^m$.

4.8.10 Theorem (Singular Value)**.** *Let $A \in M_{m\times n}(F)$. Let $\operatorname{rank}(A) = r$. Say the singular values of A are $\sigma_1 \ge \sigma_2 \ge \cdots \sigma_r > 0$. Then there exist unitary $U \in M_m(F)$ and unitary $V \in M_n(F)$ such that $A = UDV^*$, where $D = (d_{ij})$ and*

$$d_{ij} = \begin{cases} \sigma_i & i = j \\ 0 & i \ne j \end{cases}.$$

5 Tensors

July 12

5.1 Quotient Spaces

5.1.1 Notation. Throughout this section, F denotes an arbitrary field (no longer restricted to $\mathbb{R}$ or $\mathbb{C}$) and V denotes a vector space over F.

5.1.2 Definition. Let V be a vector space over F. Let $W \le V$, $v \in V$. The *coset* of W in V, containing v, is $v + W := \{v + w : w \in W\}$. We use the notation $V/W := \{v + W : v \in V\}$. Additionally, we shall sometimes denote $v + W = \overline{v}$ when the subspace W is clear.

[4] Definition 4.8.7 was presented on July 8, before the proof of the Singular Value Decompositionl Theorem.
[5] We avoid using σ_i to avoid confusion with singular values.

5.1.3 Remark. Let $W \leq V$. Note that $a + W = b + W \iff a - b \in W$, and in particular $a + W = 0 + W \iff a \in W$.

5.1.4 Proposition. *Let $W \leq V$. Then V/W is a vector spaces over F when equipped with the operations*

$$(a + W) + (b + W) = (a + b) + W \qquad\qquad \alpha(a + W) = (\alpha a) + W.$$

Proof. These operations satisfy the vector space axioms since V is a vector space. We just need to check that they are well-defined. Accordingly, suppose $a, b, a', b' \in V$ such that $a+W = a'+W$ and $b+W = b'+W$. Then $a-a' \in W$ and $b-b' \in W$, so $(a+b)-(a'+b') = (a-a')+(b-b') \in W$, and therefore $(a+b)+W = (a'+b')+W$. Furthermore, $\alpha a-\alpha a' = \alpha(a-a') \in W$, so $(\alpha a)+W = (\alpha a')+W$. Therefore these operations are independent of the choice of coset representative, so we're done. ∎

5.1.5 Example. In $P_3(\mathbb{R})/P_2(\mathbb{R})$, $\overline{6x^3 - 5x^2 + 2x - 1} = \overline{6x^3}$.

5.1.6 Definition. Let V, W be vector spaces over F. We say that V and W are *isomorphic* and write $V \cong W$ when there exists an invertible linear transformation $T\colon V \to W$. We call such an invertible linear transformation an *isomorphism.*

5.1.7 Theorem (First Isomorphism Theorem for Vector Spaces)**.** *Let V, W be vector spaces over F. Let $T\colon V \to W$ be linear. Then $V/\operatorname{Null}(T) \cong T(V) \leq U$ via the isomorphism $\overline{v} \mapsto T(v)$.*

Proof. Define $\varphi\colon V/\operatorname{Null}(T) \to U$ by $\varphi(\overline{v}) = T(v)$. We claim that φ is a well-defined injective linear transformation. Note that if $\overline{u} = \overline{v} \in V/\operatorname{Null}(T)$, then $u - v \in \operatorname{Null}(T)$, so $T(u - v) = 0$. Thus $\varphi(\overline{u}) = T(u) = T(v) = \varphi(\overline{v})$, so φ is well-defined. Let $\overline{x}, \overline{y} \in V/\operatorname{Null}(T)$, $\alpha \in F$. Then

$$\varphi(\alpha\overline{x} + \overline{y}) = \varphi(\overline{\alpha x + y}) = T(\alpha x + y) = \alpha T(x) + y = \alpha\varphi(\overline{x}) + \varphi(\overline{y}),$$

so φ is linear. Finally, suppose that $\overline{v} \in \operatorname{Null}(\varphi)$. Then $\varphi(\overline{v}) = T(v) = 0$, so $v \in \operatorname{Null}(T) = \overline{0}$. Thus φ is injective. It follows that $V/\operatorname{Null}(T) \cong T(V)$, as required. ∎

July 15

5.1.8 Example. Let $V = M_2(\mathbb{R})$. Let $W = \{A \in V : A = A^T\}$. Then

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + W = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + W.$$

5.1.9 Proposition. *Let V be a finite-dimensional vector space over F. Let W be a subspace of V. Say $\{v_1, v_2, \dots, v_m\}$ is a basis for W. Then extend this basis to a basis for V, $\{v_1, v_2, \dots, v_m, v_{m+1}, \dots, v_n\}$, where $n = \dim V$. Then $\{\overline{v_{m+1}}, \overline{v_{m+2}}, \dots, \overline{v_n}\}$ is a basis for V/W. In particular,* $\dim(V/W) = \dim V - \dim W$.

Proof. Let $\overline{v} = v + W \in V/W$. Say $v = \sum_{i=1}^n a_i v_i$, where $a_1, a_2, \dots, a_n \in F$. Then

$$\overline{v} = \overline{a_{m+1}v_{m+1} + a_{m+2}v_{m+2} + \cdots + a_n v_n} = a_{m+1}\overline{v_{m+1}} + a_{m+2}\overline{v_{m+2}} + \cdots + a_n\overline{v_n},$$

so this set spans V. Now suppose $b_{m+1}\overline{v_{m+1}} + b_{m+2}\overline{v_{m+2}} + \cdots + b_n\overline{v_n} = 0$ for some $b_{m+1}, b_{m+2}, \dots, b_n \in F$. Then $b_{m+1}v_{m+1} + b_{m+2}v_{m+2} + \cdots + b_n v_n \in W$, so $b_{m+1} = b_{m+2} = \cdots = b_n = 0$. Therefore $\{\overline{v_{m+1}}, \overline{v_{m+2}}, \dots, \overline{v_n}\}$ is a linearly independent spanning set, i.e., a basis, for V/W. ∎

5.2 Tensor Products

5.2.1 Example. Let $(a, b) \in \mathbb{C}^2$. Let $S = \{(a, b) - (b, a) : a, b \in \mathbb{C}\}$ and let $W = \operatorname{Span}(S)$. In $\mathbb{C}^2/W$, $\overline{(a, b) - (b, a)} = \overline{0}$, so $\overline{(a, b)} = \overline{(b, a)}$.

5.1 Remark. Similarly to the above example, our goal is to turn a vector space V into a ring $(T(v), +, \otimes)$ with an additional bilinear scalar multiplication operation.

5.2.2 Definition. Let X be a set of algebraically independent symbols. We define the free vector space on X over F by

$$V = \operatorname{Free}(X) = \{\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n : \alpha_i \in F,\ x_i \in X\},$$

with addition defined by $\sum \alpha_i x_i + \sum \beta_i x_i = \sum(\alpha + \beta)x_i$ and scalar multiplication by $\lambda \sum \alpha_i x_i = \sum \lambda_i \alpha_i x_i$.

5.2.3 Example. Let $F = \mathbb{R}$ and $X = \{\star, \natural, \smiley\}$. Then in Free$(X)$,

$$\left(-\star + 2\natural - \frac{15}{2}\smiley\right) + \left(2\star - 2\natural + \frac{1}{2}\smiley\right) = \star + 0\natural - 7\smiley,$$

which we denote simply by $\star - 7\smiley$.

5.2.4 Remark. By construction, X is a basis for Free(X).

5.2.5 Definition. Let V, W be finite-dimensional vector spaces over F. Let $X = V \times W$, treated as a set of symbols. Let S be the set of vectors in Free(X) of one of the forms

- $(x + y, z) - (x, z) - (y, z)$
- $(z, x + y) - (z, x) - (x, y)$
- $\alpha(x, y) - (\alpha x, y)$
- $\alpha(x, y) - (x \alpha y)$.

We define the *tensor product* of V and W to be $V \oplus W \colon \operatorname{Free}(X) / \operatorname{Span}(S)$.

5.2.6 Notation. In $V \otimes W$, we denote $\overline{(v, w)} = v \otimes w$. Elements of this form are called *pure tensors*. *July 17*

5.2.7 Remark. In $V \otimes W$, note that $(v+w) \otimes z - v \otimes z - w \otimes z = 0 \otimes 0 =: 0$. Therefore $(v+w) \otimes z = v \otimes z + w \otimes z$. Also, $\alpha(v \otimes w) = (\alpha v) \otimes w = v \otimes (\alpha w)$.

5.2.8 Example. Consider $\mathbb{C}^2 \otimes_{\mathbb{C}} \mathbb{C}^3$. (This notation means that we are using the field of scalars $\mathbb{C}$.) Let the standard bases for $\mathbb{C}^2$ and $\mathbb{C}^3$ be $\sigma_2 = \{a_1, a_2\}$ and $\sigma_3 = \{b_1, b_2, b_3\}$ respectively. Then

$$\begin{aligned}(1, 2) \otimes (1, 2, 3) &= (a_1 + 2a_2) \otimes (b_1 + 2b_2 + 3b_3) \\ &= a_1 \otimes (b_1 + 2b_2 + 3b_3) + 2(a_2 \otimes (b_1 + 2b_2 + 3b_3)) \\ &= (a_1 \otimes b_1) + 2(a_1 \otimes b_2) + 3(a_1 \otimes b_3) + 2(a_2 \otimes b_1) + 4(a_2 \otimes b_2) + 6(a_2 \otimes b_3).\end{aligned}$$

5.2.9 Proposition. *Let V, W be finite-dimensional vector spaces over F. Suppose that $\{v_1, v_2, \ldots, v_n\}$ and $\{w_1, w_2, \ldots, w_m\}$ are bases for V and W respectively. Then a basis for $V \otimes_F W$ is*

$$\{v_i \otimes w_j : 1 \le i \le n,\ 1 \le j \le m\}.$$

In particular, $\dim_F (V \otimes_F W) = nm = \dim_F (V) \dim_F (W)$.

5.2.10 Theorem (Universal Property of Tensor Products)**.** *Let V, W, Z be vector spaces over some field F. Let $\varphi \colon V \times W \to Z$ be bilinear. Then there exists a unique linear transformation $T \colon V \otimes W \to Z$ defined by $T(v \otimes w) = \varphi(v, w)$. Moreover, all linear transformations $V \otimes W \to Z$ can be constructed in this way.*

5.2.11 Remark. Let V be a finite-dimensional vector space over F. Let $\{v_1, v_2, \ldots, v_n\}$ be a basis for V. For each $1 \le i \le n$, define $v_i^* \colon V \to F$ by $v_i^*(v_j) = \delta_{ij}$ for $1 \le j \le n$. Then $\{v_1^*, v_2^*, \ldots, v_n^*\}$ is a basis for V^*.

5.2.12 Notation. Let V, W be finite-dimensional vector spaces over F. We denote the collection of linear transformations from V to W by $L(V, W)$. Note that $L(V, W)$ is a vector space over F.

5.2.13 Example. Let V, W be finite-dimensional vector spaces over F. We show that $V^* \otimes_F W \cong L(V, W)$.

Define $\varphi \colon V^* \times W \to L(V, W)$ by $\varphi(f, w)(v) = f(v)w$, where $v \in V$ is arbitrary. Confirm that φ is bilinear and well-defined. By the Universal Property, there is a linear transformation $T \colon V^* \otimes_F W \to L(V, W)$ such that $T(f \otimes w) = \varphi(f, w)$. We will show that T is an isomorphism by explicitly constructing its inverse. *July 19*

Let $\{w_1, w_2, \ldots, w_m\}$ be a basis for W. Define a basis for W^* by $\{w_1^*, w_2^*, \ldots, w_m^*\}$ as in Remark 5.2.11. Define $U \colon L(v, w) \to V^* \otimes_F W$ by $U(F) = \sum_{i=1}^m (w_i^* \circ F) \otimes w_i$. Let $v \in V$. Say $F(v) = \sum_{i=1}^m \alpha_i w_i$, where $\alpha_1, \alpha_2, \ldots, \alpha_m \in F$. Then

$$(TU)(F)(v) = T\left(\sum_{i=1}^m (w_i^* F) \otimes w_i\right)(v) = \sum_{i=1}^m w_i^*(F(v)) w_i = \sum_{i=1}^m w_i^* \left(\sum_{j=1}^m \alpha_j w_j\right) w_i = \sum_{i=1}^m \alpha_i w_i = F(v),$$

so $U = T^{-1}$, and therefore T is an isomorphism, as claimed.

5.3 Tensor and Exterior Algebras

5.3.1 Definition. Let F be a field. An F-algebra is a vector space A over F equipped with a multiplication map $\cdot\colon A \times A \to A$ such that

- $a \cdot (b \cdot c) = (a \cdot b) \cdot c$
- $a \cdot (b + c) = a \cdot b + a \cdot c$
- $(a + b) \cdot c = a \cdot c + b \cdot c$
- $\alpha(a \cdot b) = (\alpha a) \cdot b = a \cdot (\alpha b)$

for all $a, b, c \in A$, $\alpha \in F$.

5.3.2 Definition. Let V be a vector space over F. For $k \in \mathbb{N}$, we define $T^k(V) = \bigotimes_{i=1}^k V$. Elements of $T^k(V)$ are called *k-tensors*. We also define $T^0(V) = F$.

5.3.3 Aside. Let V be a vector space over F. Let $W_1, W_2, W_3, \ldots \leq V$. We define the direct product of $W_1, W_2, W_3 \ldots$ to be

$$\prod_{i=1}^{\infty} w_i = \{(a_1, a_2, a_3, \ldots) : a_i \in W_i\}$$

and the direct sum of $W_1, W_2, W_3 \ldots$ to be

$$\bigoplus_{i=1}^{\infty} W_i = \{(a_1, a_2, a_3, \ldots) : a_i \in W_i,\ a_i = 0 \text{ for all but finitely many } i\}.$$

We denote $(a_1, a_2, a_3, \ldots) \in \bigoplus_{i=1}^\infty W_i$ by $a_1 + a_2 + a_3 + \cdots$. Note however, that this is notation only; we are not using addition in V.

5.3.4 Example. For $i = 0, 1, 2, \ldots$, define $W_i = \operatorname{Span}_{\mathbb{R}}\{x^i\}$. Then $\bigoplus_{i=1}^\infty W_i = \mathbb{R}[x]$.

5.3.5 Definition. We define the *tensor algebra* of V by $T(V) = \bigoplus_{i=0}^\infty T^i(V)$. Elements of $T(V)$ look like finite linear combinations of k-tensors.

5.3.6 Example. Let $F = \mathbb{R}$ and let V be a vector space over $\mathbb{R}$. Let $x, y \in V$. Then

$$3 + 2(x \otimes y) - \frac{1}{7}(x \otimes x \otimes y) + 87(x \otimes x \otimes x \otimes x \otimes x) \in T(V).$$

5.3.7 Definition. In $T(V)$, multiplication is defined by

$$(v_1 \otimes v_2 \otimes \cdots \otimes v_k)(u_1 \otimes u_2 \otimes \cdots \otimes u_\ell) = v_1 \otimes v_2 \otimes \cdots \otimes v_k \otimes u_1 \otimes u_2 \otimes \cdots \otimes u_\ell$$

and then extended by distributivity.

5.3.8 Definition. Let V be a vector space over F. Let $A(V)$ in $T(V)$ be the ideal generated by elements of the form $v \otimes v$, where $v \in V$. We define the *exterior algebra* of V by $\bigwedge(V) = T(V)/A(V)$, equipped with operations given by $\overline{x} + \overline{y} = \overline{x + y}$, $\alpha\overline{x} = \overline{\alpha x}$, and $\overline{x}\,\overline{y} = \overline{xy}$ for all $x, y \in T(V)$, $\alpha \in F$.

5.3.9 Notation. In $\bigwedge(V)$, we denote $\overline{v_1 \otimes v_2 \otimes \cdots \otimes v_k}$ by $v_1 \wedge v_2 \wedge \cdots \wedge v_k$.

5.3.10 Example. In $\bigwedge(V)$,

$$0 = (x + y) \wedge (x + y) = x \wedge x + x \wedge y + y \wedge x + y \wedge y = x \wedge y + y \wedge x,$$

so $x \wedge y = -(y \wedge x)$. Similarly, $a \wedge b \wedge c \wedge a \wedge e = 0$. Note that $0 \otimes v = 0(0 \otimes v) = 0$ for any $v \in V$.

6 Functional Analysis

July 22

6.1 Definition. Let $(V, \|\cdot\|)$ be a normed vector space.

(1) We say that a sequence (x_n) in V *converges* to $x \in V$, denoted $x_n \to x$, if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, $n \geq N \implies \|x_n - x\| < \varepsilon$.

(2) We say that a sequence (x_n) in V is *Cauchy* if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n, m \in \mathbb{N}$, $n, m \geq N \implies \|x_n - x_n\| < \varepsilon$.

(3) We say that V is *complete* if every Cauchy sequence in V converges in V.

(4) If V is complete, we call it a *Banach space*.

(5) If V is a Banach Space and the norm on V is defined by $\|v\| = \sqrt{\langle v, v\rangle}$ for some inner product $\langle\cdot,\cdot\rangle$ on V, we call V a *Hilbert space*.

6.2 Example. $(\mathbb{R}^n, \|\cdot\|)$ and $(\mathbb{C}^n, \|\cdot\|)$ are Hilbert spaces. In fact, they are the only finite-dimensional Hilbert spaces, up to isomorphism.

6.3 Definition. Define

$$c_{00} = \{(x_n)_{n=1}^{\infty} : x_n \in \mathbb{R},\ x_n = 0 \text{ for all but finitely many } n \in \mathbb{N}\}$$
$$c_0 = \left\{(x_n)_{n=1}^{\infty} : x_n \in \mathbb{R},\ \lim_{n\to\infty} x_n = 0\right\}$$

A norm on c_{00} and c_0 is given by $\|(x_n)\|_\infty = \max_{n\in\mathbb{N}}\{|x_n|\}$.

6.4 Example. Let $x_n = \left(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, 0, 0, \ldots\right)$. We claim that $(x_n) \in c_{00}$ is Cauchy. Accordingly, let $\varepsilon > 0$. Choose $n \in \mathbb{N}$ such that $\frac{1}{N} < \varepsilon$. Suppose $n, m \geq N$; without loss of generality assume $n < m$. Then

$$\|x_n - x_m\|_\infty = \frac{1}{n+1} < \frac{1}{n} \leq \frac{1}{N} < \varepsilon.$$

However, it is clear that $x_n \to x$, where $x = \left(1, \frac{1}{2}, \frac{1}{3}, \ldots\right) \notin c_{00}$. By the uniqueness of limits, it follows that c_{00} is not a Banach space.

6.5 Example. $(c_0, \|\cdot\|_\infty)$ is a Banach space, for reasons apparent from the preceding example.

6.6 Definition. Define

$$\ell^\infty = \left\{(a_n)_{n=1}^{\infty} : a_n \in \mathbb{R},\ \sup_{n\in\mathbb{N}} |a_n| < \infty\right\}.$$

A norm on ℓ^∞ is given by $\|(a_n)\|_\infty = \sup_{n\in\mathbb{N}}\{|a_n|\}$.

6.7 Example. We claim that $(\ell^\infty, \|\cdot\|_\infty)$ is a Banach space. Let (x_n) be a Cauchy sequence in ℓ^∞. We write

$$x_n = \left(x_n^{(1)}, x_n^{(2)}, x_n^{(3)}, \ldots\right).$$

Let $\varepsilon > 0$. Then there exists $N_1 \in \mathbb{N}$ such that $\|x_n - x_m\|_\infty < \frac{\varepsilon}{2}$ for $n, m \geq N_1$. Then for every $i \in \mathbb{N}$, for $n, m \geq N_1$, we have

$$\left|x_n^{(i)} - x_m^{(i)}\right| \leq \|x_n - x_m\|_\infty < \frac{\varepsilon}{2} < \varepsilon.$$

Therefore the component sequences are Cauchy, hence convergent. Say each $\left(x_n^{(i)}\right)$ converges to a_i as $n \to \infty$.

Let $x = (a_1, a_2, a_3, \ldots)$. We claim that $x \in \ell^\infty$. Note that there exists $N_2 \in \mathbb{N}$ such that $\|x_n - x_m\|_\infty < 1$ for $n, m \geq N_2$. Then for $n, m \geq N_2$, $i \in \mathbb{N}$,

$$\left|x_n^{(i)} - x_m^{(i)}\right| \leq \|x_n - x_m\|_\infty < 1.$$

Now,

$$\left|x_n^{(i)} - a_i\right| = \lim_{m\to\infty} \left|x_n^{(i)} - x_m^{(i)}\right| \leq 1,$$

so for $n \geq N_2$,

$$\sup_{i\in\mathbb{N}} |a_i| = \sup_{i\in\mathbb{N}} \left|a_i - x_n^{(i)} + x_n^{(i)}\right| \leq \sup_{i\in\mathbb{N}} \left\{\left|a_i - x_n^{(i)}\right| + \left|x_n^{(i)}\right|\right\} \leq 1 + \|x_n\|_\infty < \infty.$$

Therefore $x \in \ell^\infty$.

Now claim that $x_n \to x$. For $i \in \mathbb{N}$, $n, m \geq N_1$,

$$\left|x_m^{(i)} - x_n^{(i)}\right| \leq \|x_m - x_n\|_\infty < \frac{\varepsilon}{2}.$$

Then

$$\left|x_m^{(i)} - a_i\right| = \lim_{n\to\infty} \left|x_m^{(i)} - x_n^{(i)}\right| \leq \frac{\varepsilon}{2}.$$

For $m \geq N_1$,

$$\|x_m - x\|_\infty = \sup_{i\in\mathbb{N}} \left|x_m^{(i)} - a_i\right| \leq \frac{\varepsilon}{2} < \varepsilon.$$

Therefore $x_n \to x$. It follows that $(\ell^\infty, \|\cdot\|_\infty)$ is a Banach space.

6.8 Remark. A closed subset of a Banach space is also a Banach space.

6.9 Example. $c_0 \subseteq \ell^\infty$ is a Banach space.

July 24

6.10 Definition. Let $p \in [1, \infty)$. Let $(a_n)_{n=1}^\infty$ be a sequence in $\mathbb{R}$. Define the $\|\cdot\|_p$ to be

$$\|(a_n)\|_p = \left(\sum_{n=1}^\infty |a_n|^p\right)^{\frac{1}{p}}.$$

Define $\ell^p = \{(a_n)_n : \|(a_n)\|_p < \infty\}$.

6.11 Fact. *$(\ell^p, \|\cdot\|_p)$ is a Banach space. When $p = 2$, ℓ^2 is a Hilbert space, where $\langle(a_n), (b_n)\rangle = \sum_{n=1}^\infty a_n b_n$.*

6.12 Fact. *Let $(V, \|\cdot\|)$ be a normed vector space. Then the parallelogram law holds in V if and only if $\|\cdot\|$ is induced by an inner product.*

6.13 Example. We claim that ℓ^∞ is not a Hilbert space. Let $x = (1, 0, 0, \ldots)$, $y = (0, 1, 0, \ldots)$. Then

$$\|x + y\|_\infty^2 + \|x - y\|_\infty^2 = 1^2 + 1^2 = 2 \neq 4 = 2(1 + 1) = 2(\|x\|_\infty^2 + \|y\|_\infty^2).$$

Therefore the parallelogram law does not hold in ℓ^∞, so ℓ^∞ is not a Hilbert space by Fact 6.12.

6.14 Fact. *$(\ell^p, \|\cdot\|_p)$ is a Hilbert space if and only if $p = 2$.*

6.15 Remark. Using the same x, y as in Example 6.13, but in ℓ^p instead of ℓ^∞, we have

$$\|x + y\|_p^2 + \|x - y\|_p^2 = (1^p + 1^p)^{\frac{2}{p}} + (1^p + 1^p)^{\frac{2}{p}} = 2^{\frac{2}{p}+1}$$

while

$$2(\|x\|_p^2 + \|y\|_p^2) = 2(1 + 1) = 4.$$

Clearly equality does not hold unless $p = 2$, which gives some insight into Fact 6.14.

6.16 Definition. Let V, W be normed vector spaces. Let $T\colon V \to W$ be linear. We say that

(1) T is *continuous at* $v \in V$ if for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in V$,

$$\|x - v\| < \delta \implies \|T(x) - T(v)\| < \varepsilon.$$

(2) T is *continuous* if it is continuous at every $v \in V$.

(3) T is bounded if there exists $C \geq 0$ such that $\|T(x)\| \leq C\|x\|$ for all $x \in V$.

6.17 Theorem. *Let V, W be normed vector spaces. Let $T\colon V \to W$ be linear. The following are equivalent:*

(1) *T is continuous.*

(2) *T is continuous at 0.*

(3) *T is bounded.*

(4) *$n_1 = \sup\{\|T(x)\| : \|x\| \leq 1\} < \infty$.*

(5) *$n_2 = \sup\{\|T(x)\| : \|x\| = 1\} < \infty$.*

Proof. (1) $\Rightarrow$ (2) Trivial.

(2) $\Rightarrow$ (3) Suppose T is continuous at 0. Then there exists $\delta > 0$ such that $\|x\| < \delta \implies \|T(x)\| < 1$. For $0 \neq x \in V$,

$$\frac{\delta}{2\|x\|}\|T(x)\| = \left\|T\left(\frac{\delta}{2\|x\|}x\right)\right\| < 1,$$

since $\left\|\frac{\delta}{2\|x\|}x\right\| < \delta$. Therefore

$$\|T(x)\| < \frac{2}{\delta}\|x\|,$$

so we set $C = \frac{2}{\delta}$ and the result follows.

(3) $\Rightarrow$ (4) Suppose T is bounded. Say $\|T(x)\| \leq C\|x\|$, $C \geq 0$. Then for $x \in V$ with $\|x\| \leq 1$, $\|T(x)\| \leq C\|x\| \leq C$, so $n_1 \leq C < \infty$.

(4) $\Rightarrow$ (5) Trivial.

(5) $\Rightarrow$ (1) Suppose $n_2 < \infty$. Let $v \in V$. Let $\varepsilon > 0$. Choose $\delta = \frac{\varepsilon}{n_2+1}$. Suppose $x \in V$ with $\|x - v\| < \delta$. If $x = v$ then $\|T(x) - T(v)\| = 0 < \varepsilon$. Otherwise,

$$\|T(x) - T(v)\| = \|T(x - v)\| = \left\|T\left(\frac{x-v}{\|x-v\|}\right)\right\| \|x - v\| \leq n_2\|x - v\| < n_2\delta < \varepsilon,$$

so T is continuous. This completes the proof. ∎

6.18 Remark. Suppose $T\colon V \to W$ is continuous and n_1, n_2 are defined as in Theorem 6.17. Clearly $n_2 \leq n_1$. If $x \in V$ with $0 < \|x\| \leq 1$, we have

$$\left\|T\left(\frac{x}{\|x\|}\right)\right\| = \frac{1}{\|x\|}\|T(x)\| \leq n_2,$$

so $\|T(x)\| \leq n_2\|x\| \leq n_2$. Therefore $n_1 \leq n_2$, so $n_1 = n_2$. We can use n_2 to define the *operator norm* given by

$$\|T\| = \sup_{\|x\|=1} \|T(x)\|.$$

This completes the course. ∎