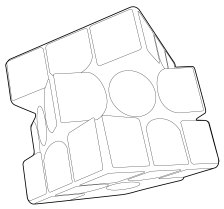
MATH 245

Linear Algebra II



This page intentionally left blank.

Contents

This page intentionally left blank.

4.
Polynomials

Power Series

(4.1) Examples of Fields

Throughout this course, we are going to write $\mathbb{F}$ do denote an arbitrary field. Here are some examples:

(a) $\mathbb{F} = \mathbb{Q}$ is the *field of rationals*.

(b) $\mathbb{F} = \mathbb{R}$ is the *field of reals*.

(c) $\mathbb{F} = \mathbb{C}$ is the *field of complex numbers*.

(d) $\mathbb{F} = \mathbb{Z}/p\mathbb{Z}$ for any prime $p \in \mathbb{N}$ is the *field of integers modulo p*.

The key property of these fields is that, for any $x \in \mathbb{F}$, there exists $x^{-1} \in \mathbb{F}$ such that $xx^{-1} = 1_{\mathbb{F}}$, the unity of $\mathbb{F}$.

Def'n. Power Series over a Field
Let $\mathbb{F}$ be a field. A ***power series*** over $\mathbb{F}$ is a *formal sum*

$$\sum_{i=0}^{\infty} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots,$$

where x is an ***indeterminate*** and the $a_i \in \mathbb{F}$ are called the ***coefficients***.

(4.2) Power Series as a Formal Sum

We say that the power sum is *formal* because we are not really summing anything. Of course, the notation suggests that we are adding infinitely many numbers up. However, such operation only makes sence when its partial sums converges, as done in calculus. But we are using an abstract algebraic setting, and, therefore, there is no notion of convergence. Whenever it is clear from the context, we are going to drop the indices and write

$$\sum a_i x^i.$$

It is simply an alternative form of writing an infinite sequence $(a_0, a_1, \ldots, a_n, \ldots)$ in an intuitive way. To highlight why this sum is formal, let us discuss when we consider two power series to be equal. That is, we say two power series $\sum a_i x^i$ and $\sum b_i x^i$ are equal, or

$$\sum a_i x^i = \sum b_i x^i$$

if and only if $a_i = b_i$ for all $i \in \mathbb{N}$, where we include 0 as a natural number for convenience as well.

Def'n. Set of Power Series over a Field
We denote the ***set of power series*** over $\mathbb{F}$ by $\mathbb{F}[\![x]\!]$.

(4.3)

We expect $\mathbb{F}[\![x]\!]$ to be an algebraic structure. That is, we may carefully choose operations for $\mathbb{F}[\![x]\!]$ such that it has pleasing algebraic properties.

(a) *addition*:

$$\left(\sum a_i x^i\right) + \left(\sum b_i x^i\right) = \sum (a_i + b_i) x^i.$$

This is called the *coefficient-wise* addition.

(b) *additive identity*:

$$0_{\mathbb{F}[\![x]\!]} = \sum 0_{\mathbb{F}} x^i.$$

That is, the additive identity, or the zero, of $\mathbb{F}[\![x]\!]$ is the formal sum which every coefficient $a_i = 0$, the zero of the corresponding field $\mathbb{F}$.

(c) *additive inverse*:
$$-\left(\sum a_i x^i\right) = \sum (-a_i) x^i.$$

(d) *scalar multiplication*: For any $c \in \mathbb{F}$,
$$c\left(\sum a_i x^i\right) = \sum (ca_i) x^i.$$
That is, $\mathbb{F}[\![x]\!]$ is a *vector space* over $\mathbb{F}$.

(e) *multiplication*: Unlike the above operations, multiplication of power series is not coefficient-wise. Here is where the intuition of sum helps. That is, if we think each power series to be a sum, then a term $a_i x^i$ would be multiplied to each term $b_j x^j$, $j \in \mathbb{N}$. In other words, we define multiplication such that if we are to write
$$\left(\sum a_i x^i\right)\left(\sum b_i x^i\right) = \sum c_i x^i,$$
then
$$c_i = \sum_{j=0}^{i} a_i b_{i-j}.$$
Notice that the above sum is not formal, but an actual, finite sum inside the field $\mathbb{F}$.

(f) *unity*:
$$1_{\mathbb{F}[\![x]\!]} = 1 + 0x + 0x^2 + 0x^3 + \cdots.$$
That is, $\mathbb{F}[\![x]\!]$ is an *algebra* over $\mathbb{F}$.

(4.4) Here are some notational conventions when we are writing a power series:

(a) Write x instead of x^1.

(b) Omit x^0.

(c) Omit any terms with zero coefficient.

Now we are ready to define the algebra of polynomials.

Polynomial Algebra

Def'n. Algebra of Polynomials
We denote by $\mathbb{F}[x] \subseteq \mathbb{F}[\![x]\!]$ the subalgebra of power series but finitely many of whose coefficients are nonzero, which we call the ***algebra of polynomials***.

Def'n. Degree, Leading Coefficient of a Polynomial
Let
$$f = \sum a_i x^i \in \mathbb{F}[x]$$
be nonzero. We define the ***degree*** of f, denoted by $\deg(f)$, to be the largest $d \in \mathbb{N}$ such that $a_d \neq 0$. We call a_d the ***leading coefficient*** of f.

(4.5) Here are few remarks:

(a) We do not assign degree to $f = 0 \in \mathbb{F}[x]$.

(b) In case of $\deg(f) = d$, we may safely write

$$f = \sum^{d} a_i x^i.$$

Def'n. Constant, Linear Polynomials
A ***constant*** polynomial is 0 or a degree zero polynomial. A ***linear*** polynomial is a degree one polynomial.

Proposition 4.1. *Let $f, g \in \mathbb{F}[x]$ be nonzero. Then $fg \neq 0$, $fg \in \mathbb{F}[x]$, and in particular,*

$$\deg(fg) = \deg(f) + \deg(g).$$

Furthermore, the leading coefficient of fg is the product of leading coefficients of f and g.

Proof. Let $m = \deg(f)$ and $f = \sum^m f_i x^i$ and let $n = \deg(g)$ and $g = \sum^n g_i x^i$. Then,

$$(fg)_{m+n} = \sum_i^{m+n} f_i g_{m+n-i} x^i$$

by definition in $\mathbb{F}[\![x]\!]$. Notice that,

$$i > m \implies f_i = 0.$$

Similarly,

$$i < m \implies m+n-i > n = \deg(g) \implies g_{m+n-1} = 0.$$

Therefore,

$$(fg)_{m+n} = f_m g_n.$$

By similar argument, we may show that $(fg)_k = 0$ for all $k > m+n$. The result easily follows. ∎

Proposition 4.2. *Let $f, g \in \mathbb{F}[x]$ be nonzero and satisfy $f + g \neq 0$. Then* $\deg(f+g) \leq \max(\deg(f), \deg(g))$.

Lagrange Polynomials

(4.6) Polynomials as Functions

Polynomials is distinguished from general power series by their finitude. Because of this nature, one can assign some *semantics* to polynomials. In particular, they can be interpreted as functions, functions on the field $\mathbb{F}$ on which they are defined. First, fix

$$f = \sum^{d} f_i x^i \in \mathbb{F}[x].$$

Then one can obtain an induced function $f : \mathbb{F} \to \mathbb{F}$ by evaluation

$$a \mapsto f(a) = \sum^{d} f_i a^i,$$

where the sum here is an actual sum - it is not formal. In fact, we may also evaluate f on any algebra over $\mathbb{F}$, the same field which f is defined on.

(EX 4.7) Here are some examples. Fix $f \in \mathbb{Q}[x]$.

(a) Since $\mathbb{C}$ is an algebra over $\mathbb{Q}$, so for a polynomial f on $\mathbb{Q}$, it totally makes sense to evaulate $f:\mathbb{C}\to\mathbb{C}$ on complex numbers.

(b) We can even evaluate $f:M_{n\times n}(\mathbb{Q})\to M_{n\times n}(\mathbb{Q})$ on $M_{n\times n}(\mathbb{Q})$, since multiplication is well-defined in $M_{n\times n}(\mathbb{Q})$ with closure (in fact, for any field $\mathbb{F}$, $M_{n\times n}(\mathbb{F})$ is an algebra over $\mathbb{F}$).

(c) Fix V to be a vector space over $\mathbb{Q}$. Recall that $\mathcal{L}(V)$ is a set of linear operators on V. It is a easy verification that $\mathcal{L}(V)$ is an algebra over $\mathbb{Q}$, that it has well-behaved notion of addition, scalar multiplication, and multiplication (e.g. interprete multiplication as composition of operators). Therefore, f defines a function $f:\mathcal{L}(V)\to\mathcal{L}(V)$.

(4.8)

One should be careful when viewing polynomials as functions. That is, polynomials are not determined uniquely by their interpretations as functions. For instance, suppose that $\mathbb{F}=\mathbb{Z}/2\mathbb{Z}=\{0,1\}$ and define $f=x,g=x^2\in\mathbb{F}[x]$. It is clear that $f\neq g$ from the defintion of power series. But as functions of $\mathbb{F}$, they are the same! Namely, $f(0)=0,f(1)=1$ and $g(0)=0,g(1)=1$. But, a good news is that, when $\mathbb{F}$ is infinite, then polynomials on $\mathbb{F}$ are uniquely determined by their interpretations as functions.

Theorem 4.3.

If $\mathbb{F}$ is infinite and $f,g\in\mathbb{F}[x]$ which agree as functions on $\mathbb{F}$, then $f=g$.

But to prove Theorem 4.3, we need the following.

Theorem 4.4.
Lagrange Interpolation

Suppose that $t_0,t_1,\dots,t_n\in\mathbb{F}$ are $n+1$ distinct elements (that is, we are assuming that $\mathbb{F}$ has more than $n+1$ elements). For each $i\in\{0,\dots,n\}$, let

$$P_i=\prod_{j\neq i}\frac{x-t_j}{t_i-t_j}\in\mathbb{F}[x]. \tag{4.1}$$

Then for any $f\in\mathbb{F}[x]$ with $\deg(f)\leq n$ or $f=0\in\mathbb{F}[x]$, we may write

$$f=\sum^{n}f(t_i)P_i.$$

That is, any polynomial $f\in\mathbb{F}[x]$ with $\deg(f)\leq n$ (the case of zero polynomial is trivial), one can write f as a linear combination of $P_0,\dots,P_n$ (each with degree n), where the coefficients are given by the evaluations of f at n distinct elements of $\mathbb{F}$, $t_0,t_1,\dots,t_n\in\mathbb{F}$. In other words, $P_0,P_1,\dots,P_n$ form a basis of the vector space over $\mathbb{F}$ which the degree of polynomials are bounded by n.

Proof. Let

$$V=\{f\in\mathbb{F}[x]:f=0\vee\deg(f)\leq n\}.$$

We claim that V is a subspace of $\mathbb{F}[x]$, and this claim can be easily verified. Notice that V is not a *subalgebra* of $\mathbb{F}[x]$, as V is not closed under multiplication. One can easily find a basis for V, that $\{1,x,x^2,\dots,x^n\}$ is a basis for V, and $\dim(V)=n+1$. Now the claim is that, $P_0,\dots,P_n$ also form a basis for V. To verify this, we only need to show that $P_0,\dots,P_n$ are linearly independent, as we already have $n+1=\dim(V)$ elements of V. Now, write

$$\sum^{n}c_iP_i=0$$

for some $c_0,\dots,c_n\in\mathbb{F}$. Fix $k\in\{0,\dots,n\}$ and consider evaluating the above polynomial at t_k. Then we obtain

$$0=\sum^{n}c_iP_i(t_k)=\sum_i^n\prod_{j\neq i}\frac{t_k-t_j}{t_i-t_j}.$$

But it is evident from (4.1), the definition of P_i, that,

$$P_i(t_k) = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{otherwise} \end{cases}. \tag{4.2}$$

In other words,

$$0 = c_k,$$

or that $c_0, \ldots, c_n = 0$. Therefore $P_0, \ldots, P_n$ form a basis for V as claimed. Now, if $f \in V$, then

$$f = \sum^n a_i P_i.$$

Then by evaluating f at t_k gives

$$f(t_k) = \sum^n a_i P_i(t_k) = a_k$$

by (4.2), as desired. ∎

Proof of Theorem 4.3. Let $n = \max(\deg(f), \deg(g))$ and fix $n+1$ distinct elements $t_0, \ldots, t_n \in \mathbb{F}$. Then by the Lagrange interpolation,

$$f = \sum^n f(t_i) P_i$$

where

$$P_i = \prod_{j \neq i} \frac{x - t_i}{t_j - t_i} \in \mathbb{F}[x].$$

Similarly,

$$g = \sum^n g(t_i) P_i.$$

But f and g agree as functions on $\mathbb{F}$, we have $f(t_i) = g(t_i)$ for all i. Thus $f = g$, as desired. ∎

Division on Polynomials

Theorem 4.5. Division Algorithm

If $f, g \in \mathbb{F}[x]$ and $g \neq 0$, then there exists unique $q, r \in \mathbb{F}[x]$ such that

(a) $f = qg + r$; and

(b) $r = 0$ or $\deg(r) < \deg(g)$.

We call q the ***quotient*** and r the ***remainder***. If $r = 0$, then we say g ***divides*** f, and write $g \mid f$. To prove Theorem 4.5, we utilize the following lemma.

Lemma 4.5.1. *Let $f, g \in \mathbb{F}[x]$ be nonzero such that $\deg(g) \leq \deg(f)$. Then there exists $h \in \mathbb{F}[x]$ such that $f - gh = 0$ or $\deg(f - gh) < \deg(f)$.*

Proof. Let $m = \deg(f), n = \deg(g)$, and write

$$f = a_m x^m + \sum^{m-1} a_i x^i, g = b_n x^n + \sum^{n-1} b_i x^i.$$

Let $h = \frac{a_n}{b_n} x^{m-n}$. Then one sees that $\deg(f) > \deg(f - gh)$. ∎

Proof of Theorem 4.5. We verify the existence first. Here are some basic cases:

(a) If $f = 0$, take $q = 0, r = 0$.

(b) If $\deg(f) = 0$,

(i) if $\deg(g) = 0$, take $q = \frac{f}{g}, r = 0$; and

(ii) if $\deg(g) > 0 = \deg(f)$, take $q = 0, r = f$.

(c) If $\deg(f) > 0$ and $\deg(f) < \deg(g)$, take $q = 0, r = f$.

So we may assume that $\deg(f) \geq \deg(g)$. By Lemma 4.5.1, we obtain h such that $f - gh = 0$ or $\deg(f - gh) < \deg(f)$, so take $q = h$ and $r = f - gh$. (Prove uniqueness) ∎

Corollary 4.5.2.

Let $f \in \mathbb{F}[x]$ and let $c \in \mathbb{F}$. Then

$$(x - c) \mid f \iff f(c) = 0.$$

Corollary 4.5.3.

Let $f \in \mathbb{F}[x]$ and let $n = \deg(f)$. Then there exists at most n roots of f.

Ideal

Def'n. Ideal in $\mathbb{F}[x]$

An ***ideal*** in $\mathbb{F}[x]$ is a subspace $I \subseteq \mathbb{F}[x]$ such that

$$f \in I \wedge g \in \mathbb{F}[x] \implies fg \in I.$$

(EX 4.9) Examples of Ideal

Two easy examples of ideal in $\mathbb{F}[x]$ is $\{0\}$ and $\mathbb{F}[x]$ itself. Moreover, we observe the following: Let $I \subseteq \mathbb{F}$ be an ideal. Then $I = \mathbb{F}[x]$ if and only if $1 \in I$.

Proof. The forward direction is trivial. For the backward direction, observe that $1 \in I$ means that

$$\forall f \in \mathbb{F}[x] \, [f = 1f \in I].$$ ∎

Def'n. Principal Ideal Generated by a Polynomial

Let $g \in \mathbb{F}[x]$. The ***principal*** *ideal generated by* g is

$$(g) = g\mathbb{F}[x] = \{gf : f \in \mathbb{F}[x]\}.$$

In particular, we see that every ideal of $\mathbb{F}[x]$ is principal.

Theorem 4.6. Principal Ideal Theorem

If I is a nonzero ideal of $\mathbb{F}[x]$, then there exists unique monic polynomial $g \in \mathbb{F}[x]$ such that

$$I = (g).$$

Proof. Let $g \in I$ be nonzero and of minimal degree. Notice that

$$\frac{1}{l} g \in I,$$

where l is the leading coefficient of g, is of minimal degree and monic. So without loss of generality, suppose that g is monic. Of course

$$(g) \subseteq I.$$

To show the opposite direction, suppose $f \in I$. By division algorithm,

$$f = qg + r$$

for some $q, r \in \mathbb{F}[x]$ with $r = 0$ or $\deg(r) < \deg(g)$. But, observe that

$$r = f - qg,$$

where $f, qg \in I$, so $r \in I$. This exactly means that r cannot be of degree less than g, $r = 0$. Thus $f \in (g)$, as desired. to prove the uniqueness, suppose that there exists another monic $g' \in I$ that generates I. Then

$$g' = gh$$

for some $h \in \mathbb{F}$, and, similarly,

$$g = g'h'$$

for some $h' \in \mathbb{F}$. It follows that

$$g = (g'h') = ghh'.$$

So by cancellative property of $\mathbb{F}[x]$, $hh' = 1$, which means $\deg(h) = \deg(h') = 0$. But

$$g = g'h'.$$

where both g and g' are monic, which means $h' = 1$. Thus $g = g'$, as desired. ∎

Corollary 4.6.1.

Suppose that polynomials $p_1, \ldots, p_n \in \mathbb{F}[x]$ are not all zero. Then tehre is a unique monic polynomial $g \in \mathbb{F}[x]$ such that

(a) $g \in (p_1) + \cdots + (p_n) = \{f_1 + \cdots + f_n : \forall i \in \{1, \ldots, n\} [f_i \in (p_i)]\}$;

(b) for all $i \in \{1, \ldots, n\}$, $g \mid p_i$; and

(c) if $h \mid p$ for all $i \in \{1, \ldots, n\}$, then $h \mid g$.

In fact, if g satisfies (a) and (b), then it also satisfies (c).

Def'n. Gratest Common Divisor of Polynomials

Let $p_1, \ldots, p_n \in \mathbb{F}[x]$. Then the unique monic polynomial $g \in \mathbb{F}[x]$ which satisfies (a) and (b) of Corollary 1.6.1 is called the ***greatest common divisor*** of $p_1, \ldots, p_n$, denoted as

$$g = \gcd(p_1, \ldots, p_n).$$

Proof. Let

$$I = \sum_{i=1}^{n} (p_i),$$

then I is a nonzero ideal, since, given $f_1, \ldots, f_n, h \in \mathbb{F}[x]$, one has

$$\left(\sum_{i=1}^{n} f_i p_i\right) h = \sum_{i=1}^{n} f_i p_i h,$$

where each $f_i p_i h \in (p_i)$, and, in particular, some nonzero $p_i \in I$. By Theorem 4.6, $I = (g)$ for some monic $g \in \mathbb{F}[x]$. We claim that this g is the greatest common divisor of $p_1, \ldots, p_n$. Observe that (a) is already satisfied. To prove (b), observe that each $p_i \in I = (g)$, so $p_i = gh$ for some $h \in \mathbb{F}[x]$. But this exactly means that $g \mid p_i$. To show that (c) follows from (a) and (b), suppose that $h \in \mathbb{F}[x]$ is such that, for all $i \in \{1, \ldots, n\}$,

$$p_i = hf_i$$

for some $f_i \in \mathbb{F}[x]$. By (a),

$$g = \sum_{i=1}^{n} p_i g_i = \sum_{i=1}^{n} h f_i g_i = h \sum_{i=1}^{n} f_i g_i,$$

which means $h \mid g$. To show uniqueness, suppose there exists a monic $g' \in \mathbb{F}[x]$ which satisfies (a) and (b). Then by (c), $g \mid g'$ and $g' \mid g$. But both g and g' are monic, so $g = g'$, as desired. ∎

Polynomial Factorization

Def'n. Reducible, Irreducible Polynomial
Let $f \in \mathbb{F}[x]$ be nonconstant. We say f is ***reducible*** if there exists a nonconstant polynomials $g, h \in \mathbb{F}[x]$ such that $f = gh$. Otherwise, we say f is ***irreducible***.

(4.10) Degree 1 polynomials are irreducible. The following proposition shows an analogy between prime numbers and irreducible polynomials.

Proposition 4.7. *Let $p \in \mathbb{F}[x]$ be irreducible. Then, for any $f, g \in \mathbb{F}[x]$,*

$$p \mid fg \implies p \mid f \vee p \mid g.$$

Proof. Let $d = \gcd(p, f)$. Then $p = dh$ for some $h \in \mathbb{F}[x]$. if $\deg(d) \geq 1$, then $\deg(h) = 0$, since p is irreducible. Moreover, there exists $q \in \mathbb{F}[x]$ such that

$$f = dq = \left(\frac{1}{h} p\right) q = p\left(\frac{1}{h} q\right),$$

where $\frac{1}{h} \in \mathbb{F}[x]$, since $\deg(h) = 0$. Thus it follows that $p \mid f$. On the other hand, suppose that $\deg(d) = 0$. But this means that $d = 1$, since d is monic. Observe that $d \in (p) + (f)$, so we have

$$1 = pr_1 + fr_2$$

for some $r_1, r_2 \in \mathbb{F}[x]$. Then by multiplying both sides by g,

$$g = gpr_1 + gfr_2,$$

where $p \mid gf$, so it follows that $p \mid g$. ∎

Theorem 4.8. Unique Factorization of Polynomials

Every nonconstant monic polynomial in $\mathbb{F}[x]$ can be factored as a product of monic irreducible polynomials in $\mathbb{F}[x]$. Moreover, this factorization is unique up to reordering the factors.

Proof. Exercise. ∎

We present the following theorem without proof.

Theorem 4.9. Fundamental Theorem of Algebra

Every irreducible polynomial in $\mathbb{C}[x]$ have degree 1.

Corollary 4.9.1. Factorization of Polynoimials over the Complex Field

Let $f \in \mathbb{C}[x]$ be nonconstant. Then there exist $c, c_1, \ldots, c_n \in \mathbb{C}$ such that

$$f = c\prod_{i=1}^{n}(x-c_i).$$

Proof. Let c be the leading coefficient of f and $n = \deg(f)$. Then $f = cg$ for some monic $g \in \mathbb{C}[x]$, and by the unique factorization theorem (Theorem 4.8),

$$g = \prod_{i=1}^{n} p_i$$

for some irreducible $p_1, \ldots, p_n$. It follows from the fundamental theorem of algebra (Theorem 4.9) that each p_i is of degree 1, so one can find $c_1, \ldots, c_n$ such that

$$f = c\prod_{i=1}^{n}(x-c_i).$$ ∎

(4.11)

In fact, in

$$f = c\prod_{i=1}^{n}(x-c_i),$$

some c_i's can be repeated, and one can write

$$f = c\prod_{j=1}^{k}(x-c_j)^{r_j}$$

for some distinct $c_1, \ldots, c_k \in \mathbb{C}$ and $r_1, \ldots, r_k \in \mathbb{N}$. Note that $c_1, \ldots, c_k$ are distinct roots of f.

Def'n. Multiplicity of a Root of a Polynomial

Let $f \in \mathbb{F}[x]$ and let $c \in \mathbb{F}$ be a root of f. We define the ***multiplicity*** of c to be the largest positive integer r such that

$$(x-c)^r \mid f.$$

Proposition 4.10. Characterization of Multiplicity

Let $f \in \mathbb{F}[x]$ and let $c \in \mathbb{F}$. Then c is a root of f of multiplicity r if and only if

$$f = (x-c)^r g$$

for some $g \in \mathbb{F}[x]$ with $g(c) \neq 0$.

Proof. For the forward direction, suppose that the multiplicity of c is r. Then

$$f = (x-c)^r g.$$

For the sake of contradiction, suppose that $g(c) = 0$. Then there exists $h \in \mathbb{F}[x]$ such that

$$g = (x-c)h,$$

which contradicts the maximality of r. For the reverse directiono, suppose $f = (x-c)^r$ and $g(c) \neq 0$. Then

$$(x-c)^r \mid f.$$

For the sake of contradiction, suppose that

$$(x-c)^k \mid f$$

for some $k > r$. Then

$$f = (x-c)^k h,$$

so by cancellation, $g = (x-c)^{k-r} h$. But $k - r > 0$, which contradicts the fact $g(c) \neq 0$. ∎

(4.12) It follows from Proposition 4.10 that, if

$$f = c\prod_{i=1}^{l} (x-c_i)^{r_i},$$

where $c_1, \ldots, c_l \in \mathbb{F}$ are distinct and $r_1, \ldots, r_l \in \mathbb{N}$. Then r_i is the multiplicity of c_i, since

$$f = (x-c_i)^{r_i} g,$$

where

$$g = c\prod_{j \neq i} (x-c_j)^{r_j},$$

so $g(c_i) \neq 0$.

Def'n. Formal Derivative of a Polynomial
Given $f = \sum_{i=0}^{n} a_i x^i$, then we define the ***formal derivative*** of f to be

$$Df = \sum_{i=1}^{n} i a_i x^{i-1}.$$

Theorem 4.11. *Let $\mathbb{F}$ be a field of characteristic 0, let $f \in \mathbb{F}[x]$ be nonzero, and let $c \in \mathbb{F}$. Then c is a root of f with multiplicity r if the following conditions hold:*

(a) $D^k f(c) = 0$ if $0 \leq k \leq r-1$; and

(b) $D^r f(c) \neq 0$,

where $D : \mathbb{F}[x] \to \mathbb{F}[x]$ is the formal derivative operator.

This page intentionally left blank.

6.
Elementary Canonical Forms

Direct Sum Decompoisitions

(6.1) Throughout this chapter, fix an arbitrary field $\mathbb{F}$, and whenever a vector field is introduced without mentioning underlying field, it is understood that the field is $\mathbb{F}$. Recall that, when we have a vector space W over $\mathbb{F}$ and subspaces $W_1, \dots, W_k \subseteq W$, the sum of the subspaces is defined as

$$\sum_{i=1}^{k} W_i = \left\{ \sum_{i=1}^{k} w_i : \forall i \in \{1, \dots, k\} \, [w_i \in W_i] \right\},$$

which is also a subspace of W. We introduce the notion of independence of subspaces.

Def'n. Independent Subspace
Let W be a vector space and let $W_1, \dots, W_k \subset W$ be subspaces. We call $W_1, \dots, W_k$ is ***independent*** if whenever $w_1 \in W_1, \dots, w_k \in W_k$ are such that

$$\sum_{i=1}^{k} w_i = 0,$$

then each $w_i = 0$.

(EX 6.2) Given subspaces $W_1, W_2 \subseteq W$, W_1, W_2 are independent if and only if $W_1 \cap W_2 = \{0\}$.

Proof. For the forward direction, let $w \in W_1 \cap W_2$. Then we have $-w \in W_2$ such that

$$w + (-w) = 0.$$

Thus by independence, $w = -w = 0$. For the reverse direction, let $w_1 \in W_1, w_2 \in W_2$ be such that

$$w_1 + w_2 = 0.$$

This implies $w_1 = -w_2 \in W_1 \cap W_2 = \{0\}$. Thus $w_1 = w_2 = 0$. ∎

But when we have more than two subspaces $W_1, \dots, W_k$, $k > 2$, the condition $\bigcap_{i=1}^{k} W_i = \{0\}$ is not strong enough to imply the independence of subspaces. Instead, we have the following proposition.

Proposition 6.1. *Let V be a vector space and let $W_1, \dots, W_k \subseteq W$ be finite subspaces. Let*

$$W = \sum_{i=1}^{k} W_i.$$

Then the following are equivalent.

(a) $W_1, \dots, W_k$ are independent.

(b) For each $j \geq 2$,

$$\left(\bigcap_{i=1}^{j-1} W_i \right) \cap W_j = \{0\}.$$

(c) If β_i is a basis for W_i for all $i \in \{1, \dots, k\}$, then

$$\beta = (\beta_1, \dots, \beta_k)$$

is a basis for W.

(d) For all $w \in W$, there exists unique $w_1 \in W_1, \ldots, w_k \in W_k$ such that

$$w = \sum_{i=1}^{k} w_i.$$

Proof. We assume $k \geq 2$ for convenience. We proceed in a cycle.
(a) $\implies$ (b): Exercise.
(b) $\implies$ (c): Fix a basis β_i for each W_i and write

$$\beta_i = \{v_{i1}, \ldots, v_{il_i}\}.$$

Clearly

$$\beta = (\beta_1, \ldots, \beta_k)$$

spans W by definition. To verify that β is linearly independent, suppose that we have a linear combination in β

$$\sum_{i=1}^{k} \sum_{j}^{l_i} a_{ij} v_{ij} = 0$$

for some $a_{ij} \in \mathbb{F}$. For convenience, write

$$w_i = \sum_{j}^{l_i} a_{ij} v_{ij}.$$

Then it sufflices to show that $w_i = 0$ for all $i \in \{1, \ldots, k\}$. For the sake of contradiction, suppose not. Let j be the largest index such that $w_j \neq 0$. If $j = 1$, then we have a contradiction immediately, as $w_2 = \cdots = w_k = 0$. So we may assume that $j \geq 2$ without loss of generality. Then

$$w_{j+1} = \cdots = w_k = 0,$$

which means

$$\sum_{i=1}^{j} w_i = 0.$$

In other words,

$$w_j = -\sum_{i=1}^{j-1} w_i.$$

But $w_j \in W_j$ and $-\sum_{i=1}^{j-1} w_i \in \sum_{i=1}^{j-1} W_i$, which means

$$0 \neq w_j \in \left(\sum_{i=1}^{j-1} W_i \right) \cap W_j,$$

which contradicts (b).
(c) $\implies$ (d): Fix a basis $\beta_i = \{v_{i1}, \ldots, v_{il_i}\}$ for all W_i. Let $w \in W$ and let $w_1, w_1' \in W_1, \ldots, w_k, w_k' \in W_k$ be such that

$$w = \sum_{i=1}^{k} w_i = \sum_{i=1}^{k} w_i'.$$

Then $w_i - w_i' \in W_i$, so

$$w_i - w_i' = \sum_{j=1}^{l_i} a_{ij} v_{ij}$$

for some $a_{ij} \in \mathbb{F}$. So one can write

$$\sum_{i=1}^{k} \sum_{j=1}^{l_i} a_{ij} v_{ij} = 0,$$

where (c) implies that each $a_{ij} = 0$. Thus $w_i = w_i'$ for all $i \in \{1, \ldots, k\}$.
(d) $\implies$ (a): Since (d) guarantees unique representation of every $w \in W$, we have unique representation of $0 \in W$. Moreover

$$\sum_{i=1}^{k} 0 = 0$$

is the trivial representation of 0, and (a) follows. ∎

Def'n. Direct Sum of Subspaces

Suppose that $W_1, \ldots, W_k \subseteq W$ are subspaces of a finite dimensional vector space W. We say W is a ***direct sum*** of $W_1, \ldots, W_k$ if

(a) $W = \sum_{i=1}^{k} W_i$; and

(b) $W_1, \ldots, W_k$ are independent.

We denote this by

$$W = \bigoplus_{i=1}^{k} W_i.$$

(6.3) Direct Sum Decomposition

By Proposition 6.1,

$$W = \bigoplus_{i=1}^{k} W_i$$

if and only if every vector in W can be written uniquely as a sum of vectors from $W_1, \ldots, W_k$. We sometimes call

$$W = \bigoplus_{i=1}^{k} W_i$$

a *direct sum decomposition* of W.

(EX 6.4) Let V be a finite dimensional vector space with basis $\beta = \{v_1, \ldots, v_n\}$. Then

$$V = \bigoplus_{i=1}^{k} \operatorname{span}(v_i).$$

(EX 6.5) Let V be a finite dimensional vector space, let $T : V \to V$ be a linear operator, and let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues. Let W_i be the eigenspace associated to λ_i,

$$W_i = \ker(T - \lambda_i I).$$

Then $W_1, \ldots, W_k$ are independent. Moreover, T is diagonalizable if and only if

$$V = \bigoplus_{i=1}^{k} W_i = \bigoplus_{i=1}^{k} \ker(T - \lambda_i I).$$

Proof. Exercise. ∎

(EX 6.6) Let V be a vector space and let $W_1, \ldots, W_n \subseteq V$ be finite dimensional subspaces. Then $W_1, \ldots, W_n$ are independent if and only if whenever we have linearly independent $\beta_i \subseteq W_i$ for all $i \in \{1, \ldots, n\}$, then

$$\beta = (\beta_1, \ldots, \beta_n)$$

is linearly independent in V.

Proof. Exercise. ∎

Projections

(6.7) To motivate the upcoming definition, suppose

$$W = W_1 \oplus W_2.$$

Then we may define a function $E : W \to W$ as follows. Given $w \in W$, we can uniquely write w as

$$w = w_1 + w_2$$

for some $w_1 \in W_1, w_2 \in W_2$. By uniqueness, defining E by the mapping

$$w \mapsto w_1$$

is well defined. Moreover, we claim that E is linear.

Proof. Let $w = w_1 + w_2, v = v_1 + v_2 \in W$ and let $c \in \mathbb{F}$. Then

$$E(cv + w) = E((cw_1 + v_1) + (cw_2 + vw)) = cw_1 + v_1 = cE(w) + E(v). \qquad \blacksquare$$

Another property of E is that it is *idempotent*, $E^2 = E$.

Proof. Let $w = w_1 + w_2 \in W$ for some $w_1 \in W_1, w_2 \in W_2$. Then

$$E^2(w) = E(E(w)) = E(w_1) = E(w_1 + 0) = w_1 = E(w). \qquad \blacksquare$$

Def'n. Projection on a Vector Space
A ***projection*** on a vector space V is a linear operator $T : V \to V$ that is idempotent, $T^2 = T$.

Proposition 6.2. *Let V be a vector space and let $E : V \to V$ to be a projection. Then*

$$V = \text{image}(E) \oplus \ker(E).$$

Proof. Let $v \in V$. Then

$$v = E(v) + (v - E(v)),$$

where $E(v) \in \text{image}(E)$ and $E(v - E(v)) = E(v) - E^2(v) = 0$, so $(v - E(v)) \in \ker(E)$. So we have

$$V = R + N.$$

To verify that R and N are independent, suppose $v \in R \cap N$. Then $E(v) = 0$ since $v \in N$, and $v = E(w)$ for some $w \in V$ since $v \in R$. Thus

$$0 = E(v) = E^2(w) = E(w) = v,$$

as desired. $\blacksquare$

(6.8) If E is a projection and $v \in \text{image}(E)$, then $E(v) = v$. Now we turn our attention to when we have any number of subspaces which direct sum is the whole space.

Theorem 6.3.

Let V be a finite dimensional vector space and let $W_1, \ldots, W_n \subseteq V$ be such that

$$V = \bigoplus_{i=1}^{n} W_i.$$

Then there exist projections $E_1, \ldots, E_n : V \to V$ satisfying:

(a) for all $i \in \{1, \ldots, n\}$, $W_i = \text{image}(E_i)$;

(b) if $i \neq j$, then $E_i E_j = 0$; and

(c) $\sum_{i=1}^{n} = I$, the identity.

Conversely, given projections

$$E_1, \ldots, E_n : V \to V$$

such that (b) and (c) hold, then

$$V = \bigoplus_{i=1}^{n} \text{image}(E_i).$$

Proof. Suppose that

$$V = \bigoplus_{i=1}^{n} W_i.$$

Given $v \in V$ with the unique representation

$$v = \sum_{i=1}^{n} w_i,$$

where $w_i \in W_i$ for all $i \in \{1, \ldots, n\}$. Define $E_i : V \to V$ by the mapping

$$v \mapsto w_i$$

for each $i \in \{1, \ldots, n\}$. Then each E_i is a projection (verify this). To show (a), it is clear that $\text{image}(E_i) \subseteq W_i$. Conversely, suppose $v \in W_i$. Then

$$v = 0 + \cdots + 0 + v + 0 + \cdots + 0$$

which means $v = E_i(v) \in \text{image}(E_i)$. To verify (b), suppose $i \neq j$ and write unique representation

$$v = \sum_{i=1}^{n} w_i.$$

Then

$$E_i E_j(v) = E_i(w_j) = E_i(0 + \cdots + 0 + w_j + 0 + \cdots + 0) = 0.$$

To verify (c), write unique representation

$$v = \sum_{i=1}^{n} w_i.$$

Then

$$\left(\sum_{i=1}^{n} E_i\right)(v) = \sum_{i=1}^{n} E_i(v) = \sum_{i=1}^{n} w_i = v = I(v).$$

For the converse statement, suppose that we have projections

$$E_1, \ldots, E_n : V \to V$$

such that (b) and (c) hold. Fix $v \in V$. Then

$$v = I(v) = \left(\sum_{i=1}^{n} E_i\right)(v) = \sum_{i=1}^{n} E(v) \in \sum_{i=1}^{n} \text{image}(E_i).$$

This shows that

$$V = \sum_{i=1}^{n} \text{image}(E_i).$$

To show that the above sum is direct, we show that

$$v = \sum_{i=1}^{n} E(v) \qquad [6.1]$$

is the unique representation of v as a sum of elements of $\text{image}(E_1), \ldots, \text{image}(E_n)$. Suppose that

$$v = \sum_{i=1}^{n} w_i$$

for some $w_1 \in \text{image}(E_1), \ldots, w_n \in \text{image}(E_n)$. Then

$$w_i = E_i(v_i)$$

for some $v_i \in V$, and

$$v = \sum_{i=1}^{n} E_i(v_i).$$

Now, fix $j \in \{1, \ldots, n\}$. Then,

$$E_j(v) = E_j \sum_{i=1}^{n} E_i(v_i) = \sum_{i=1}^{n} E_j E_i(v_i) = E_j E_j(v_j) = E_j(v_j) = w_j,$$

which shows that [2.1] is the unique representation of v, and the result follows. ∎

Invariant Decompositions

Def'n. T**-invariant** Subspace
Let $T : V \to V$ be a linear operator and let $W \subseteq V$ be subspace. Then we say W is ***T*-invariant** if $T(W) \subseteq W$.

(6.9)
T-invariant Decomposition

By a T-invariant direct sum decomposition of V, we mean a direct sum decomposition

$$V = \bigoplus_{i=1}^{n} W_i$$

such that $W_1, \ldots, W_n \subseteq V$ are T-invariant. Note that if we let

$$T_i = T|_{W_i} : W_i \to W_i$$

for each $i \in \{1, \ldots, n\}$ to be the restriction operator of T on W_i and, given $v \in V$, write

$$v = \sum_{i=1}^{n} w_i$$

with each $w_i \in W_i$, then

$$T(v) = \sum_{i=1}^{n} T_i(w_i).$$

Def'n. Direct Sum of Operators
Consider the case in (6.9). We say that T is the ***direct sum*** of $T_1, \ldots, T_n$.

The idea is that infromation about $T_1, \ldots, T_n$ will provide us information about T. We may also discuss this idea in terms of matices. Suppose that V is finite dimensional and fix basis β_i for W_i for all $i \in \{1, \ldots, n\}$. Then

$$\beta = (\beta_1, \ldots, \beta_n)$$

is a basis for V. Then we claim that

$$[T]_\beta = \begin{bmatrix} [T_1]_{\beta_1} & & & \\ & [T_2]_{\beta_2} & & \\ & & \ddots & \\ & & & [T_n]_{\beta_n} \end{bmatrix},$$

a block diagonal matrix.

Proof. Exercise. ∎

Now we proceed to characterize T-invariant decompositions in terms of projections.

Proposition 6.4. Characterization of T-invariant Decomposition by Commutativity

Let $T : V \to V$ be a linear operator and suppose

$$V = \bigoplus_{i=1}^{n} W_i \tag{6.2}$$

is a direct sum decomposition with corresponding projections $E_1 : V \to V, \ldots, E_n : V \to V$. Then [2.2] is a T-invariant decomposition if and only if T commutes with E_i,

$$TE_i = E_i T,$$

for all $i \in \{1, \ldots, n\}$.

Proof. Recall that $W_i = \text{image}(E_i)$, and if we restrict E_i to its range, then

$$E_i|_{W_i} = I_{W_i} : W_i \to W_i,$$

the identity on W_i. For the reverse direction, suppose that T commutes with each E_i. Fix $i \in \{1, \ldots, n\}$ and let $w \in W_i$. Then $W = E_i(w)$, so

$$T(w) = T(E_i(w)) = E_i T(w) \in \text{image}(E_i) = W_i.$$

Thus W_i is T-invariant. For the forward direction, suppose that [2.2] is a T-invariant decomposition. Let $v \in V$ and write its unique representation

$$v = \sum_{i=1}^{n} E_i(v).$$

Then

$$T(v) = T \sum_{i=1}^{n} E_i(v) = \sum_{i=1}^{n} TE_i(v),$$

where each $TE_i(v) \in W_i$ by T-invariance. Fix $i \in \{1, \ldots, n\}$. Then

$$E_j T(v) = \sum_{i=1}^{n} E_j TE_i(v) = E_j TE_j(v) = TE_j(v),$$

v since $E_j(w_i) = 0$ whenever $w_i \in W_i$ with $i \neq j$, and E_j is an identity on W_j. ∎

Theorem 6.5. Characterization of Diagonalizability

Let $T : V \to V$ be a linear operator on a finite dimensional vector space V.

(a) If T is diagonalizable with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, then there is a T-invariant decomposition

$$V = \bigoplus_{i=1}^{k} W_i$$

with corresponding projections $E_1, \ldots, E_n$ such that

(i) $T = \sum_{i=1}^{k} \lambda_i E_i$; and

(ii) $W_i = \ker(\lambda_i I - T)$, the eigenspace corresponding to λ_i.

(b) Suppose $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ are distinct scalars and $E_1, \ldots, E_k : V \to V$ are nonzero linear operators on V such that

(i) $T = \sum_{i=1}^{k} \lambda_i E_i$;

(ii) $I = \sum_{i=1}^{k} E_i$; and

(iii) $E_i E_j = 0$ for all $i \neq j$.

Then T is diagonalizable, $\lambda_1, \ldots, \lambda_k$ are the only eigenvalues of T, $E_1, \ldots, E_k$ are projections, and each image(E_i) *is the eigenspace corresponding to λ_i.*

Proof. (ii) of (a) is left as an exercise. To prove (i) of (a), Let $v \in V$ and write

$$v = \sum_{i=1}^{k} E_i(v),$$

then

$$T(v) = \sum_{i=1}^{k} TE_i(v) = \sum_{i=1}^{k} \lambda_i E_i(v).$$

Thus it follows that $T = \sum_{i=1}^{k} E_i$. For (b), we make several claims. First, we claim that each E_i is a projection. To verify this, observe that

$$I = \sum_{i=1}^{k} E_i,$$

so by multiplying E_j to both sides,

$$E_j = \sum_{i=1}^{k} E_j E_i = E_j E_j,$$

which means $E_1, \ldots, E_k$ are projections. Now, by Theorem 6.3, we also obtain

$$V = \bigoplus_{i=1}^{k} \text{image}(E_i). \qquad [6.3]$$

The next claim that each λ_i is an eigenvalue of T and image$(E_i) \subseteq \ker(\lambda_i I - T)$, the eigenspace corresponding to λ_i. To verify this, notice that

$$T = \sum_{i=1}^{k} \lambda_i E_i,$$

so

$$TE_j = \sum_{i=1}^{k} \lambda_i E_i E_j = \lambda_j E_j^2 = \lambda_j E_j.$$

Now let $v \in$ image(E_j), or, $v = E_j(w)$ for some $w \in V$. Then

$$T(v) = TE_j(w) = \lambda_j E_j(w) = \lambda_j v.$$

Since $E_j \neq 0$, image$(E_j) \neq \{0\}$, so there exists nonzero eigenvalue v corresponding to λ_j in image(E_j). The third claim is that T is diagonalizable. This immediately follows from [2.3] and the second claim, since they together implies that every vector in V can be written as a sum of eigenvectors. Next, we verify that $\lambda_1, \ldots, \lambda_k$ are the only eigenvalues of T. Let $\lambda \in \mathbb{F}$ and $v \in V$ be nonzero such that $T(v) = \lambda v$. Then

$$(\lambda I - T)(v) = 0.$$

But

$$T = \sum_{i=1}^{k} \lambda_i E_i$$

and

$$\lambda I = \lambda \sum_{i=1}^{k} E_i = \sum_{i=1}^{k} \lambda E_i.$$

Thus

$$T - \lambda I = \sum_{i=1}^{k} (\lambda_i - \lambda) E_i(v) = 0.$$

Then by [2.3], each

$$(\lambda_i - \lambda) E_i(v) = 0.$$

But

$$v = \sum_{i=1}^{k} E_i(v) \neq 0,$$

so some $E_i(v) \neq 0$. So it concludes that $\lambda_i = \lambda$, and such index i is unique, since $\lambda_1, \ldots, \lambda_k$ are distinct. Lastly, we claim that image$(E_j) \supseteq \ker(\lambda_i I - T)$. To verify this, suppose $v \in V$ is such that

$$T(v) = \lambda_j v$$

for some λ_j. Then

$$0 = (T - \lambda_j I)(v) = \sum_{i=1}^{k} (\lambda_i - \lambda_j) E_j(v).$$

So, again, by [2.3], each $(\lambda_i - \lambda_j) E_i(v) = 0$. Therefore, if $j \neq i$, then $E_i(v) = 0$. But

$$v = Iv = \sum_{i=1}^{k} E_i(v) = E_j(v) \in \text{image}(E_i),$$

as required. ∎

(6.10) Here is the setting for the remaining part of the chapter. Let V be a finite dimensional vector space over $\mathbb{F}$ and let $T : V \to V$ be a linear operator. Given this, we know that there is an ideal of annihilating polynomials of T,

$$I_T = \{f \in \mathbb{F}[x] : f(T) = 0\}.$$

Then by the principal ideal theorem, there exists a unique monic polynomial $p \in \mathbb{F}[x]$ which generates I_T. In fact, p is the monic polynomial of minimal degree in I_T, or, p is the polynomial of minimal degree which annihilates T.

> **Def'n. Minimal Polynomial** of a Linear Operator
> Let V be a vector space and let $T : V \to V$ be linear. We say the unique monic generator of the ideal of polynomials which annihilate T the ***minimal polynomial*** of T. Equivalently, the minimal polynomial of T is the polynomial of minimal degree which annihilates T.

Given this, what we now desire to prove is the *primary decomposition theorem*.

Theorem 6.6. Primary Decomposition Theorem

Let $T : V \to V$ be a linear operator on a finite dimensional vector space V over $\mathbb{F}$ and let $p \in \mathbb{F}[x]$ be the minimal polynomial of T. Write p as

$$p = \prod_{i=1}^{k} p_i^{r_i}$$

for some distinct monic irreducibles $p_1, \ldots, p_k \in \mathbb{F}[x]$ and for some positive integers $r_1, \ldots, r_k \in \mathbb{N}$. Let

$$W_i = \ker\left(p_i^{r_i}(T)\right)$$

for each $i \in \{1, \ldots, k\}$. Then

$$V = \bigoplus_{i=1}^{k} W_i$$

is a T-invariant decomposition of V and the minimal polynomial of the restriction of T onto W_i,

$$T_i = T\,|_{W_i} \colon W_i \to W_i$$

is $p_i^{r_i}$ for all $i \in \{1, \ldots, k\}$.

(6.11)

In connection to diagonalizable linear operator $T : V \to V$, what the primary decomposition theorem says is that,

$$p = \prod_{i=1}^{k} x - \lambda_i$$

is the prime decomposition of p where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues. We also obtain each W_i as the eigenspace corresponding to λ_i.

Proof of Theorem 6.6. For each $i \in \{1, \ldots, k\}$, let

$$f_i = \prod_{j \neq i} p_j^{r_j}.$$

The point is that, we desire to make $f_1, \ldots, f_k$ such that

$$p_i^{r_i} f_i = p$$

for all $i \in \{1, \ldots, k\}$. Then one can easily verify that $f_1, \ldots, f_k$ are coprime (prove as exercise). This means that

$$\sum_{i=1}^{k} \langle f_i \rangle = \mathbb{F}[x].$$

In particular, there exists $g_1, \ldots, g_k \in \mathbb{F}[x]$ such that

$$1 = \sum_{i=1}^{k} g_i f_i.$$

For each $i \in \{1, \ldots, k\}$, let

$$h_i = g_i f_i$$

and $E_i = h_i(T) : V \to V$. We now claim that each E_i is a projection. To verify this, observe the following.

(a) $\sum_{i=1}^{k} E_i = I$.

Proof of (a). Observe that

$$\sum_{i=1}^{k} E_i = \sum_{i=1}^{k} h_i(T) = \left(\sum_{i=1}^{k} f_i g_i\right)(T) = 1(T) = I.$$

(b) $E_iE_j = 0$ whenever $i \neq j$.

Proof of (b). Observe that

$$E_iE_j = h_ih_j(T) = f_ig_if_jg_j(T) = \prod_{l\neq i} p_l^{r_l} g_i \prod_{l\neq j} p_l^{r_l} g_j(T),$$

where $p \mid \prod_{l\neq i} p_l^{r_l} \prod_{l\neq j} p_l^{r_l}$. So it follows that $p \mid h_ih_j$ and so $E_iE_j = 0$.

By (a) and (b), we have $E_1, \ldots, E_k$ are projections. So we obtain a direct sum decomposition

$$V = \bigoplus_{i=1}^{k} \text{image}(E_i).$$

We also have few more claims.

(c) Each $W_i = \text{image}(E_i)$.

Proof of (c). Let $v \in \text{image}(E_i)$, then $E_i(v) = v$, and so one has

$$p_i^{r_i}(T)(v) = p_i^{r_i}(T)E_i(v) = p_i^{r_i} f_ig_i(T)(v) = g_ip(T)(v) = 0,$$

so $\text{image}(E_i) \subseteq W_i$. Conversely, let $v \in W_i$, $p_i^{r_i}(T)(v) = 0$. If $j \neq i$, then $p_i^{r_i} \mid f_jg_j$, so $f_j(T)g_j(T)(v) = 0$ for all $j \neq i$. Therefore,

$$v \in \ker(E_j)$$

for all $j \neq i$. But this means that

$$v = Iv = \sum_{j=1}^{k} E_j(v) = E_i(v) \in \text{image}(E_i).$$

It follows from (c) that we obtain a direct sum decomposition

$$V = \bigoplus_{i=1}^{k} \text{image}(E_i) = \bigoplus_{i=1}^{k} W_i. \qquad [6.4]$$

To show that [2.4] is indeed a T-invariant direct sum, one only has to notice that each E_i is a polynomial in T, since any polynomials in T commute with T (in fact, any polynomials of T commute with each other). The last claim is as follows.

(d) $p_i^{r_i}$ is the minimal polynomial of $T_i = T\,|_{W_i}$.

Proof of (d). By definition,

$$p_i^{r_i}(T_i) = 0,$$

since $W_i = \ker(p_i^{r_i}(T))$. Let $q_i \in \mathbb{F}[x]$ be the minimal polynomial of T_i. Then $q_i \mid p_i^{r_i}$. But both p_i and q_i are monic and p_i is an irreducible, so it follows that $q_i = p_i^{s_i}$ for some $s_i \leq r_i$. Consider

$$q = \prod_{i=1}^{k} q_i \in \mathbb{F}[x]$$

and fix $v \in V$, to compute $q(T)(v)$. By [2.4], we have $w_1 \in W_1, \ldots, w_k \in W_k$ such that

$$v = \sum_{i=1}^{k} w_i.$$

Then

$$q(T)(v) = \sum_{i=1}^{k} q(T)(w_i) = \sum_{i=1}^{k} q(T_i)(w_i) = 0,$$

where the last equality holds by the fact that q_i annihilates T_i. Therefore, q annihilates T and so $p \mid q$. But

$$q = \prod_{i=1}^{k} p_i^{s_i}$$

for some $s_1 \leq r_1, \ldots, s_k \leq r_k$, so the only case which

$$\prod_{i=1}^{k} p_i^{r_i} \mid \prod_{i=1}^{k} p_i^{s_i}$$

is when $r_i = s_i$ for all $i \in \{1, \ldots, k\}$. Thus $q_i = p_i$ is the minimal polynomial of T_i for all $i \in \{1, \ldots, k\}$. ∎

Diagonalizable and Nilpotent Parts

(6.12) Diagonalizable and Nilpotent Parts

By the primary decomposition theorem, we can obtain a T-invariant direct sum decomposition

$$V = \bigoplus_{i=1}^{k} W_i$$

given a linear operator $T : V \to V$ on a finite dimensional vector space V. Moreover, we obtain a polynomial

$$p = \prod_{i=1}^{k} p_i^{r_i} \in \mathbb{F}[x]$$

for some irreducible monic $p_1, \ldots, p_k \in \mathbb{F}[x]$ and for some positive integers $r_1, \ldots, r_k \in \mathbb{N}$, such that

(a) p is the minimal polynomial of T;

(b) each $W_i = \ker(p_i^{r_i}(T))$; and

(c) we have corresponding projections $E_1, \ldots, E_k$, which are polynomials of T.

We now consider a special case. Suppose each p_i is linear, say $p_i = x - \lambda_i$ for some $\lambda_i \in \mathbb{F}$. For instance, if $\mathbb{F} = \mathbb{C}$, this is always the case by the fundamental theorem of algebra. Then one obtains

(a) $p = \prod_{i=1}^{k} (x - \lambda_i)^{r_i}$;

(b) $W_i = \ker((T - \lambda_i I)^{r_i})$;

(c) each E_i is a projection with image$(E_i) = W_i$;

(d) $I = \sum_{i=1}^{k} E_i$; and

(e) $E_i E_j = 0$ whenever $i \neq j$.

However, we do not know that

$$T = \sum_{i=1}^{k} \lambda_i E_i, \tag{6.5}$$

as [4.5] would imply that T is diagonalizable. So intead, we define

$$D = \sum_{i=1}^{k} \lambda_i E_i : V \to V,$$

which is a diagonalizable linear operator with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ on V (but not necessarily $T = D$). We call this D the diagonalizable part of T.

Def'n. Diagonalizable Part of a Linear Operator
Let $T : V \to V$ be such that its minimal polynomial p can be written as a product of linear polynomials over $\mathbb{F}$. Then one can define a diagonalizable linear operator $D : V \to V$ as described in (6.12), which we call the ***diagonalizable part*** of T.

Now, define $N = T - D : V \to V$. What are some important properties of N? Since

$$T = TI = T\sum_{i=1}^{k} E_i = \sum_{i=1}^{k} TE_i,$$

we have

$$N = T - D = \sum_{i=1}^{k} TE_i - \sum_{i=1}^{k} \lambda_i E_I = \sum_{i=1}^{k} (T - \lambda_i I) E_i.$$

Now, observe that

$$N^2 = \sum_{i,j=1}^{k} (T - \lambda_i I) E_i (T - \lambda_j I) E_j.$$

But E_i, E_j are polynomials in T, they commute with T, so whenever $i \neq j$, we have

$$(T - \lambda_i I) E_i (T - \lambda_j I) E_j = 0.$$

So,

$$N^2 = \sum_{i=1}^{k} (T - \lambda_i I)^2 E_i.$$

By an induction argument,

$$N^l = \sum_{i=1}^{k} (T - \lambda_i I)^l E_i.$$

Now, let $r = \max(r_1, \ldots, r_k)$. Then one can write

$$(T - \lambda_i I)^r E_i = (T - \lambda_i I)^{r - r_i} (T - \lambda_i I)^{r_i} = 0,$$

since $\text{image}(E_i) = W_i = \ker((T_i - \lambda_i I)^{r_i})$. Thus

$$N^r = 0,$$

and so we call N the nilpotent part of T.

Def'n. Nilpotent Part of a Linear Operator
Let $T : V \to V$ be such that its minimal polynomial p can be written as a product of a linear polynomials over $\mathbb{F}$. Then we define the ***nilpotent part*** $N : V \to V$ of T by $N = T - D$, where D is the diagonalizable part of T.

We prove the results so far more formally.

Theorem 6.7. Diagonalizable and Nilpotent Parts

Let $T : V \to V$ be a linear operator on a finite dimensional vector space V such that the minimal polynomial $p \in \mathbb{F}[x]$ of T can be written as a product of linear polynomial polynomials. Then there exists unique linear operators $D, N : V \to V$ such that

(a) D is diagonalizable;

(b) N is nilpotent;

(c) $T = D + N$; and

(d) $DN = ND$.

Proof. In (6.12), we constructed D, N with the mentioned properties (a), (b), (c). For (d), observe that D, N we constructed are polynomials in T. To show uniqueness, we utilize *simultaneous diagonalization* - diagonalizing linear operators simultaneously by the common basis.

Lemma 6.7.1. Simultaneously Diagonalizable Operators

Let $T, S : V \to V$ be diagonalizable linear operators. Then T, S are simultaneously diagonalizable if and only if T, S commute.

Proof. Forward direction is straightforward, as any diagonal matrices commute. For the reverse direction, suppose that T, S commute (prove as an exercise). ∎

Proof of Theorem 6.7. We utilize the fact that the only diagonalizable nilpotent linear operator is $0 : V \to V$. That is, given $D', N' : V \to V$ which satisfies (a), (b), (c), (d), we have that

$$D + N = T = D' + N' \implies D - D' = N' - N.$$

But this means that $D - D'$ is a nilpotent operator, so it follows that $D - D' = 0$ and so $N' - N = 0$, as desired. ∎

This page intentionally left blank.

7.
Rational and Jordan Form

Cyclic Subspaces

(7.1)

We begin by introducing a very natural notion, a cyclic subspace generated by a vector.

Def'n. Cyclic Subspace
Let V be a vector space and let $T : V \to V$ be linear. Fix $v \in V$. We define the ***T*-cyclic** subspace generated by v by

$$Z(v;T) = \{g(T)(v) : g \in \mathbb{F}[x]\}.$$

The significance of this definition can be shown as follows.

Proposition 7.1. Properties of Cyclic Subspacec

Let $T : V \to V$ be a linear operator on a vector space V and let $v \in V$. Then

(a) $Z(v;T)$ is the smallest T-invariant subpsace of V that contains V;

(b) $Z(v;T) = \operatorname{span}\{T^i(v) : i \geq 0\}$.

Proof. Observe the following.

(a) Notice that $Z(v;T)$ is a T-invariant subspace. That is,

$$cg(T)(v) + h(T)(v) = (cg+h)(T)(v),$$

and

$$T(g(T)(v)) = T(g(T))(v),$$

where $T(g(T))$ is again a polynomial in T. To show the minimality, suppose that $W \subseteq V$ is a T-invariant subspace with $v \in W$. Let $u \in Z(v;T)$, so write

$$u = g(T)(v) = \sum_{i=0}^{n} a_i T^i(v)$$

for some $a_0, \ldots, a_n \in \mathbb{F}$. But each $a_i T^i(v) \in W$, so it follows that $u \in W$.

(b) This is straightforward, as $\mathbb{F}[x]$ is generated by $\{x^0, x^1, \ldots\}$. ∎

(EX 7.2) Examples of Cyclic Subspaces

Let $T : V \to V$ be linear on a vector space V.

(a) $Z(0;T) = \{0\}$.

(b) If $v \in V$ is an eigenvector of T, then $Z(v;T) = \operatorname{span}(v)$. Conversely, if $\dim(Z(v;T)) = 1$, then v is an eigenvector.

(b) shows that the dimension of $Z(v;T)$ somehow measures how for v is from begin an eigenvector.

Def'n. Cyclic Vector for a Linear Operator
Let $T : V \to V$ be a linear operator. We say $v \in V$ is a ***cyclic vector*** for T if

$$Z(v;T) = V.$$

(EX 7.3)

(a) Given a linear operator, it may not have a cyclic vector at all. For instance, $I : V \to V$, the identity operator on V, has no cyclic vector whenever $\dim(V) \geq 2$. This is because every nonzero $v \in V$ is an eigenvector of I.

(b) Let $V = \mathbb{F}^2$ and $\beta = \{e_1, e_2\}$ be the standard basis for V. Suppose that $T : V \to V$ is represented by the matrix

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

Then one has

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

so $\{e_1, T(e_1)\}$ is a basis for V (i.e. e_1 is a cyclic vector for V).

(EX 7.4) Given an example of $T : V \to V$ and nonzero $v \in V$ such that v is neither eigenvector nor cyclic vector for T.

Proof. Exercise. ∎

Def'n. T-annihilator Ideal of a Vector
Let $T : V \to V$ be a linear operator and fix $v \in V$. We define the ***T-annihilator ideal*** of v to be the polynomial ideal

$$M(v;T) = \{f \in \mathbb{F}[x] : f(T)(v) = 0\}.$$

(7.5) T-annihilator Polynomial

It is immediate from the above definition that

$$M(v;T) = \{f \in \mathbb{F}[x] : f(T)(v) = 0\} \subseteq \{f \in \mathbb{F}[x] : f(T) = 0\} = I_T.$$

We can also verify that $M(v;T)$ is an ideal analogously to how we verified I_T is an ideal. Now, $\mathbb{F}[x]$ is a principal ideal domain, so $M(v;T)$ is generated by a unique monic polynomial $p_v \in M(v;T)$ which generates $M(v;T)$. We call p_v the ***T-annihilator*** of v. One sees that, if V is finite dimensional, then $M(v;T) \neq \{0\}$ for any nonzero $v \in V$. Moreover, if p_v is the T-annihilator of v and p is the minimal polynomial of T, then $p_v \mid p$. That is, we are restricting the study of the minimal polynomial on a single vector v.

Lemma 7.2. *Let $T : V \to V$ be linear and let $v \in V$. If $v \neq 0$, then*

$$\deg(p_v) > 0.$$

Proof. Suppose $\deg(p_v) = 0$, for the sake of contradiction. Then $p_v = 1$, so

$$0 = p_v(T)(v) = v,$$

which is a contradiction. ∎

(7.6) Now, fix a linear operator $T : V \to V$ on a finite dimensional vector space V over $\mathbb{F}$. Recall that, given a vector $v \in V$, the *T-annihilator ideal* is given by

$$M(v;T) = \{f \in \mathbb{F}[x] : f(T)(v) = 0\},$$

and the *T-annihilator* p_v of v is the unique monic generator of $M(v;T)$. We also defined the *T-cyclic subspace generated by v* to be

$$Z(v;T) = \operatorname{span}\left\{T^l(v) : l \in \mathbb{N}\right\},$$

which is the smallest T-cyclic subspace which has v as an element. We now relate these notions by the following theorem.

Theorem 7.3.

Let $v \in V$.

(a) $\deg(p_v) = \dim(Z(v;T))$.

(b) If $k = \deg(p_v)$, then

$$\left\{v, T(v), \ldots, T^{k-1}(v)\right\}$$

is a basis for $Z(v;T)$.

(c) The minimal polynomial of $T|_{Z(v;T)} : Z(v;T) \to Z(v;T)$ is p_v.

Proof. Observe that (a) follows from (b).

(b) For linear independence, suppose

$$\sum_{i=0}^{k-1} a_i T^i(v) = 0$$

for some $a_0, \ldots, a_{k-1} \in \mathbb{F}$. Let $g = \sum_{i=0}^{k-1} a_i x^i$. So $g(T)(v) = 0$, which means $g \in M(v;T)$. Since p_v generates $M(v;T)$, $p_v | g$. If $g \neq 0$, then $\deg(g) \geq \deg(p_v) = k$, which is a contradiction (as $\deg(g) = k-1$ by construction). Thus $g = 0$, which means $a_0, \ldots, a_{k-1} = 0$, so $\left\{v, \ldots, T^{k-1}v\right\}$ is linearly independent. For spanning part, fix $w \in Z(v;T)$. Then $w = f(T)(v)$ for some $f \in \mathbb{F}[x]$ by definition of $Z(v;T)$. Now, we can divide f by p_v to obtain $q, r \in \mathbb{F}[x]$ with

$$f = p_v q + r$$

and $r = 0$ or $\deg(r) < \deg(p_v) = k$. Then

$$w = f(T)(v) = p_v(T)\,q(T)(v) + r(T)(v) = q(T)\,p_v(T)(v) + r(T)(v) = r(T)(v),$$

since p_v is the T-annihilator of v. Write

$$r = \sum_{i=0}^{k-1} a_i x^i$$

with $a_0, \ldots, a_{k-1} \in \mathbb{F}$. That is,

$$w = \sum_{i=0}^{k-1} a_i T^i(v),$$

which exactly means that $\left\{v, \ldots, T^{k-1}(v)\right\}$ spans $Z(v;T)$.

(c) Since $Z(v;T)$ is a T-invariant subspace,

$$T|_{Z(v;T)} : Z(v;T) \to Z(v;T)$$

is well defined. Moreover, one also knows that

$$p_v\left(T|_{Z(v;T)}\right) = p_v(T)|_{Z(v;T)}. \qquad [7.1]$$

Every $w \in Z(v;T)$ is of the form $w = f(T)(v)$ for some $f \in \mathbb{F}[x]$, so

$$p_v(T)(w) = p_v(T)\,f(T)(v) = f(T)\,p_v(T) = 0.$$

But this means the restriction of $p_v(T)$ onto $Z(v;T)$ vanishes at every $w \in Z(v;T)$, so

$$p_v \in I\left(T|_{Z(v;T)}\right)$$

by [3.1]. To show that p_v is of the minimal degree, suppose that $h \in I\left(T|_{Z(v;T)}\right)$. Then

$$0 = h\left(T|_{Z(v;T)}\right) = h(T)|_{Z(v;T)}.$$

In particular,

$$h(T)(v) = 0,$$

so $h \in M(v;T)$. But p_v is the generator of $M(v;T)$, so $p_v | h$. Thus $I\left(T|_{Z(v;T)}\right) = \langle p_v \rangle$. ∎

Corollary 7.3.1. *p_v is the characteristic polynomial of $T|_{Z(v;T)}$.*

Proof. By (a) of Theorem 7.3, one knows

$$\deg(p_v) = \dim(Z(v;T)).$$

Moreover, (c) provides that p_v divides the characteristic polynomial of $T|_{Z(v;T)}$. But p_v and the characteristic polynomial $T|_{Z(v;T)}$ are monic, so p_v is the characteristic polynomial of T. ∎

Corollary 7.3.2. *Suppose that there exists a T-cyclic vector $v \in V$. Then the minimal polynomial of v is the characteristic polynomial of T.*

Proof. Since v is cyclic, $Z(v;T) = V$, and the result follows from Corollary 3.3.1. ∎

Corollary 7.3.3. *Let $p = x^k + \sum_{i=0}^{k-1} a_i x^i \in \mathbb{F}[x]$ be the minimal polynomial of T. Then T has a cyclic vector if and only if there exists a basis β for V such that*

$$[T]_\beta = \begin{bmatrix} 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & 1 & -c_{k-1} \end{bmatrix}$$

(7.7) Companion Matrix

We call $[T]_\beta$ the ***companion matrix*** of the minimal polynomial of T.

Proof. Suppose $[T]_\beta$ is of the above form for some ordered basis

$$\beta = (v_0, \ldots, v_k).$$

Then from the matrix representation, we know that

$$v_i = T^i(v_0)$$

for all $i \in \{1, \ldots, k\}$. So

$$\beta = \left(v_0, \ldots, T^{k-1}(v_0)\right)$$

is a basis for V and v_0 is a T-cyclic vector. Conversely, suppose that T has a cyclic vector, say $v \in V$. From Theorem 7.3,

$$\beta = \left(v, \ldots, T^{k-1}\right)$$

is a basis for $Z(v;T)$, but v is cyclic, so β is a basis for V. Now, when we compute $[T]_\beta$, we obtain

$$[T]_\beta = \begin{bmatrix} 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & 1 & -c_{k-1} \end{bmatrix},$$

where the last column follows from the fact that

$$x^k + \sum_{i=0}^{k-1} a_i x^i$$

is the minimal polynomial for T, so

$$T^k(v) = -\sum_{i=0}^{k-1} a_i T^i(v). \qquad \blacksquare$$

Cyclic Decomposition and Rational Form

(7.8) Cyclic Decomposition

Throughout this section, fix a finite vector space V over $\mathbb{F}$ and fix a linear operator $T : V \to V$. Then the previous section provides that, for any $v \in V$,

$$T|_{Z(v;T)}$$

is very well understood:

(a) its minimal polynomial is its characteristic polynomial;

(b) it is represented by the companion matrix of its minimal polynomial.

Now, what we would like to get is a *T-cyclic decomposition*

$$V = \bigoplus_{i=1}^{r} Z(v_r;T),$$

as we have seen that any direct sum decomposition of a vector space allows the study of T in terms of its restrictions onto subspaces. So let us work towards this. The first step would be writing V as a sum of canonically chosen T-cyclic subspaces. We begin by choosing $v_1 \in V$. To do so, consider

$$\{p_v : v \in V \wedge v \neq 0\},$$

the set of T-annihilators of nonzero v. Each p_v with nonzero v satisfies $0 < \deg(p_v) \leq \dim(V)$, so

$$\max_{v \in V} \deg(p_v)$$

is well-defined, and we choose v_0 such that $\deg(p_0) = \max_{v \in V} \deg(p_v)$. Next, to choose v_2, we proceed as follows. Denote $W_1 = Z(v_1;T)$. If v_1 is cyclic, then $W_1 = V$, and the whole process is done. So suppose that v_1 is not cyclic. Then we consider, for each $v \in V$, the collection of polynomials

$$S(v;W_1) = \{f \in \mathbb{F}[x] : f(T)(v) \in W\}$$

which we call the T-conductor ideal of v into W_1.

Def'n. T**-conductor Ideal** of a Vector into a Subspace
Let $W \subseteq V$ be a T-invariant subspace and let $T : V \to V$ be linear. Then we define the ***T*-conductor ideal** of v into W by

$$S(v;W) = \{f \in \mathbb{F}[x] : f(T)(v) \in W\}.$$

Here are some properties of T-conductor ideal.

(a) If $W = \{0\}$, then $S(v;W) = M(v;T)$, the T-annihilator ideal of v.

(b) $S(v;W)$ is a nonzero ideal. In particular, the characteristic polynomial is in $S(v;W)$.

(c) $v \notin W$ if and only if $S(v;W) \neq \mathbb{F}[x]$.

(d) If $v, u \in V$ are such that $v - u \in S(v;W) = S(u;W)$.

Proof. Exercise. ∎

Then by the principal ideal theorem, there exists a unique monic polynomial which generates $S(v;W)$.

Def'n. T**-conductor** of a Vector into a Subspace
Given linear $T : V \to V$ and T-invariant $W \subseteq V$, we define the ***T*-conductor** of $v \in V$ into W, $s(v;W)$, to be the unique monic polynomial that generates $S(v;W)$.

It is immediate from the definition that

(a) if $W = \{0\}$, then $s(v;W) = p_v$; and

(b) if $v \notin W$, then $0 < \deg(s(v;W)) \leq \dim(V)$.

Back to our construction, we let $v_2 \in V$ such that $\deg(s(v_2, W_1))$ is maximal among all $s(v;W_1)$, $v \in V \setminus W_1$. We then define

$$W_2 = Z(v_1;T) + Z(v_2;T).$$

Similarly, given

$$W_{k-1} = \sum_{i=1}^{k-1} Z(v_i;T) \neq V$$

for some $v_1, \ldots, v_{k-1} \in V$, define v_k such that $\deg(s(v_k;W_{k-1})) = \max_{v \in V \setminus W_{k-1}} \deg(s(v;T))$. This process has to halt somewhere, as each $Z(v_i;T)$ is a nonzero subspace. Therefore, there exists $r \in \mathbb{N}$ and $v_1, \ldots, v_r \in V$ such that

$$V = \sum_{i=1}^{r} Z(v_i;T)$$

and for each $i \in \{1, \ldots, r\}$, v_i is such that

$$\deg(s(v_i;W_{i-1})) = \max_{v \in V \setminus W_{i-1}} \deg(s(v;W_{i-1})).$$

In fact, the above maximal property for v_1 follows from the fact that $\{0\}$ is a T-invariant subspace and $s(v_1;\{0\}) = p_{v_1}$. Moreover, each W_i is T-invariant, and if $i < r$, then W_i is proper, and $s(v_i;W_{i-1})$ is nonconstant for all $i \in \{1, \ldots, r\}$. We now prove some lemma about this construction.

Lemma 7.4. *Let $k \in \{2, \ldots, r\}$ and let $v \in V$. For convenience, denote $f = s(v;W_{k-1})$. If $f(T)(v) = \sum_{i=1}^{k-1} g_i(T)(v_i)$ for some $g_1, \ldots, g_{k-1} \in \mathbb{F}[x]$ (i.e. $g_i(T)(v) \in Z(v_i;T)$), then $f | g_i$ for all $i \in \{1, \ldots, k-1\}$.*

Proof. For each $i \in \{1,\dots,k-1\}$, devide g_i by f to get $q_i, r_i \in \mathbb{F}[x]$ such that

$$g_i = fq_i + r_i$$

with $r_i = 0$ or $\deg(r_i) < \deg(f_i)$. Let

$$w = v - \sum_{i=1}^{k-1} q_i(T)(v_i),$$

then $\sum_{i=1}^{k-1} q_i(T)(v_i) \in W_{k-1}$, as each $v_i \in W_i$ and W_i is T-invariant. Therefore,

$$s(w;W_{k-1}) = s(v;W_{k-1}) = f$$

where

$$f(w) = f\left(v - \sum_{i=1}^{k-1} q_i(T)(v_i)\right) = \sum_{i=1}^{k-1} g_i(T)(v_i) - (fq_i)(T)(v_i) = \sum_{i=1}^{k-1} r_i(T)(v_i).$$

If some $r_i \neq 0$, let $j \leq k-1$ be maximal such that

$$f(T)(w) = \sum_{i=1}^{j} r_i(T)(v_i).$$

Let $p = s(w;W_{j-1})$. Since $W_{j-1} \subseteq W_{k-1}$Since $W_{j-1} \subseteq W_{k-1}$,

$$p \in S(w;W_{k-1}) = \langle f \rangle,$$

which means $fg = p$ for some $g \in \mathbb{F}[x]$. Now,

$$p(T)(w) = gf(T)(w) = \sum_{i=1}^{j} gr_i(T)(v_i) = gr_j(T)(v_j) + \sum_{i=1}^{j-1} gr_i(T)(w_i),$$

where $\sum_{i=1}^{j-1} gr_i(T)(w_i) \in W_{j-1}$. But $p(T)(w) \in W_{j-1}$ as well, so $gr_j(T)(v_j) \in W_{j-1}$. That is, $gr_j \in S(v_j;W_{j-1})$, so $\deg(gr_j) \geq \deg(s(v_j;W_{j-1}))$. But $\deg(s(v_j;W_{j-1})) \geq \deg(s(w;W_{j-1})) = \deg(p) = \deg(fg)$, which is a contradiction, as $\deg(r) < \deg(f)$. Therefore each $r_i = 0$ and so $f|g_i$ for all $i \in \{1,\dots,k-1\}$. ∎

The next step is to modify $v_1,\dots,v_r$ such that they have additional properties

(a) each $s(v_k;W_{k-1}) = p_{v_k}$, the T-annihilator of v; and

(b) $W_i = \bigoplus_{j=1}^{i} Z(v_j;T)$ for all $i \in \{1,\dots,r\}$.

To verify this, we proceed as follows.

(a) **Proof.** Suppose that we have already modified $v_1,\dots,v_{k-1}$ in this way. To modify v_k, let $f = s(v_k;W_{k-1})$ for convenience. Then

$$f(T)(v_k) = \sum_{i=1}^{k-1} g_i(T)(v_{k-1})$$

for some $g_1,\dots,g_{k-1} \in \mathbb{F}[x]$. Then by Lemma 7.4, there exists $h_1,\dots,h_{k-1} \in \mathbb{F}[x]$ such that

$$f(T)(v_k) = \sum_{i=1}^{k-1} fh_i(T)(v_i).$$

Let

$$w_k = v_k - \sum_{i=1}^{k-1} h_i(T)(v_i).$$

Since $\sum_{i=1}^{k-1} h_i(T)(v_i) \in W_{k-1}$,

$$s(w_k; W_{k-1}) = s(v_k; W_{k-1}) = f.$$

But,

$$f(T)(w_k) = f(T)(v_k) - \sum_{i=1}^{k-1} f h_i(T)(v_i) = 0.$$

That is, f annihilates w_k, so $p_{w_k} | f$. But $f | p_{w_k}$, as $p_{w_k}(T)(w_k) = 0 \in W_{k-1}$ so $p_{w_k} \in S(w_k; W_{k-1})$. Therefore $f = p_{w_k}$, as f and p_{w_k} are monic. ∎

(b) We continue from the above result. We split the remaining verification into two steps.

(i) We claim that

$$W_k = W_{k-1} + Z(w_k; T).$$

Proof. $W_k \supseteq W_{k-1} + Z(w_k; T)$ is clear. Conversely, suppose that $w \in W_k$. By definition,

$$W_k = W_{k-1} + Z(v_k; T),$$

so

$$w = w' + g(T)(v_k) = w' + g(T)(v_k - w_k) + g(T)(w_k)$$

for some $w' \in W_{k-1}, g \in \mathbb{F}[x]$. But $w' + g(T)(v_k - w_k) \in W_{k-1}$, and $g(T)(w_k) \in Z(w_k; T)$, so $w \in W_{k-1} + Z(w_k; T)$. ∎

(ii) The next claim is

$$W_{k-1} \cap Z(w_k; T) = \{0\}.$$

Proof. Let $v \in W_{k-1} \cap Z(w_k; T)$. Then $v = g(T)(w_k) \in W_{k-1}$. Therefore, $s(w_k; W_{k-1}) | g$, so $g = hf$ (recall that $f = s(w_k; W_{k-1})$) for some $f \in \mathbb{F}[x]$. Thus

$$v = g(T)(w_k) = hf(T)(w_k) = 0$$

by property (a). ∎

Now we have

$$W_k = W_{k-1} \oplus Z(w_k; T) = \left(\bigoplus_{i=1}^{k-1} Z(v_i; T) \right) \oplus Z(w_k; T),$$

so by replacing v_k by w_k, we have the desired result. We also have the following additional result.

Lemma 7.5.

For all $k \in \{1, \ldots, r\}$,

$$p_{v_k} | p_{v_{k-1}}.$$

Proof. By definition,

$$p_{v_k}(T)(v_k) = 0 = \sum_{i=1}^{k-1} 0 = \sum_{i=1}^{k-1} p_{v_1}(T)(v_i).$$

It follows from Lemma 7.4 that $p_{v_k} | p_{v_i}$ for all $i \in \{1, \ldots, k-1\}$, as desired. ∎

Here is the theorem which summarizes the work so far.

Theorem 7.6. Cyclic Decomposition Theorem

Let $T : V \to V$ be a linear operator on a finite dimensional vector space V. Then there exists $v_1, \ldots, v_r \in V$ such that

(a) $V = \bigoplus_{i=1}^{r} Z(v_i; T)$;

(b) if we denote p_i to be the T-annihilator of v_i, then $p_k | p_{k-1}$ for all $k \in \{2, \ldots, r\}$; and

(c) r and $p_1, \ldots, p_r$ are uniquely determined by (a) and (b).

Notice that we have proved the whole theorem except for (c) (left as reading).

Def'n. Invariant Factors of a Linear Operator

Let $T : V \to V$ be a linear operator on a finite dimensional vector space V and let $p_1, \ldots, p_r \in \mathbb{F}[x]$ be associated with T as described in Theorem 7.6. We call $p_1, \ldots, p_r$ the ***invariant factors*** of T.

(7.9) Invariant Factors

Here are some immediate properties of invariant factors. Let $T : V \to V$ be a linear operator on a finite dimensional V and let $p_1, \ldots, p_r \in \mathbb{F}[x]$ be the invariant factors.

(a) p_1 is the minimal polynomial of T.

Proof. Recall that $p_k | p_{k-1}$ for all $k \in \{2, \ldots, r\}$, so we have $p_k | p_1$ for all $k \in \{1, \ldots, k\}$. Therefore,

$$p_1(T)(v_i) = 0$$

for all $i \in \{1, \ldots, r\}$. But we have a cyclic decomposition $V = \bigoplus_{i=1}^{r} Z(v_i; T)$, which means

$$p_1(T)(v) = 0$$

for any $v \in V$. But we already know that $p | p_1$, so $p = p_1$. ∎

(b) If T has a cyclic vector $v \in V$, then

$$V = Z(v; T)$$

is a cyclic decomposition of V. Therefore, by (a), the T-annihilator of v is the minimal polynomial of T.

Corollary 7.6.1. Rational Form

Let $T : V \to V$ be a linear operator on a finite dimensional vector space V and let $p_1, \ldots, p_r \in \mathbb{F}[x]$ be the invariant factors of T. Then there exists a basis β for V such that

$$[T]_\beta = \begin{bmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_n \end{bmatrix},$$

where each A_i is the companion matrix of p_i, which is of the form

$$A_i = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{bmatrix}$$

given that $p_i = x^k + \sum_{i=0}^{k-1} a_i x^i$.

Proof. By cyclic decomposition theorem, there exists $v_1, \ldots, v_r \in V$ such that $p_1, \ldots, p_r \in \mathbb{F}[x]$ are the corresponding T-annihilators, respectively, and

$$V = \bigoplus_{i=1}^{r} Z(v_i; T).$$

For each $i \in \{1, \ldots, r\}$,

$$\beta_i = \left(v_i, T(v_i), \ldots, T^{k_i - 1}(v_i)\right)$$

is a basis for $Z(v_i; T)$, where $k_i = \deg(p_i)$. Hence $\beta = (\beta_1, \ldots, \beta_r)$ is a basis for V, and

$$[T]_\beta = \begin{bmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_n \end{bmatrix},$$

where each $A_i = \left[T|_{Z(v;T)}\right]_{\beta_i}$. Now we only have to verify that each A_i is the companion matrix of p_i, so fix $i \in \{1, \ldots, r\}$. But v_i is a cyclic vector for $T|_{Z(v_i;T)}$, so the $T|_{Z(v_i;T)}$-annihilator of v_i is the minimal polynomial of $T|_{Z(v_i;T)}$. Thus

$$\left[T|_{Z(v_i;T)}\right]_{\beta_i} = \text{companion matrix of minimal polynomial of } T = \text{companion matrix of } p_i,$$

as desired. ∎

Def'n. Rational Form

Let $T : V \to V$ be a linear operator on a finite dimensional vector space V and let $p_1, \ldots, p_r \in \mathbb{F}[x]$ be the invariant factors of T. We call the matrix

$$\begin{bmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_n \end{bmatrix},$$

where each A_i is the companion matrix of p_i, the ***rational form*** of T. Moreover, any matrix which satisfies the above description is said to be in ***rational form***.

Corollary 7.6.2. *Let $T : V \to V$ be a linear polynomial on a finite dimensional vector space V and let $p_1, \ldots, p_r \in \mathbb{F}[x]$ be the invariant factors of T. Then*

$$\prod_{i=1}^{r} p_i$$

is the characteristic polynomial of T. In particular, the minimal polynomial and characteristic polynomial of T have the same irreducible factors.

Proof. There exists $v_1, \ldots, v_r \in V$ such that

$$V = \bigoplus_{i=1}^{r} Z(v_i; T)$$

and that each p_i is the T-annihilator of v_i. Fix $i \in \{1,\dots,r\}$ and it is immediate that p_i is the minimal polynomial of $T|_{Z(v_i;T)}$. Therefore, the minimal polynomial and characteristic polynomial of $T|_{Z(v_i;T)}$ agrees. But whenever we have a T-invariant decomposition, the characteristic polynomial of a linear operator is the product of its restrictions. Therefore, if we denote $f \in \mathbb{F}[x]$ be the characteristic polynomial of T, then

$$f = \prod_{i=1}^{r} p_i.$$

But p_1 is the minimal polynomial of T and $p_2,\dots,p_r$ divide p_1, so f and p_1 have the same irreducible factors. ∎

Corollary 7.6.3.

For any $A \in M_{n\times n}(\mathbb{F})$, there exists $B \in M_{n\times n}(\mathbb{F})$ in rational form such that A and B are similar.

Proof. Apply Corollary 3.6.1 to $T : \mathbb{F}^n \to \mathbb{F}^n$ where $T(v) = Av$ for any $v \in V$ (i.e. A is the matrix representation of T in the standard ordered basis for $\mathbb{R}^n$). ∎

Theorem 7.7. Generalized Cayley-Hamilton Theorem

Let $T : V \to V$ be a linear operator on a finite dimensional vector space V and let $p \in \mathbb{F}[x]$ be the minimal polynomial of T. Let $f \in \mathbb{F}[x]$ be the characteristic polynomial of T. Write

$$p = \prod_{i=1}^{k} p_i^{r_i}$$

is the prime factorization of p ($r_1,\dots,r_k \in \mathbb{N}$), then the prime factorization of f is

$$f = \prod_{i=1}^{k} p_i^{d_i},$$

where each

$$d_i = \frac{\dim\left(\ker\left(p_i^{r_i}(T)\right)\right)}{\deg(p_i)}.$$

Proof. For convenience, denote

$$V_i = \ker\left(q_i^{r_i}(T)\right).$$

Then by the primary decomposition theorem,

$$V = \bigoplus_{i=1}^{k} V_i$$

and that the mimimal polynomial of $T|_{V_i}$ is $q_i^{r_i}$. Then by Corollary 3.6.2, the characteristic polynomial of each $T|_{V_i}$ is $q_i^{d_i}$ for some $d_i \geq r_i$. Notice that

$$\dim\left(\ker\left(q_i^{r_i}(T)\right)\right) = \dim(V_i) = \deg\left(q_i^{d_i}\right) = d_i \deg(q_i),$$

and the result follows. ∎

Jordan Form

(7.10) Recall that one of the consequences of the primary decomposition theorem is that, given any linear operator $T : V \to V$ on a finite dimensional vector space V, a diagonalizable operator $D : V \to V$ and a nilpotent operator $N : V \to V$ are canonically associated with T (i.e. $T = D + N$ and $DN = ND$). But any diagonalizable operator is easy to understood, so we can easily reduce the study of general linear operators to the study of nilpotent operators, and we can use the cyclic decomposition theorem to understand nilpotent operators better. Let $N : V \to V$ be a nilpotent operator. Then there exists $m \in \mathbb{N}$ such that $N^m = 0$. By cyclic decomposition theorem, there exists $v_1, \dots, v_r \in V$ such that

$$V = \bigoplus_{i=1}^{r} Z(v_i; N)$$

with corresponding invariant factors $p_1, \dots, p_r \in \mathbb{F}[x]$. Since N is nilpotent, the minimal polynomial of N is x^{k_1} for some $k_1 \in \mathbb{N}$, which means $p_1 = x^{k_1}$. It easily follows that $p_2 = x^{k_2}, \dots, p_r = x^{k_r}$ for some $k_2, \dots, k_r \in \mathbb{N}$ with $k_1 \geq k_2 \geq \cdots \geq k_r$. On the other hand, $\prod_{i=1}^{r} p_i = \prod_{i=1}^{r} x^{k_i}$ is the characteristic polynomial of N, so it follows that $\dim(V) = \deg\left(\prod_{i=1}^{r} x^{r_i}\right) = \sum_{i=1}^{r} k_i$. Moreover, notice that the companion matrix of each $p_i = x^{k_i}$ is

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \in M_{k_i \times k_i}(\mathbb{F}),$$

denote which by A_i. Then by taking the appropirate basis β for V, we get the rational form of N

$$N = \begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_r \end{bmatrix}.$$

Actually, the number of blocks, r, has a meaning, as shown by the following proposition.

Proposition 7.8. *Let $N : V \to V$ be a nilpotent operator on a finite dimensional vector space V and let $p_1, \dots, p_r \in \mathbb{F}[x]$ be the invariant factors of N. Then $r = \dim(\ker(N))$.*

Proof. We claim that

$$\alpha = \left(N^{k_1 - 1}(v_1), \dots, N^{k_r - 1}(v_r)\right)$$

is a basis for $\ker(N)$, where $v_1, \dots, v_r \in V$ are chosen such that $V = \bigoplus_{i=1}^{r} Z(v_i; N)$ and that each p_i is the T-annihilator of v_i. It is immediate from the choice that α is linearly independent, as each $N^{k_i - 1}(v_i) \in Z(v_i; N)$. Moreover, each $N^{k_i - 1}(v_i) \in \ker(N)$, as

$$N\left(N^{k_i - 1}(v_i)\right) = N^{k_i}(v_i) = p_i(N)(v_i) = 0.$$

So it suffices to show that α spans $\ker(N)$. Let $v \in \ker(N)$. Then there exists $f_1, \dots, f_r \in \mathbb{F}[x]$ with $\deg(f_i) \leq k_i$ such that

$$v = \sum_{i=1}^{r} f_i(N)(v_i).$$

So

$$0 = N(v) = \sum_{i=1}^{r} N(f_i(N)(v_i)),$$

which means each $N(f_i(N)(v_i)) = 0$ by the cyclic decomposition. Therefore, the N-annihilator x^{k_i} of v_i divides xf_i, which means $xf_i = c_i x^{k_i}$ for some $c_i \in \mathbb{F}$, since the $\deg(f_i) \leq k_i = \deg\left(x^{k_i}\right)$. Therefore, $f_i = c_i x^{k_i - 1}$ and so

$$v = \sum_{i=1}^{r} c_i N^{k_i - 1}(v_i).$$

Thus, every $v \in \ker(N)$ satisfies $v \in \operatorname{span}(\alpha)$, so α is a basis for $\ker(n)$ and $r = \dim(N)$ in particular. ∎

(7.11) Jordan Form

Now that we know the rational form of a nilpotent matrix is particularly simple, let us use this fact to analyze more general class of linear operators. Fix linear $T : V \to V$ on a finite dimensional vector space V. The assumption here is that the characteristic polynomial $f \in \mathbb{F}[x]$ splits over the field. That is, it can be written as a product of linear factors,

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

where $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ are distinct and $d_1, \ldots, d_k \in \mathbb{N}$. Note that this is always the case when $\mathbb{F} = \mathbb{C}$, but for other fields, this may not be true for all polynomials. Then by the generalized Cayley-Hamilton theorem,

$$p = \prod_{i=1}^{k} (x - \lambda_i)^{r_i}$$

for some $r_1, \ldots, r_k \in \mathbb{N}$ such that $r_1 \leq d_1 \cdots r_k \leq d_k$. Furthermore, each $d_i = \dim(\ker((T - \lambda_i I)^{r_i}))$, as each factor is linear. For convenience, denote each $W_i = \ker((T - \lambda_i I)^{r_i})$, then by the primary decomposition theorem,

$$V = \bigoplus_{i=1}^{k} W_i$$

is a T-invariant decomposition. These exponents also have the following meaning: each d_i is the multiplicity of λ_i in f and r_i is the multiplicity of λ_i in p. The primary decomposition theorem provides even more, that $(x - \lambda_i)^{r_i}$ is the minimal polynomial of $T|_{W_i}$ and that we have nilpotent part $N_i = T|_{W_i} - \lambda_i I : W_i \to W_i$ of $T|_{W_i}$. Then from cyclic decomposition theorem, there exists a basis β_i of W_i such that

$$[N_i]_{\beta_i} = \begin{bmatrix} A_{i1} & & & 0 \\ & A_{i2} & & \\ & & \ddots & \\ 0 & & & A_{il_i} \end{bmatrix},$$

where $l_i = \dim(\ker(N_i))$ and

$$A_{ij} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ddots & 1 & 0 \end{bmatrix} \in M_{k_{ij} \times k_{ij}}(\mathbb{F})$$

for all $i \in \{1, \ldots, k\}$. But $N_i = T_i - \lambda_i I$, so

$$[T_i]_{\beta_i} = \begin{bmatrix} J_{i1} & & 0 \\ & \ddots & \\ 0 & & J_{il_i} \end{bmatrix}$$

where

$$J_{ij} = \begin{bmatrix} \lambda_i & 0 & \cdots & 0 & 0 \\ 1 & \lambda_i & \cdots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & 0 \\ \vdots & \vdots & \ddots & \lambda_i & \vdots \\ 0 & 0 & \cdots & 1 & \lambda_i \end{bmatrix}.$$

Moreover, the integers $k_{i1}, \cdots, k_{il_i}$ satisfies that $k_{i1} \geq \cdots \geq k_{il_i} \geq 1$ and that $\sum_{j=1}^{l_i} k_{ij} = \dim(W_i) = d_i$. Now we put the bases $\beta_1, \ldots, \beta_k$ together: let $\beta = (\beta_1, \ldots, \beta_k)$. Then we get

$$[T]_\beta = \begin{bmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_k \end{bmatrix}, \tag{7.2}$$

where each $B_i = [T_i]_{\beta_i}$. We call each J_{ij} an ***elementary Jordan block*** and $[T]_\beta$ the ***Jordan form*** of T. In summary, whenever the characteristic polynomial of a linear operator splits over the field, the operator has a block matrix representation as shown in [3.2], where k is the number of distinct eigenvalues, the block matrices $B_1, \ldots, B_k$ decrease in size as the index runs from 1 to k, and each B_i is a block matrix with elementary Jordan matrices. So we have proved the following theorem (except for the uniqueness).

Theorem 7.9. Jordan Form

Let $T : V \to V$ be a linear operator on a finite dimensional vector space V. If the characteristic polynomial $f \in \mathbb{F}[x]$ of T splits over $\mathbb{F}$,

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

where $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ are distinct eigenvalues of T and $d_1, \ldots, d_k \in \mathbb{N}$, then there exists a basis β for V such that

$$[T]_\beta = \begin{bmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_k \end{bmatrix},$$

where each $B_i \in M_{d_i \times d_i}(\mathbb{F})$ and each B_i are made of elementary Jordan blocks $J_{i1}, \ldots, J_{il_i}$ with λ_i on the main diagonal and 1 on the off diagonal below the main diagonal,

$$B_i = \begin{bmatrix} J_{i1} & & 0 \\ & \ddots & \\ 0 & & J_{il_i} \end{bmatrix},$$

where $l_i = \dim(\ker(T - \lambda_i I)^{r_i})$, r_i is the multiplicity of λ_i in the minimal polynomial of T, and each $J_{ij} \in M_{k_i \times k_i}(\mathbb{F})$ with

$$r_i = k_{i1} \geq \cdots \geq k_{il_i}$$

and

$$\sum_{j=1}^{l_i} k_{ij} = d_i.$$

Moreover, the Jordan form T is unique is up to a reordering of $B_1, \ldots, B_k$.

The uniqueness part can be proved by using the uniqueness parts of the primary decomposition theorem and cyclic decomposition theorem. Also note that the choice of basis is not unique.

Corollary 7.9.1. *Every matrix over $\mathbb{C}$ is similar to a matrix in Jordan form.*

Proof. Fix $A \in M_{n\times n}(\mathbb{C})$ and let ε be the standard ordered basis for $\mathbb{C}^n$. Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be the linear operator such that $[T]_\varepsilon = A$. Since $\mathbb{C}$ is an algebraically closed field, every polynomial splits over $\mathbb{C}$, so by the Jordan form, there exists a basis β for V such that $[T]_\beta$ is in Jordan form. This $[T]_\beta$ is similar to A. ∎

(7.12) Suppose that we A is a diagonal matrix with the eigenvalues grouped together. That is,

$$A = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \vdots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \lambda_k & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_k \end{bmatrix},$$

where $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ are distrinct eigenvalues of A. Then A is in Jordan form, and by the uniqueness part, A is the Jordan form of A. Thus, given a diagonalizable linear operator $T : V \to V$ on a finite dimensional vector space V, the Jordan form of T is a diagonal matrix (with the eigenvalues grouped together). The point is that the Jordan form is a generalization of diagonal form for diagonalizable matrices.

8.
Inner Product Spaces

Inner Products

(8.1) Motivation

The motivation for inner products is the *dot product*. That is, given any $v,u \in \mathbb{R}^n$, we define the dot product of v,u by

$$v \cdot u = \sum_{i=1}^{n} v_i u_i.$$

It is also well known that

$$v \cdot u = \|v\| \, \|u\| \cos(\theta),$$

where $\|\cdot\|$ is the norm (i.e. length; more on this later) of a vector and θ is the angle between v,u. That is, the dot product on $\mathbb{R}^n$ somehow encodes information of length and angle. But on (suitable) abstract vector spaces, we shall proceed in the reverse direction. In other words, we are going talk about inner products, and use them to define the notion of length and, although we would not be able to fully define the notion of angle, the notion of perpendicularity.

(8.2)

Throughout this chapter, we are going to work only with vector spaces over $\mathbb{R}$ or $\mathbb{C}$. That is, if we write $\mathbb{F}$, it is either $\mathbb{R}$ or $\mathbb{C}$.

Def'n. Inner Product on a Vector Space

Let V be a vector space. An ***inner product*** on V is a function

$$\langle \cdot, \cdot \rangle : V^2 \to \mathbb{F}$$

satisfying, for any $v,u,w \in V$ and $c \in \mathbb{F}$,

(a) *linearity in the first argument*: $\langle cv+u, w \rangle = c\langle v,w \rangle + \langle u,w \rangle$.

(b) *conjugate symmetry*: $\langle v,w \rangle = \overline{\langle w,v \rangle}$.

(c) *positive definiteness*: $\langle v,v \rangle \geq 0$ and $\langle v,v \rangle = 0 \implies v = 0$.

(8.3)

(a) Linearity in the first argument means that, the map $\langle \cdot, w \rangle : V \to \mathbb{F}$ for some fixed $w \in V$ is a linear transformation.

(b) When $\mathbb{F} = \mathbb{R}$, we have a symmetry: $\langle v,w \rangle = \langle w,v \rangle$ for any $v,w \in V$. Moreover, even when $\mathbb{F} = \mathbb{C}$, then

$$\langle v,v \rangle = \overline{\langle v,v \rangle}$$

for all $v \in V$, which means that $\langle v,v \rangle \in \mathbb{R}$. This allows the third property of inner product, positive definiteness.

(8.4)

Here are some immediate consequence of the defintion. Fix $v,u,w \in V$ and $c \in F$.

(a) $\langle v, u+w \rangle = \langle v,u \rangle + \langle v,w \rangle$.

(b) $\langle v, cw \rangle = \overline{c}\langle v,w \rangle = \langle \overline{c}v, w \rangle$.

(c) $\langle 0,v \rangle = \langle v,0 \rangle = 0$.

(EX 8.5) Here are some basic examples of inner products.

(a) *standard inner product on* $\mathbb{F}^n$: The standard inner product on $\mathbb{F}^n$ is analogous to the dot product. That is, given any $v=(v_1,\ldots,v_n), u=(u_1,\ldots,u_n)\in\mathbb{F}^n$, we have

$$\langle v,u\rangle=\sum_{i=1}^{n} v_i\overline{u_i}.$$

This is indeed an inner product. Moreover, notice that, when $\mathbb{F}=\mathbb{R}$, we have $\langle v,u\rangle=v\cdot u$ for all $v,u\in\mathbb{F}^n$.

(b) *standard inner product on* $M_{n\times n}(\mathbb{F})$: Given any $A,B\in M_{n\times n}(\mathbb{F})$,

$$\langle A,B\rangle=\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\overline{B_{ij}}.$$

This indeed defines an inner product on $M_{n\times n}(\mathbb{F})$, which is called the standard inner product on $M_{n\times n}(\mathbb{F})$.

The two examples above are indeed identical (up to isomorphism). That is, we have $M_{n\times n}(\mathbb{F})\cong\mathbb{F}^{n^2}$ in a natural way, and for such natural isomorphism, the standard inner products coincides. A cleaner formulation of (b) would be that, given $A\in M_{n\times n}(\mathbb{F})$, call A^* be the ***conjugate transpose*** of A, such that

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

Notice that, when $\mathbb{F}=\mathbb{R}$, $A^*=A^T$. Then it turns out that, given any $A,B\in M_{n\times n}(\mathbb{F})$,

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

Proof. Observe that

$$\langle A,B\rangle=\sum_{i,j}A_{ij}\overline{B}_{ji}=\sum_{i,j}A_{ij}(B^*)_{ji}=\sum_i\left(\sum_j A_{ij}(B^*)_{ji}\right)=\sum_i(AB^*)_{ii}=\operatorname{tr}(AB^*).$$ ∎

(EX 8.6) Let $T:V\to W$ be injective and linear and suppose that an inner product $\langle\cdot,\cdot\rangle_W$ on W is given. Then we get an inner product $\langle\cdot,\cdot\rangle_V$ on V by

$$\langle v,u\rangle_V=\langle T(v),T(u)\rangle_W.$$

Proof. Exercise. ∎

(EX 8.7) We can also define inner products for infinite dimensional vector spaces. Let

$$V=\left\{f\in\mathbb{R}^{[0,1]}:f\text{ is continuous}\right\}.$$

Then we can define an inner product $\langle\cdot,\cdot\rangle$ on V by

$$\langle f,g\rangle=\int_0^1 f(t)\,g(t)\,\mathrm{d}t.$$

Matrix Representation of Inner Products

Def'n. Matrix Representation of an Inner Product

Let V be a finite dimensional vector space and let $\langle \cdot, \cdot \rangle$ be an inner product on V. Suppose that a basis $\beta = (v_1, \ldots, v_n)$ for V is given. Define a matrix $G \in M_{n\times n}(\mathbb{F})$ by

$$G_{ij} = \langle v_j, v_i \rangle .$$

We call this G the ***matrix representation*** of $\langle \cdot, \cdot \rangle$ with respect to β.

Proposition 8.1. *Let V be a finite dimensional vector space and let $\langle \cdot, \cdot \rangle$ be a vector space on V. Suppose that we have a basis $\beta = (v_1, \ldots, v_n)$ for V and let G be the matrix of $\langle \cdot, \cdot \rangle$ with respect to β.*

(a) $\langle \cdot, \cdot \rangle$ is uniquely determined by G (i.e. if $\langle \cdot, \cdot \rangle'$ is represented by G with respect to β, then $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle'$). In fact, if

$$v = \sum_{i=1}^{n} a_i v_i, w = \sum_{i=1}^{n} b_i v_i,$$

for some $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{F}$, then

$$\langle v, w \rangle = \begin{bmatrix} \overline{b_1} \cdots \overline{b_n} \end{bmatrix} G \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

(b) $G = G^*$.

(c) For all $X \in M_{n\times 1}(\mathbb{F})$,

$$X^* G X \geq 0.$$

(d) G is invertible.

Proof.

(a) Observe that

$$\langle v, w \rangle = \left\langle \sum_{i=1}^{n} a_i v_i, b_i v_i \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} a_j \overline{b_i} \langle v_j, v_i \rangle = \sum_{i,j} \overline{b_i} a_j G_{ij} = \begin{bmatrix} \overline{b_1} \cdots \overline{b_n} \end{bmatrix} G \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}. \blacksquare$$

(b) Observe that

$$\overline{G_{ji}} = \overline{\langle v_i, v_j \rangle} = \langle v_j, v_i \rangle = G_{ij}. \blacksquare$$

(c) Let

$$X = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

and let $v = \sum_{i=1}^{n} a_i v_i \in V$. Then

$$X^* G X = \begin{bmatrix} \overline{a_1} \cdots \overline{a_n} \end{bmatrix} G \begin{bmatrix} a_1 \\ \vdots \end{bmatrix} = \langle v, v \rangle \geq 0$$

by (a). $\blacksquare$

(d) For the sake of contradiction, suppose that G is not invertible. Then there exists some $X \in M_{n\times 1}(\mathbb{F})$ such that $X \neq 0$ but $GX = 0$. But this contradicts (c), since this implies $X^*GX = 0$ but $X \neq 0$. Thus G is invertible. ∎

Def'n. Hermitian Matrix
We call $A \in M_{n\times n}(\mathbb{F})$ a ***Hermitian*** matrix if $A = A^*$.

Proposition 8.2. *Let V be a finite dimensional vector space and let $\beta = (v_1, \ldots, v_n)$ is a basis for V. Let $G \in M_{n\times n}(\mathbb{F})$ be Hermitian and satisfying $X^*GX > 0$ for all nonzero $X \in M_{n\times 1}(\mathbb{F})$. Then there exists an inner product $\langle \cdot, \cdot \rangle$ on V such that G is the matrix representation with respect to β.*

Proof. Define $\langle \cdot, \cdot \rangle$ by, given any $v = \sum_{i=1}^n a_i v_i, w = \sum_{i=1}^n b_i v_i$,

$$\langle v, w \rangle = \begin{bmatrix} \overline{b}_1 \cdots \overline{b}_n \end{bmatrix} G \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

Then we can verify that $\langle \cdot, \cdot \rangle$ is indeed an inner product on V, and clearly G is the matrix representation of $\langle \cdot, \cdot \rangle$. ∎

(EX 8.8) Let β be the standard basis for $\mathbb{F}^n$ and let $\langle \cdot, \cdot \rangle$ be the standard inner product on $\mathbb{F}^n$. Then the matrix representation of $\langle \cdot, \cdot \rangle$ with respect to β is I, the identity matrix.

Inner Product Spaces

Def'n. Inner Product Space over a Field
An ***inner product space*** over $\mathbb{F}$ is a vector space V over $\mathbb{F}$ with a fixed inner product $\langle \cdot, \cdot \rangle$ on V. We usually use an ordered pair $(V, \langle \cdot, \cdot \rangle)$ (or simply V for convenience) to denote an inner product space.

Def'n. Norm on an Inner Product Space
Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then we define the ***norm*** on V, denoted by $\|\cdot\| V \to \mathbb{F}$ by

$$\|v\| = \sqrt{\langle v, v \rangle}$$

for all $v \in V$.

(8.9) Here are some immediate properties following from the above definition.

(a) *positive definiteness*: For any $v \in V$, $\|v\| \geq 0$. In particular, if $v \neq 0$, then $\|v\| > 0$.

(b) *absolute homogeniety*: For any $v \in V$ and for any $c \in \mathbb{F}$, $\|cv\| = |c| \, \|v\|$.

Notice that the above properties show that a norm is an abstract notion of length. An important inequality that follows from the defintion of the norm is the following.

Theorem 8.3. Cauchy-Schwartz Ineqality

Let V be an inner product space. Then for any $v, w \in V$,

$$|\langle v, w\rangle| \leq \|v\| \, \|w\| .$$

Proof. Notice that the ineqality is immediate when $v = 0$. So suppose $v \neq 0$. Consider

$$u = w - \frac{\langle w, v\rangle}{\|v\|^2} v \in V.$$

Then notice that

$$\langle u, v\rangle = \left\langle w - \frac{\langle w, v\rangle}{\|v\|^2} v, v \right\rangle = \langle w, v\rangle - \frac{\langle w, v\rangle}{\|v\|^2} \langle v, v\rangle = \langle w, v\rangle - \langle w, v\rangle = 0.$$

Now, we have

$$0 \leq \|u\|^2 = \left\langle u, w - \frac{\langle w, v\rangle}{\|v\|^2} v \right\rangle = \langle u, w\rangle - \frac{\overline{\langle w, v\rangle}}{\|v\|^2} \langle u, v\rangle = \langle u, w\rangle .$$

But

$$\langle u, w\rangle = \left\langle w - \frac{\langle w, v\rangle}{\|v\|}, w \right\rangle = \langle w, w\rangle - \frac{\langle w, v\rangle}{\|v\|^2} \langle v, w\rangle = \|w\|^2 - \frac{|\langle v, w\rangle|^2}{\|v\|^2}.$$

Thus

$$|\langle v, w\rangle| \leq \|v\| \, \|w\|$$

by some rearranging. ∎

(8.10) Now probabily the question to ask is that, when does we get an eqality? It turns out the Cauchy-Schwartz ineqality holds witn an eqality if and only if v, w are linearly dependent.

Proof. Say $v = cw$ for some $c \in \mathbb{F}$. So

$$|\langle v, w\rangle| = |\langle cw, w\rangle| = |c| \, \|w\|^2 = \|w\| \, (|c| \, \|w\|) = \|w\| \, \|v\| .$$

Conversely, suppose that $|\langle v, w\rangle| = \|v\| \, \|w\|$. From the presented proof of Cauchy-Schwartz ineqality, we get (by assuming $v \neq 0$)

$$0 = \|u\| = \left\| w - \frac{\langle w, v\rangle}{\|v\|^2} v \right\|,$$

which means that $w - \frac{\langle w, v\rangle}{\|v\|^2} v = 0$ by positive definiteness. Thus v, w are linearly dependent. ∎

(8.11) The Cauchy-Schwartz inequality has various forms, depending on which vector space we are working with.

(a) If $V = \mathbb{F}^n$ and $\langle \cdot, \cdot \rangle$ is the standard inner product on $\mathbb{F}^n$, then the Cauchy-Schwartz ineqality is

$$\left| \sum_{i=1}^n a_i \overline{b_i} \right| \leq \sqrt{\sum_{i=1}^n |a_i|^2} \sqrt{\sum_{i=1}^n |b_i|^2}.$$

(b) If $V = M_{n\times n}(\mathbb{F})$ and $\langle \cdot, \cdot \rangle$ is the standard inner product (i.e. $\langle A, B\rangle = \operatorname{tr}(AB^*)$), then the Cauchy-Schwartz inequality is

$$|\operatorname{tr}(AB^*)| \leq \sqrt{\operatorname{tr}(AA^*)} \sqrt{\operatorname{tr}(BB^*)}.$$

(c) If $V = \left\{ f \in \mathbb{R}^{[0,1]} : f \text{ is continuous} \right\}$ and $\langle \cdot, \cdot \rangle$ is such that

$$\langle f, g \rangle = \int_0^1 fg,$$

then the Cauchy-Schwartz ineqality is

$$\left| \int_0^1 fg \right| \leq \sqrt{\int_0^1 f^2} \sqrt{\int_0^1 g^2}.$$

Another important ineqality involving norm is the triangle ineqality.

Theorem 8.4. Triangle Ineqality

Let V be an inner product space and let $v, w \in V$. Then

$$\|v+w\| = \|v\| + \|w\|.$$

Proof. Notice that

$$\|v+w\| = \langle v+w, v+w \rangle = \|v\|^2 + \langle v, w \rangle + \langle w, v \rangle + \|w\|^2 = \|v\|^2 + 2\Re(\langle v, w \rangle) + \|w\|^2.$$

But $\Re(z) \leq |z|$ for any $z \in \mathbb{C}$, so by the Cauchy-Schwartz ineqality,

$$\|v+w\| \leq \|v\|^2 + 2|\langle v, w \rangle| + \|w\|^2 \leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 = (\|v\| + \|w\|)^2.$$

Thus by taking the positive square roots, we have the desired result. ∎

(8.12) We have

$$\|v+w\| = \|v\| + \|w\|$$

if and only if v, w are $\mathbb{R}$-linearly dependent. That is, there exists real $c \in \mathbb{R}$ such that $v = cw$.

Proof. The reverse direction is clear. Conversely, suppose that $\|v+w\| = \|v\| + \|w\|$. Then by the preseted proof of the triangle ineqality, it must be the case that $\langle v, w \rangle = \|v\|\|w\|$, so v, w are linearly dependent by (8.10). So say $v = cw$. But we also have the eqality, coming from the proof,

$$\Re(\langle cw, w \rangle) = |\langle cw, w \rangle|,$$

which exactly means that $\langle cw, w \rangle \in \mathbb{R}$. Thus $c \in \mathbb{R}$, as desired. ∎

Orthogonality

(8.13) Now that we have discussed about the notion of length coming from inner products (i.e. norm), let us discuss about the notion of angles.

Def'n. Orthogonal Vectors
Let V be an inner product space and let $v, w \in V$. We say v, w are ***orthogonal*** if $\langle v, w \rangle = 0$.

Moreover, we say a subset $S \subseteq V$ is ***orthogonal*** if S is a set of pairwise orthogonal vectors (i.e. $v, w \in S$ are such that $v \neq w$, then $\langle v, w \rangle = 0$). If, in addition, $\|v\| = 1$ for all $v \in S$, we say S is an ***orthonormal*** set of vectors.

Theorem 8.5. Orthogonality Implies Linear Independence

Let V be an inner product space and let $S \subseteq V$ be orthogonal. Then

(a) S is linearly independent; and

(b) if $v_1, \dots, v_n \in S$ are distinct elements and

$$w = \sum_{i=1}^{n} a_i v_i$$

for some $a_1, \dots, a_n \in \mathbb{F}$, then

$$a_i = \frac{\langle w, v_i \rangle}{\|v_i\|^2}$$

for all $i \in \{1, \dots, n\}$.

Proof. Notice that (b) implies in particular that (a). To prove (b), simply calculate

$$\langle w, v_i \rangle = \left\langle \sum_{j=1}^{n} a_j v_j, v_i \right\rangle = \sum_{j=1}^{n} a_j \langle v_j, v_i \rangle = a_i \langle v_i, v_i \rangle = a_i \|v_i\|^2 . \qquad \blacksquare$$

(8.14)

Notice that Theorem 8.5 justifies the following geometrical intuition: namely, given an n-dimensional inner product space, there maximum number of orthogonal vectors is n, where the maximum number of orthogonal vectors is the geometric intuition of dimension. That is, the geometric dimension is bounded above by the algebraic dimension. In fact, they actually agree, as the following theorem shows.

Theorem 8.6. Gram-Schmidt Orthogonalization

Let V be an inner product space and suppose that we have a finite number of linearly independent vectors $v_1, \dots, v_n \in V$. Then $\operatorname{span}\{v_1, \dots, v_n\}$ *has an orthogonal basis.*

Proof. This is a recursive construction of an orthogonal basis $\{w_1, \dots, w_n\}$.

(a) Define $w_1 = v_1$.

(b) Suppose that we have constructed $w_1, \dots, w_m$ for some $m < n$, such that $\{w_1, \dots, w_m\}$ is orthogonal and

$$\operatorname{span}\{w_1, \dots, w_m\} = \operatorname{span}\{v_1, \dots, v_m\} .$$

Define

$$w_{m+1} = v_{m+1} + \sum_{k=1}^{m} \frac{\langle v_{m+1}, w_k \rangle}{\|w_k\|^2} w_k .$$

Observe the following.

(i) If we assume $w_{m+1} = 0$, then $v_{m+1} \in \operatorname{span}\{w_1, \dots, w_m\} = \operatorname{span}\{v_1, \dots, v_m\}$, so we have a contradiction. So w_{m+1} is nonzero.

(ii) Fix $j \in \{1, \dots, m\}$ and observe that

$$\langle w_{m+1}, w_j \rangle = \langle v_{m+1}, w_j \rangle - \left\langle \sum_{k=1}^{m} \frac{\langle v_{m+1}, w_k \rangle}{\|w_k\|^2} w_k, w_j \right\rangle ,$$

where

$$\begin{aligned} \left\langle \sum_{k=1}^{m} \frac{\langle v_{m+1}, w_k \rangle}{\|w_k\|^2} w_k, w_j \right\rangle &= \sum_{k=1}^{m} \frac{\langle v_{m+1}, w_k \rangle}{\|w_k\|^2} \langle w_k, w_j \rangle \\ &= \frac{\langle v_{m+1}, w_j \rangle}{\|w_j\|^2} \langle w_j, w_j \rangle = \langle v_{m+1}, w_j \rangle . \end{aligned}$$

Thus

$$\langle w_{m+1}, w_j \rangle = \langle v_{m+1}, w_j \rangle - \langle v_{m+1}, w_j \rangle = 0,$$

which means $w_1, \ldots, w_{m+1}$ are pairwise orthogonal.

(iii) Clearly $w_1, \ldots, w_{m+1} \in \operatorname{span}\{v_1, \ldots, v_{m+1}\}$. But

$$\dim(\operatorname{span}\{v_1, \ldots, v_{m+1}\}) = m+1$$

and $w_1, \ldots, w_{m+1}$ are linearly independent by Theorem 8.5, we have

$$\operatorname{span}\{w_1, \ldots, w_{m+1}\} = \operatorname{span}\{v_1, \ldots, v_{m+1}\}.$$ ∎

Corollary 8.6.1. *Every finite dimensional inner product space has an orthonormal basis.*

Proof. Let V be a finite dimensional inner product space and let β be a basis for V. Then by Gram-Schmidt orthogonalization, there exists orthogonal γ for V. Given that $\gamma = \{v_1, \ldots, v_n\}$, define

$$\alpha = \left\{ \frac{v_1}{\|v_1\|}, \ldots, \frac{v_n}{\|v_n\|} \right\}.$$

Then clearly α is an orthonormal basis for V. ∎

Def'n. Orthogonal Projection
Let V be an inner product space and let $W \subseteq V$ be a finite dimensional subspace. Let $\beta = \{w_1, \ldots, w_n\}$ be an orthonormal basis for W and let $v \in V$. Then the ***orthogonal projection*** of v onto W is

$$\sum_{i=1}^{n} \langle v, w_i \rangle w_i \in W.$$

(8.15) Notice that, in the above defintion, if $v \in W$, then

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

by Theorem 8.5.

Proposition 8.7. *Let V be an inner product space and let $W \subseteq V$ finite dimensional subspace. Then for any $v \in V$, the orthogonal projection $w \in W$ of v onto W is the unique vector in W such that $v - w$ is orthogonal to every $w' \in W$.*

Proof. Fix an orthonormal basis $(w_1, \ldots, w_n)$ for W. Then the orthogonal projection of v onto W is

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

Let $w' \in W$ and write

$$w' = \sum_{i=1}^{n} a_i w_i.$$

Then notice that

$$\begin{aligned} \langle v - w, w' \rangle &= \sum_{i=1}^{n} \langle v - w, a_i w_i \rangle = \sum_{i=1}^{n} \overline{a_i} \langle v - w, w_i \rangle = \sum_{i=1}^{n} \overline{a_i} \left(\langle v, w_i \rangle - \sum_{j=1}^{n} \langle v, w_j \rangle \langle w_j, w_i \rangle \right) \\ &= \sum_{i=1}^{n} \overline{a_i} \left(\langle v, w_i \rangle - \langle v, w_i \rangle \langle w_i, w_i \rangle \right) = \sum_{i=1}^{n} \overline{a_i} \left(\langle v, w_i \rangle - \langle v, w_i \rangle \right) = 0, \end{aligned}$$

so w, w' are orthogonal. For the uniqueness part, suppose that $u \in W$ has the same property, that $v-u$ is orthogonal to every vector in W. Fix $i \in \{1,\dots,n\}$ and consider $w_i \in W$. Then

$$0 = \langle v-u, w_i\rangle = \langle v, w_i\rangle - \langle u, w_i\rangle,$$

so $\langle v, w_i\rangle = \langle u, w_i\rangle$ for all $i \in \{1,\dots,n\}$. Since $(w_1,\dots,w_n)$ is an orthonormal basis for W, and $u \in W$, so we know that

$$u = \sum_{i=1}^{n} \langle u, w_i\rangle w_i.$$

But then,

$$u = \sum_{i=1}^{n} \langle u, w_i\rangle w_i = \sum_{i=1}^{n} \langle v, w_i\rangle w_i = w$$

by definition, so $u = w$. ∎

Corollary 8.7.1. *Let V be an inner product space and let $W \subseteq V$ be finite dimensional subspace. Then for any $v \in V$, the orthogonal projection of v onto W is independent of the choice of orthonormal basis for W.*

Proposition 8.8. *Let V be an inner product space and let $W \subseteq V$ be finite dimensional subspace. Let $E : V \to W$ be such that, for every $v \in V$, $E(v)$ is the orthogonal projection of v onto W. Then E is a projection onto W.*

Proof. Fix an orthonormal basis $(w_1,\dots,w_n)$ for W. Then

$$E(v) = \sum_{i=1}^{n} \langle v, w_i\rangle w_i$$

for all $v \in V$.

(a) Let $v_1, v_2 \in V$ and let $c \in \mathbb{F}$. Then

$$\begin{aligned} E(cv_1+v_2) &= \sum_{i=1}^{n} \langle cv_1+v_2, w_i\rangle w_i = \sum_{i=1}^{n} c\langle v_1, w_i\rangle w_i + \langle v_2, w_i\rangle w_i \\ &= c\sum_{i=1}^{n} \langle v_1, w_i\rangle w_i + \sum_{j=1}^{n} \langle v_2, w_j\rangle w_j = cE(v_1) + E(v_2), \end{aligned}$$

so E is linear.

(b) For any $v \in V$,

$$\begin{aligned} E^2(v) &= E\left(\sum_{i=1}^{n} \langle v, w_i\rangle w_i\right) = \sum_{j=1}^{n} \left\langle \sum_{i=1}^{n} \langle v, w_i\rangle w_i, w_j\right\rangle w_j \\ &= \sum_{j=1}^{n}\sum_{i=1}^{n} \langle v, w_i\rangle \langle w_i, w_j\rangle w_j = \sum_{j=1}^{n} \langle v, w_j\rangle \langle w_j, w_j\rangle w_j = \sum_{j=1}^{n} \langle v, w_j\rangle w_j = E(v), \end{aligned}$$

so E is a projection.

(c) Clearly $\text{image}(E) \subseteq W$. Conversely, fix $w \in \text{image}(E)$. Since $(w_1,\dots,w_n)$ is an orthonormal basis for W, we have

$$w = \sum_{i=1}^{n} \langle w, w_i\rangle w_i = E(w) \in \text{image}(E).$$

Thus $W \subseteq W$. ∎

Def'n. Orthogonal Complement of a Set of Vectors
Let V be a inner product space and let $S \subseteq V$ be a set of vectors. We define the ***orthogonal complement*** of S to be

$$S^{\perp} = \{v \in V : \forall w \in S\,[\langle v, w\rangle = 0]\}.$$

(8.16)

(a) The orthogonal complement of the whole space is the trivial space, $V^{\perp} = \{0\}$. Similarly, $\{0\}^{\perp} = V$.

(b) For any subset $S \subseteq V$, $S^{\perp}$ is a subspace.

Proof. Let $v_1, v_2 \in S$ and $c \in S$. Then

$$\langle cv_1 + v_2, w\rangle = c\langle v_1, w\rangle + \langle v_2, w\rangle = 0 + 0 = 0$$

for all $w \in S$. Moreover, $\langle 0, w\rangle = 0$ for all $w \in S$ as well, so $0 \in S$. ∎

(c) Suppose that $W \subseteq V$ is a finite dimensional subspace. Let $E : V \to W$ be the orthogonal projection of V onto W. Then, for any $v \in V$, $E(v)$ is the unique vector in W such that $v - E(v) \in W^{\perp}$. Notice that this is a direct consequence of Proposition 8.7.

Proposition 8.9. *Let V be a inner product space and let $W \subseteq V$ be finite dimensional. Then*

$$V = W \oplus W^{\perp}.$$

Proof. Let $E : V \to W$ be the orthogonal proejection. Then we have a direct sum decomposition

$$V = \text{image}(E) \oplus \ker(E) = W \oplus \ker(E).$$

So it remains to prove that $\ker(E) = W^{\perp}$. For this, observe that

$$v \in \ker(E) \iff E(v) = 0 \iff v - 0 \in W^{\perp} \iff v \in W^{\perp}.$$ ∎

Adjoints

Def'n. Linear Functional
Let V be a vector space over $\mathbb{F}$. A ***linear functional*** on V is a linear transformation on V that maps into $\mathbb{F}$.

(EX 8.17) Let V be a vector space and suppose that $\langle \cdot, \cdot \rangle$ is an inner product on V. Then, for fixed $w \in V$,

$$\langle \cdot, w\rangle : V \to \mathbb{F}$$

is a linear functional on V. Note that $\langle w, \cdot\rangle : V \to \mathbb{F}$ is a linear functional if and only if $\mathbb{F} = \mathbb{R}$. The following proposition shows that every linear functional on a finite dimensional vector space is uniquely determined in this way.

Proposition 8.10. *Let V be a finite dimensional inner product space and let $f : V \to \mathbb{F}$ be a linear functional. Then there exists a unique $w \in V$ such that f is given by*

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

Proof. Let $\beta = (v_1, \ldots, v_n)$ be an orthonormal basis for V and let

$$w = \sum_{i=1}^{n} \overline{f(v_i)} v_i.$$

Then for each $j \in \{1, \ldots, n\}$,

$$\langle v_j, w \rangle = \left\langle v_j, \sum_{i=1}^{n} \overline{f(v_i)} v_i \right\rangle = \sum_{i=1}^{n} f(v_i) \langle v_j, v_i \rangle = f(v_j).$$

So $f(v) = \langle v, w \rangle$ for all $v \in V$. For the uniqueness part, suppose that there exists $w' \in V$ such that

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

defines f. Then

$$\langle w - w', w - w' \rangle = \langle w, w \rangle - \langle w, w' \rangle - \langle w', w \rangle + \langle w', w' \rangle = f(w) - f(w) - f(w') + f(w') = 0.$$

Thus by the positive definiteness of $\langle \cdot, \cdot \rangle$, $w - w' = 0$, which means $w = w'$. ∎

Theorem 8.11. *Let V be a finite dimensional inner product space and let $T : V \to V$ be linear. Then there exists a unique linear operator $T^* : V \to V$ such that*

$$\langle T(v), w \rangle = \langle v, T^*(w) \rangle$$

for all $v, w \in V$.

Proof. Fix $w \in V$, and consider the linear functional $f : V \to \mathbb{F}$ defined by

$$v \mapsto \langle T(v), w \rangle .$$

By Proposition 8.10, there exists a unique vector $w' \in V$ such that

$$f(v) = \langle v, w' \rangle$$

for all $v \in V$. Define $T^* : V \to V$ by putting

$$T^*(w) = w'.$$

This T is a unique function, since w' is unique. Then by construction

$$\langle T(v), w \rangle = \langle v, T^*(w) \rangle$$

for all $v, w \in V$. To show that T^* is linear, fix a vector $v \in V$ and observe that we have

$$\langle v, T^*(cw_1 + w_2) \rangle = \overline{c} \langle T(v), w_1 \rangle + \langle T(v), w_2 \rangle = \overline{c} \langle v, T^*(w_1) \rangle + \langle v, T^*(w_2) \rangle = \langle v, cT^*(w_1) + T^*(w_2) \rangle$$

for all $c \in \mathbb{F}$ and $w_1, w_2 \in V$. But this means that

$$\langle \cdot, T^*(cw_1 + w_2) \rangle = \langle \cdot, cT^*(w_1) + T^*(w_2) \rangle$$

as linear functionals, so by Proposition 8.10, $T^*(cw_1 + w_2) = cT^*(w_1) + T^*(w_2)$, as required. ∎

Def'n. Adjoint of a Linear Operator

For any linear operator $T : V \to V$ on a finite dimensional inner product space V, we define the ***adjoint*** of T, denoted by T^*, to be the unique linear operator on V such that

$$\langle T(v), w\rangle = \langle v, T^*(w)\rangle$$

for all $v, w \in V$.

Theorem 8.12. *Let $T : V \to V$ be a linear operator on a finite dimensional inner product space V and let β be an orthonormal basis for V. Then*

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

Proof. Write $\beta = (v_1, \ldots, v_n)$ and let $A = [T]_\beta, B = [T^*]_\beta \in M_{n\times n}(\mathbb{F})$. We desire to show that $B = A^*$. Since β is an orthonormal basis for V,

$$T(v_j) = \sum_{i=1}^{n} \langle T(v_j), v_i\rangle v_i$$

for any $j \in \{1, \ldots, n\}$. So by the definition of matrix representation, we have

$$A_{ij} = \langle T(v_j), v_i\rangle$$

for all $i, j \in \{1, \ldots, n\}$. Similarly, $B_{ij} = \langle T^*(v_j), v_i\rangle$ for all $i, j \in \{1, \ldots, n\}$. Now observe that

$$B_{ij} = \langle T^*(v_j), v_i\rangle = \overline{\langle v_i, T^*(v_j)\rangle} = \overline{\langle T(v_i), v_j\rangle} = \overline{A_{ji}}.$$

Thus $B = A^*$, as desired. ∎

(EX 8.18) Let $\langle \cdot, \cdot \rangle : (\mathbb{F}^n)^2 \to \mathbb{F}$ be the standard inner product on $\mathbb{F}^n$ and let $T : \mathbb{F}^n \to \mathbb{F}^n$ be linear whose matrix is A with respect to the standard basis for $\mathbb{F}$. Then $T^* : (\mathbb{F}^n)^2 \to \mathbb{F}$ is represented by A^* with respect to the standard basis.

Corollary 8.12.1. *Let V be a finite dimensional inner product space and let $T, U : V \to V$ are linear operators. Then the following hold.*

(a) $(T+U)^* = T^* + U^*$.

(b) $(\lambda T)^* = \overline{\lambda} T^*$.

(c) $(TU)^* = U^* T^*$.

(d) $(T^*)^* = T$.

Proof. Fix an orthonormal basis β for V, and prove the corresponding properties for the conjugate transpose of the matrices $[T]_\beta, [U]_\beta$. ∎

Def'n. Hermitian (Self-adjoint) Linear Operator

Let V be a finite dimensional inner product space and let $T : V \to V$ be linear. We say T is ***Hermitian*** (or ***self-adjoint***) if $T^* = T$.

Proposition 8.13. *Orthogonal projections are Hermitian.*

Proof. Let V be an inner product space, let $W \subseteq V$ be a finite dimensional subspace, and let $E : V \to W$ be an orthogonal projection. Fix $v_1, v_2 \in V$. Then

$$\langle E(v_1), v_2 \rangle = \langle E(v_1), E(v_2) + v_2 - E(v_2) \rangle = \langle E(v_1), E(v_2) \rangle + \langle E(v_1), v_2 - E(v_2) \rangle .$$

But $v_2 - E(v_2) \in W^\perp$ by definition, so

$$\langle E(v_1), v_2 \rangle = \langle E(v_1), E(v_2) \rangle + \langle E(v_1), v_2 - E(v_2) \rangle = \langle E(v_1), E(v_2) \rangle .$$

A similar argument shows that $\langle v_1, E(v_2) \rangle = \langle E(v_1), E(v_2) \rangle$. In particular,

$$\langle E(v_1), v_2 \rangle = \langle v_1, E(v_2) \rangle$$

for all $v_1, v_2 \in V$, so $E = E^*$, as desired. ∎

Proposition 8.14. *Let V be a finite dimensional inner product space over $\mathbb{C}$ and let $T : V \to V$ be linear. Then*

$$T = U_1 + iU_2$$

for some Hermitian $U_1, U_2 : V \to V$.

Proof. We proceed to construct $U_1, U_2 : V \to V$. Take

$$U_1 = \frac{1}{2}(T + T^*)$$

and

$$U_2 = \frac{1}{2i}(T - T^*) .$$

Then the result follows. ∎

Isomorphisms of Inner Product Spaces

Def'n. Isomorphism of Inner Product Spaces
Let $(V, \langle \cdot, \cdot \rangle_V), (W, \langle \cdot, \cdot \rangle_W)$ be finite dimensional inner product spaces and let $T : V \to W$ be linear. We say that T ***preserves*** the inner products if

$$\langle T(v_1), T(v_2) \rangle_W = \langle v_1, v_2 \rangle_V$$

for all $v_1, v_2 \in V$. If this is the case, we usually write $T : (V, \langle \cdot, \cdot \rangle_V) \to (W, \langle \cdot, \cdot \rangle_W)$. If in addition, T is bijective, we say T is an ***isomorphism*** of inner product spaces.

Proposition 8.15. *Let $(V, \langle \cdot, \cdot \rangle_V), (W, \langle \cdot, \cdot \rangle_W)$ be finite dimensional inner product spaces and let $T : (V, \langle \cdot, \cdot \rangle_V) \to (W, \langle \cdot, \cdot \rangle_W)$. Then* $\ker(T) = \{0\}$.

Proof. Suppose that $\ker(T) \neq \{0\}$. Then there exists a nonzero $v \in \ker(T)$ such that

$$\langle v, v\rangle_V = \langle T(v), T(v)\rangle_W = \langle 0, 0\rangle_W = 0,$$

which is a contradiction. ∎

Corollary 8.15.1. *Any surjective linear operator that preserves inner prodcut is an isomorphism.*

(8.19) Let V be a finite dimensional inner product space and let $T : V \to V$ be a linear operator. If T preserves inner product, then $\|T(v)\| = \|v\|$ for all $v \in V$.

Proposition 8.16. *Let $(V, \langle\cdot,\cdot\rangle_V), (W, \langle\cdot,\cdot\rangle_W)$ be inner product spaces and let $T : V \to W$ be an isomorphism. Then $T^{-1} : W \to V$ is also an isomorphism.*

Proof. Since T is also an isomorphism of vector spaces, there exists a linear isomorphism $T^{-1} : W \to V$. To show that T^{-1} preserves inner products, take $w_1, w_2 \in W$. Then

$$\left\langle T^{-1}(w_1), T^{-1}(w_2)\right\rangle_V = \left\langle T\left(T^{-1}(w_1)\right), T\left(T^{-1}(w_1)\right)\right\rangle_W = \langle w_1, w_2\rangle_W$$

since T preserves inner products. ∎

Theorem 8.17. Characterizations of Isomorphisms of Inner Product Spaces

Let $(V, \langle\cdot,\cdot\rangle_V), (W, \langle\cdot,\cdot\rangle_W)$ be finite dimensional inner product spaces with $\dim(V) = \dim(W)$ and let $T : V \to W$ be linear. Then the following are equivalent.

(a) T preserves inner products.

(b) T is an isomorphism.

(c) For any orthonormal basis $\beta = (v_1, \ldots, v_n)$ for V, then $T(\beta) = \{T(v_1), \ldots, T(v_n)\}$ is an orthonormal basis for W.

(d) For some orthonormal basis $\beta = (v_1, \ldots, v_n)$ for V, then $T(\beta) = \{T(v_1), \ldots, T(v_n)\}$ is an orthonormal basis for W.

Proof.

- (a) $\Longrightarrow$ (b) Suppose that T preserves inner products. Then $\ker(T) = \{0\}$ by Proposition 8.15, so $\dim(V) = \operatorname{rank}(T)$. But $\dim(V) = \dim(W)$ by assumption, so $\operatorname{rank}(T) = \dim(W)$, which means $\operatorname{image}(T) = W$. Therefore, T is an isomorphism.
- (b) $\Longrightarrow$ (c) Suppose that T is an isomorphism and let $\beta = (v_1, \ldots, v_n)$. Then $T(\beta)$ is a basis for W, and we have
$$\left\langle T(v_i), T(v_j)\right\rangle_W = \left\langle v_i, v_j\right\rangle_V$$
for all $i, j \in \{1, \ldots, n\}$, since any isomorphism of inner product spaces preserves inner products. But this means that
$$\left\langle T(v_i), T(v_j)\right\rangle_W = \left\langle v_i, v_j\right\rangle_V = \delta_{ij},$$
so $T(\beta)$ is an orthonormal basis.
- (c) $\Longrightarrow$ (d) This is clear.

- (d) $\implies$ (a) Suppose that β is an orthonormal basis for V such that $T(\beta)$ is an orthonormal basis for W. Given $v, v' \in V$, write $v = \sum_{i=1}^{n} \langle v, v_i \rangle v_i, v' = \sum_{i=1}^{n} \langle v', v_i \rangle v_i$. Then

$$\langle v, v' \rangle_V = \sum_{i,j=1}^{n} \langle v, v_i \rangle_V \overline{\langle v', v_j \rangle_V} \langle v_i, v_j \rangle_V = \sum_{i=1}^{n} \langle v, v_i \rangle_V \overline{\langle v', v_j \rangle_V},$$

since β is orthonormal. On the other hand,

$$\langle T(v), T(v') \rangle_W = \sum_{i,j=1}^{n} \langle v, v_i \rangle_V \overline{\langle v', v_j \rangle_V} \langle T(v_i), T(v_j) \rangle_V = \sum_{i=1}^{n} \langle v', v_i \rangle_V \overline{\langle v', v_j \rangle_V},$$

since $T(\beta)$ is orthonormal. Therefore, we conclude that

$$\langle v, v' \rangle_V = \langle T(v), T(v') \rangle_W$$

so T preserves inner products. ∎

Corollary 8.17.1. *Any two finite dimensional inner product spaces of the same dimension are isomorphic as inner product spaces.*

Proof. Suppose $(V, \langle \cdot, \cdot \rangle_V), (W, \langle \cdot, \cdot \rangle_W)$ are finite dimensional inner product spaces with $\dim(V) = \dim(W)$ and let $\beta = (v_1, \ldots, v_n), \gamma = (u_1, \ldots, u_n)$ be orthonormal bases for V, W, respectively. Let $T : V \to W$ be a linear isomorphism such that

$$T(v_i) = w_i$$

for all $i \in \{1, \ldots, n\}$. Then T takes an orthonormal basis β for V to an orthonormal basis γ for W, so T is an isomorphism. ∎

(EX 8.20) Let $\langle \cdot, \cdot \rangle$ be the standard inner product of $\mathbb{R}^3$ and let

$$W = \{A \in M_{n \times n}(\mathbb{R}) : A^* = -A\}.$$

This W is a subspace of $M_{n \times n}(\mathbb{R})$, and let $\langle \cdot, \cdot \rangle$ on W be such that $\langle A, B \rangle = \operatorname{tr}(AB^*)$. Then $\dim(W) = 3$: the idea of verifying this fact is that every skew-symmetric 3×3 matrix is of the form

$$\begin{bmatrix} 0 & x_1 & x_2 \\ -x_1 & 0 & x_3 \\ -x_2 & -x_3 & 0 \end{bmatrix}.$$

Therefore, by Corollary 8.17.1,

$$\left(\mathbb{R}^3, \langle \cdot, \cdot \rangle\right) \cong (W, \langle \cdot, \cdot \rangle).$$

Now, to find a concrete isomorphism, find orthonormal bases for $\mathbb{R}^3, W$, resepctively, and define a linear transformation sending one to the another. Then the presented proof of Corollary 8.17.1 tells us that this linear transformation is an isomorphism.

Theorem 8.18. *Let V, W be finite dimensional inner product spaces and let $T : V \to W$ be linear. Then T preserves inner product if and only if T preserves norm.*

Proof. The forward direction is clear. The converse direction follows from the following identities (known as ***polarization identities***)

(a) If $\mathbb{F} = \mathbb{R}$, then

$$\langle v, w \rangle = \frac{1}{4} \|v+w\|^2 - \frac{1}{4} \|v-w\|^2 .$$

(b) If $\mathbb{F} = \mathbb{C}$, then

$$\langle v, w \rangle = \frac{1}{4} \|v+w\|^2 - \frac{1}{4} \|v-w\|^2 + \frac{i}{4} \|v+iw\|^2 - \frac{i}{4} \|v-iw\|^2 .$$

Thus by (a), (b), T preserves inner product whenever T preserves the norm. ∎

Def'n. Unitary Operator on an Inner Product Space
Let V be a finite dimensional inner product space and let $T : V \to V$. If T is isomorphic, then we call T a ***unitary*** operator on V.

Theorem 8.19. Characterization of Unitary Operators by Adjoint

Let V be a finite dimensional inner product space and let $T : V \to V$. Then T is unitary if and only if T is a linear isomorphism and $T^{-1} = T^*$.

Proof. For the forward direction, suppose that T is unitary. Then T is a linear isomorphism, so there exists $T^{-1} : V \to V$. For any $v, w \in V$,

$$\begin{aligned} \langle v, T^*(w) \rangle &= \langle T(v), w \rangle \\ &= \langle T(v), T(T^{-1}(w)) \rangle \\ &= \langle v, T^{-1}(w) \rangle \qquad \text{since } T \text{ preserves inner product} \end{aligned}$$

so $T^* = T^{-1}$. Conversely, suppose that T is a linear isomorphism and $T^{-1} = T^*$. Then for any $v, w \in V$,

$$\langle T(v), T(w) \rangle = \langle v, T^*(T(w)) \rangle = \langle v, T^{-1}(T(w)) \rangle = \langle v, w \rangle$$

so T preserves inner product, which means T is unitary. ∎

(8.21) Characterizations of Unitary Operators

Let V be a finite dimensional inner product space and let $T : V \to V$ be linear. We can summarize characterizations of unitary operators as follows. The following are equivalent.

(a) T is unitary.

(b) T preserves inner product.

(c) For any orthonormal basis β for V, $T(\beta)$ is an orthonormal basis for V.

(d) For some orthonormal basis β for V, $T(\beta)$ is an orthonormal basis for V.

(e) T preserves norm (i.e. $\|T(v)\| = \|v\|$ for all $v \in V$).

(f) $T^{-1} = T^*$.

Often times (e) is taken as the most natural characterization of unitary operators.

Corollary 8.19.1. *Let V be a finite dimensional inner product space and let β be an orthonormal basis for V. Then for any linear $T : V \to V$, T is unitary if and only if $[T]_\beta$ is invertible and $[T]_\beta^{-1} = [T]_\beta^*$.*

Proof. We know that $T : V \to V$ is a linear isomorphism if and only if $[T]_\beta$ is invertible. We also know that

$$\left[T^{-1}\right]_\beta = [T]_\beta^{-1}$$

and

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

This shows that $T^{-1} = T^*$ if and only if $[T]_\beta^{-1} = [T]_\beta^*$, and the result follows. ∎

Corollary 8.19.1 motivates the following definition.

Def'n. Unitary Matrix
Let $A \in M_{n\times n}(\mathbb{F})$. We say A is ***unitary*** if $A^{-1} = A^*$.

(8.22) If $A \in M_{n\times n}(\mathbb{F})$ is unitary, then $\det(A) = 1$.

Proof. Observe that

$$1 = \det(I) = \det\left(A^{-1}A\right) = \det(A^*A) = \det(A^*)\det(A) = \overline{\det(A)}\det(A) = |\det(A)|^2 . \quad ∎$$

When $\mathbb{F} = \mathbb{R}$, then $A \in M_{n\times n}(\mathbb{F})$ is unitary if and only if $A = A^T$. In this case, $\det(A) = \pm 1$.

(EX 8.23) Let $A \in M_{2\times 2}(\mathbb{R})$, and write

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then A is unitary if and only if A is invertible and $A^{-1} = A^T$, so

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} = A^T = A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \pm 1 \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

This happens if and only if

$$A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$

or

$$A = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$$

with $a^2 + b^2 = 1$. That is, the only unitary operators on $\mathbb{R}^2$ are the rotations.

Proposition 8.20. Characterization of Unitary Matrices
Let $A \in M_{n\times n}(\mathbb{F})$. Then the following are equivalent.

(a) A is unitary.

(b) Rows of A are orthornormal with respect to the standard inner product on $\mathbb{F}^n$.

(c) Columns of A are orthonormal with respect to the standard inner product on $\mathbb{F}^n$.

Proof. It is clear that all there conditions fail when A is not invertible, so suppose that A is invertible without loss of generality. Then

$$\begin{aligned} A \text{ is unitary} &\iff AA^* = I \iff (i\text{th row of } A)\,(j\text{th column of } A^*) = \delta_{ij} \\ &\iff (i\text{th row of } A)\,\overline{(j\text{th row of } A)} = \delta_{ij} \\ &\iff \text{rows of } A \text{ are orthonormal,} \end{aligned}$$

where δ_{ij} is the Kronecker delta. Similar varification holds for (c). ∎

Corollary 8.20.1. *Let V be a finite dimensional inner product space and let $T : V \to V$ be linear. Let β, γ be orthonormal bases for V. Then*

$$[T]_\gamma = P^{-1}\,[T]_\beta\,P$$

for some unitary $P \in M_{n\times n}(\mathbb{F})$.

Proof. Write $\beta = (v_1, \ldots, v_n), \gamma = (w_1, \ldots, w_n)$ and let

$$P = [I]_\gamma^\beta,$$

the change of basis matrix from γ to β. Then one has

$$[T]_\gamma = P^{-1}\,[T]_\beta\,P$$

by definition. But then by defining $S : V \to V$ by

$$S(v_i) = w_i$$

$[S]_\beta = P$. Observe that S is unitary, since $S(\beta) = \gamma$. Thus $P = [S]_\beta$ is unitary as well. ∎

Def'n. Unitarily Equivalent Matrices
Let $A, B \in M_{n\times n}(\mathbb{F})$. If there exists a unitary $U \in M_{n\times n}(\mathbb{F})$ such that

$$B = U^{-1}AU,$$

we say A, B are ***unitarily equivalent***.

Orthonormal Diagonalization

(8.24)

Statements regarding linear operators on a finite dimensional inner product space so far are natural analogues of what we know about linear operators on a finite dimensional vector space. Yet another question we can ask is that, given a linear operator $T : V \to V$ on a finite dimensional inner product space V, can we find an orthonormal basis β for V such that $[T]_\beta$ is diagonal? For now, suppose that T is diagonalizable with respect to an orthonormal basis and observe what are the consequences. Say $\beta = (v_1, \ldots, v_n)$ is such that $T(v_i) = c_i v_i$ for some $c_1, \ldots, c_n \in \mathbb{F}$ (we are not insisting that $c_1, \ldots, c_n$ are distinct). Then

$$[T]_\beta = \begin{bmatrix} c_1 & & 0 \\ & \ddots & \\ 0 & & c_n \end{bmatrix},$$

so

$$[T^*]_\beta = [T]_\beta^* = \begin{bmatrix} \overline{c_1} & & 0 \\ & \ddots & \\ 0 & & \overline{c_n} \end{bmatrix}.$$

In particular, when $\mathbb{F} = \mathbb{R}$, we obtain that $[T]_\beta = [T^*]_\beta$, so T is Hermitian. When $F = \mathbb{C}$, we at least know that T, T^* commute, since they are simultaneously diagonalizable.

Def'n. Normal Operator
Let V be a finite dimensional inner product space and let $T : V \to V$ be linear. We say T is ***normal*** if T, T^* commute.

It is clear that any Hermitian operator is normal, and any unitary operator is also normal. We thus have the following statement.

Proposition 8.21. *Let V be a finite dimensional inner product space and let $T : V \to V$ be linear. If there exists an orthonormal eigenbasis β for V, then T is normal. In particular, when $\mathbb{F} = \mathbb{R}$, T is Hermitian.*

In fact, the fact that a linear operator is normal implies that it is diagonalizable. But first, we work with Hermitian operators, showing the existence of orthonormal eigenbasis. For the remaining part of the section, let V be a finite dimensional inner product space and let $T : V \to V$ be linear.

Proposition 8.22. *If T is Hermitian, then*

(a) every eigenvalue of T is real; and

(b) eigenvectors corresponding to distinct eigenvalues are orthogonal to each other.

Proof.

(a) Suppose that $c \in \mathbb{F}$ is an eigenvalue and let $v \in V$ be an eigenvalue corresponding to c. Then

$$c\langle v, v\rangle = \langle cv, v\rangle = \langle T(v), v\rangle = \langle v, T^*(v)\rangle = \langle v, T(v)\rangle = \langle v, cv\rangle = \overline{c}\langle v, v\rangle.$$

But $v \neq 0$, so it must be the case that $c = \overline{c}$, which means $c \in \mathbb{R}$.

(b) Suppose that $c, d \in \mathbb{F}$ are distinct eigenvalues and let $v, w \in V$ eigenvectors corresponding to c, d, respectively. Then

$$c\langle v, w\rangle = \langle T(v), w\rangle = \langle v, T^*(w)\rangle = \langle v, T(w)\rangle = \langle v, dw\rangle = \overline{d}\langle v, w\rangle = d\langle v, w\rangle.$$

Since c, d are distinct, it follows that $\langle v, w\rangle = 0$. ∎

Proposition 8.23. *If T is Hermitian, then T has an eigenvalue.*

Proof. Since every linear operator on a vector space over $\mathbb{C}$ has an eigenvalue, suppose that $\mathbb{F} = \mathbb{R}$. Let β be an orthonormal basis for V and let

$$A = [T]_\beta \in M_{n\times n}(\mathbb{R}) = M_{n\times n}(\mathbb{C}).$$

Let $\tilde{T} : \mathbb{C}^n \to \mathbb{C}^n$ be such that

$$\left[\tilde{T}\right]_\gamma = A,$$

where γ is the standard basis for $\mathbb{C}^n$. Let $c \in \mathbb{C}$ be an eigenvalue of $\tilde{T}$. Then

$$\left[\tilde{T}^*\right]_\gamma = \left[\tilde{T}\right]_\gamma^* = A^* = [T]_\beta^* = [T^*]_\beta = [T]_\beta = A = \left[\tilde{T}\right]_\gamma,$$

so $\tilde{T}$ is Hermitian. Now, by Proposition 8.22, $c \in \mathbb{R}$, which means c is a root of the characteristic polynomial of $\tilde{T}$. But

$$\text{characteristic polynomial of } \tilde{T} = \det(xI - A) = \text{characteristic polynomial of } T,$$

so c is a root of the characteristic polynomial of T. Thus c is an eigenvalue of T. ∎

Proposition 8.24.

If $W \subseteq V$ is T-invariant, then $W^\perp$ is T^*-invariant.

Proof. Let $v \in W^\perp$ and we desire to show that $T^*(v) \in W^\perp$. Then given any $w \in W$, we have

$$\langle T^*(v), w\rangle = \overline{\langle w, T^*(v)\rangle} = \overline{\langle T(w), v\rangle} = 0,$$

since $T(w) \in W$ by the T-invariance of W. ∎

Theorem 8.25.
Self-adjoint Operators Are Orthonormally Diagonalizable

If T is self-adjoint, then there is an orthonormal eigenbasis β for V.

Proof. We proceed inductively. We are given the result when $\dim(V) = 0$ freely. Let $v \in V$ be an eigenvector of T, which exists by Proposition 8.23. Since we desire to have an orthonormal basis, let

$$v_1 = \frac{v}{\|v\|}$$

and denote $W = \text{span}(v_1)$. If $W = V$, then we are done. So we may assume that $W \neq V$ (i.e. $\dim(V) > 1$). Since v_1 is an eigenvector of T, W is T-invariant. Therefore, by Proposition 8.24, $W^\perp$ is also T-invariant by the self-adjointness of T. Note that the restriction $T|_{W^\perp} : W^\perp \to W^\perp$ is also self-adjoint. But $W^\perp$ is of dimension $\dim(V) - 1$, so by an induction argument, $T|_{W^\perp}$ is orthogonally diagonalizable, say with respect to $(v_2, \ldots, v_n)$. Then v_1 is orthogonal to $v_2, \ldots, v_n$ by construction, so $\beta = (v_1, \ldots, v_n)$ is an orthonormal basis for V. Moreover, β diagonalizes T. ∎

Corollary 8.25.1.

If $\mathbb{F} = \mathbb{R}$, then T is self-adjoint if and only if there exists an orthonormal eigenbasis for V.

Proof. The forward direction is provided by Theorem 8.25 and the reverse direction is provided by (8.24). ∎

Corollary 8.25.2.

Suppose $A \in M_{n\times n}(\mathbb{F})$ is Hermitian. Then A is a unitarily equivalent to a diagonal matrix.

Proof. Let α be the standard basis for $\mathbb{F}^n$, which is an orthonormal basis with respect to the standard inner product on $\mathbb{F}^n$, and let $T : \mathbb{F}^n \to \mathbb{F}^n$ be such that T is represented by A with respect to α. Then

$$[T^*]_\alpha = [T]_\alpha^* = A^* = A = [T]_\alpha,$$

so $T^* = T$. So by Theorem 8.25, there exists an orthonormal basis β of eigenvectors of T for $\mathbb{F}^n$. So

$$D = [T]_\beta$$

is diagonal. That is,

$$D = P^{-1}AP,$$

where P is the change of basis matrix from α to β. But P is coming from a linear operator that takes an orthonormal basis α to another orthonormal basis β, so P is unitary. Thus A, D are unitarily equivalent, as claimed. ∎

We now consider orthonormal diagonalization of normal operators.

Def'n. Normal Matrix
We say $A \in M_{n\times n}(\mathbb{F})$ is ***normal*** if A, A^* commute.

Proposition 8.26. *Let β be an orthonormal basis for V. Then T is normal if and only if $[T]_\beta$ is normal.*

Proof. Observe that

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

Proposition 8.27. Properties of Normal Operators

Suppose that T is normal.

(a) For any $v \in V$, $\|T(v)\| = \|T^*(v)\|$.

(b) If $v \in V$ is an eigenvector of T corresponding to an eigenvalue c, then v is an eigenvector of T^* corresponding to $\overline{c}$.

Proof.

(a) Observe that

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

(b) Observe that

$$\begin{aligned}(cI - T)(v) = 0 &\iff \|(cI-T)(v)\| = 0 \\ &\iff \left\|(cI-T)^*(v)\right\| = 0 \iff \|(\overline{c}I - T^*)(v)\| = 0 \iff (\overline{c}I - T^*)(v) = 0.\end{aligned}$$ ∎

Proposition 8.28. *Let β be an orthonormal basis for V such that $[T]_\beta$ is upper triangular. Then T is normal if and only if $[T]_\beta$ is diagonal.*

Proof. The reverse direction is clear, since diagonal matrices commute. Conversely, suppose that T is normal. Denote $A = [T]_\beta$ and write $\beta = (v_1, \ldots, v_n)$. Then, since A is upper triangular,

$$T(v_1) = \sum_{j=1}^{n} A_{j1}v_1 = A_{11}v_1,$$

which means that v_1 is an eigenvector of T corresponding to A_{11}. Then from (a) of Proposition 8.27, v_1 is an eigenvector of T^* corresponding to $\overline{A_{11}}$ by normality of T. But $[T^*]_\beta = A^*$, so

$$T^*(v_1) = \sum_{j=1}^{n} A_{j1}^* v_j = \sum_{j=1}^{n} \overline{A_{1j}} v_j.$$

Therefore,

$$\overline{A_{11}}v_1 = \sum_{i=1}^{n} \overline{A_{1j}}v_j,$$

which means $\overline{A_{1j}} = 0$ for all $j \neq 1$. That is, we have shown that $A_{j1} = 0$ for all $j \neq i$, and by using this argument on $2, 3, \ldots, n$, we see that A is diagonal, as required. ∎

Proposition 8.29. *Consider the case which $\mathbb{F} = \mathbb{C}$. Then there exists an orthonormal basis β for V such that $[T]_\beta$ is diagonal.*

Proof. We proceed inductively. The result is clear when $\dim(V) = 1$. Now suppose that $\dim(V) > 1$. Since $\mathbb{F} = \mathbb{C}$, there exists an eigenvalue c for T^*, so let v be an eigenvector of T^* corresponding to c. This gives an T^*-invariant subspace $W = \operatorname{span}(v)$, so $W^\perp$ is T-invariant. Then by the induction hypothesis, there exists an orthonormal basis $\beta' = (v_1, \ldots, v_{n-1})$ such that $[T|_{W^\perp}]_{\beta'}$ is upper triangular. Now let

$$\beta = (v_1, \ldots, v_n),$$

where we define $v_n = \frac{v}{\|v\|}$. Then β is an orthonormal basis by construction. Moreover, we see that

$$[T]_\beta = \begin{bmatrix} [T|_{W^\perp}]_{\beta'} & * \\ 0 & * \end{bmatrix},$$

where the last column can be anything. But every entry of the last column is not below the main diagonal, $[T]_\beta$ is upper triangular, since we know that $[T|_{W^\perp}]_{\beta'}$ is upper triangular. ∎

Corollary 8.29.1. *If $\mathbb{F} = \mathbb{C}$ and T is normal, then there exists an orthonormal eigenbasis for V.*

Proof. This is a consequence of Proposition 8.28 and 8.29. ∎

Corollary 8.29.2. *If $\mathbb{F} = \mathbb{C}$, then T is normal if and only if T is orthonormally diagonalizable.*

Corollary 8.29.3. *Let $A \in M_{n\times n}(\mathbb{C})$ be normal. Then A is unitarily equivalent to a diagonal matrix.*

(EX 8.25) We showed that T is orthonormally diagonalizable if and only if T is self-adjoint when $\mathbb{F} = \mathbb{R}$. In fact, we can give a slightly more general statement. Show that T is self-adjoint if and only if T is an orthonormally diagonalizable operator with real eigenvalues.

This page intentionally left blank.

9.
Forms on Inner Product Space

Forms

(9.1) Fix a finite dimensional inner product space $(V, \langle \cdot, \cdot \rangle)$ throughout.

> **Def'n. Form** on an Inner Product Space
> A ***form*** on V is a function $f : V \times V \to \mathbb{F}$ with the following properties. Suppose that $v, u, w \in V$ and $c \in \mathbb{F}$ are given.
>
> (a) *linearity at the first argument*: $f(cv + w, u) = cf(v, u) + f(w, u)$.
>
> (b) *conjugate linearity at the second argument*: $f(v, cw + u) = \overline{c} f(v, w) + f(v, w)$.

(EX 9.2)

(a) $\langle \cdot, \cdot \rangle$ is a form.

(b) If $T : V \to V$ is a linear operator, then $f : V \times V \to \mathbb{F}$ defined by

$$f(v, w) = \langle T(v), w \rangle$$

is a form.

(9.3) If $\mathbb{F} = \mathbb{R}$, then forms are ***bilinear***. That is, forms are linear at each argument.

Theorem 9.1. *For every form $f : V \times V \to \mathbb{F}$, there exists a unique linear operator $T : V \to V$ such that $f(v, w) = \langle T(v), w \rangle$.*

Proof. Suppose that we are given $w \in V$ and consider the linear functional

$$f(\cdot, w) V \to \mathbb{F}.$$

We know that there exists a unique $w' \in V$ such that

$$f(v, w) = \langle v, w' \rangle$$

for all $v \in V$. So let $S : V \to V$ be the function such that $S(w) = w'$. That is,

$$f(\cdot, w) = \langle \cdot, S(w) \rangle$$

for all $w \in V$.

(a) We show that S is linear. Suppose that $v, w \in V$ and $c \in \mathbb{F}$ are given. Then

$$\langle \cdot, S(cv + w) \rangle = f(\cdot, cv + w) = \overline{c} f(\cdot, v) + f(\cdot, w) = \overline{c} \langle \cdot, S(v) \rangle + \langle \cdot, S(w) \rangle = \langle \cdot, cS(v) + S(w) \rangle.$$

But this means that the linear functional $f(\cdot, cv + w)$ can be written in two ways in terms of inner product fixing the second argument,

$$\langle \cdot, S(cv + w) \rangle = f(\cdot, cv + w) = \langle \cdot, cS(v) + S(w) \rangle.$$

But we know that such second argument is unique, so we conclude that $S(cv + w) = cS(v) + S(w)$, which means S is linear.

Now let $T = S^* : V \to V$. Then we have that, for all $v, w \in V$,

$$\langle T(v), w\rangle = \langle v, T^*(w)\rangle = \langle v, S(w)\rangle = f(v,w),$$

which is what we desired to construct. To show the uniqueness part, we show that we can *recover* T from f. Fix an orthonormal basis $\beta = (v_1, \ldots, v_n)$ for V. Then for any $v \in V$, we know that

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

In particular,

$$T(v_j) = \sum_{i=1}^{n} \langle T(v_j), v_i\rangle v_i.$$

Therefore, the Jth column of $[T]_\beta$ is

$$\begin{bmatrix} \langle T(v_j), v_1\rangle \\ \vdots \\ \langle T(v_j), v_n\rangle \end{bmatrix} = \begin{bmatrix} f(v_j, v_1) \\ \vdots \\ f(v_j, v_n) \end{bmatrix}.$$

That is,

$$\left([T]_\beta\right)_{ij} = (v_j, v_i),$$

so the matrix representation of T with respect to an orthonormal basis is completely determined by the form f and the basis. But this means that, if $U : V \to V$ is another linear operator such that $f(v,w) = \langle U(v), w\rangle$ for all $v, w \in V$, then T, U are represented by the same matrix with respect to the same basis, which means $T = U$, as desired. ∎

(9.4) Space of Forms

What Theorem 9.1 means is that we have a bijective function between $\mathcal{L}(V)$ and $\mathcal{F}(V)$, where we denote $\mathcal{L}(V)$ to be the vector space of linear operators on V and $\mathcal{F}(V)$ to be the set of forms on V. What we now desire to show is that $\mathcal{F}(V)$ is also a vector space, so this bijection is a natural isomorphism between $\mathcal{L}(V)$ and $\mathcal{F}(V)$. To do so, suppose that $f, g \in \mathcal{F}(V)$ and $c \in \mathbb{F}$. Then we can define addition by

$$(f+g)(v,w) = f(v,w) + g(v,w)$$

for all $v, w \in V$ and scalar multiplication by

$$(cf)(v,w) = cf(v,w)$$

for all $v, w \in V$. We can show that $\mathcal{F}(V)$ is a vector space over $\mathbb{F}$ with the following operations. That is, the function described in Theorem 9.1 is an isomorphism between $\mathcal{L}(V)$ and $\mathcal{F}(V)$.

(a) Notice that $0 \in \mathcal{F}(V)$ is such that $0(v,w) = 0$ for all $v, w \in V$.

(b) Since the identity in $\mathcal{L}(V)$ is the identity operator $I : V \to V$, the identity form in $\mathcal{F}(V)$ is the inner product, since we have that

$$\langle I(\cdot), \cdot\rangle = \langle \cdot, \cdot\rangle.$$

We can make defintions of forms by using this isomorphism.

Def'n. Hermitian Form

Let $f : V \times V \to \mathbb{F}$ be a form. We say f is ***Hermitian*** if

$$f(v,w) = \overline{f(w,v)}$$

for all $v, w \in V$.

Proposition 9.2. *Let $f : V \times V \to \mathbb{F}$ be a form and let $T_f : V \to V$ be the unique linear operator such that $\langle T(v), w\rangle = f(v,w)$ for all $v, w \in V$. Then f is Hermitian if and only if T is self-adjoint.*

Proof. Observe that

$$\begin{aligned} f \text{ is Hermitian} &\iff \forall v, w \in V \left[f(v,w) = \overline{f(w,v)} \right] \\ &\iff \forall v, w \in V \left[\langle T_f(v), w\rangle = \overline{\langle T_f(w), v\rangle} \right] \\ &\iff \forall v, w \in V \left[\langle v, T_f^*(w)\rangle = \langle v, T_f(w)\rangle \right] \\ &\iff \forall w \in W \left[T_f^*(w) = T_f^*(w) \right] \\ &\iff T_f \text{ is self-adjoint.} \end{aligned}$$

∎

Proposition 9.3. *Let $f : V \times V \to \mathbb{F}$ be a form. Then f is Hermitian if and only if $f(v,v) \in \mathbb{R}$ for all $v \in V$.*

Proof. For the forward direction, suppose that f is Hermitian. Then for any $v \in V$,

$$f(v,v) = \overline{f(v,v)},$$

which means $f(v,v) \in \mathbb{R}$. For the reverse direction, suppose that $f(v,v) \in \mathbb{R}$ for all $v \in V$. Fix $v, w \in V$. Then

$$f(v+w, v+w) \in \mathbb{R}.$$

So

$$f(v+w, v+w) = f(v,v) + f(v,w) + f(w,v) + f(w,w) \in \mathbb{R},$$

which means

$$f(v,w) + f(w,v) \in \mathbb{R}.$$

Hence

$$f(v,w) + f(w,v) = \overline{f(v,w) + f(w,v)} = \overline{f(v,w)} + \overline{f(w,v)}. \tag{9.1}$$

On the other hand,

$$f(v+iw, v+iw) \in \mathbb{R}$$

as well, and we have

$$f(v+iw, v+iw) = f(v,v) - if(v,w) + if(w,v) + f(iw, iw).$$

Therefore,

$$-if(v,w) + if(w,v) \in \mathbb{R}.$$

Hence

$$-if(v,w) + if(w,v) = \overline{-if(v,w) + if(w,v)} = i\overline{f(v,w)} - i\overline{f(w,v)}. \tag{9.2}$$

Notice that, when we combine the right hand side of [9.1] and i times the right hand side of [9.2], we obtain

$$2f(v,w) = 2\overline{f(w,v)}$$

for all $v, w \in V$, so f is Hermitian. ∎

Corollary 9.3.1. *Let $T : V \to V$ be linear. Then T is self-adjoint if and only if $\langle T(v), v\rangle \in \mathbb{R}$ for all $v \in V$.*

Proof. We know that there exists a unique form $f : V \times V \to \mathbb{F}$ such that

$$f(v,w) = \langle T(v), w\rangle$$

for all $v, w \in V$. Then

$$\begin{aligned} T \text{ is self-adjoint} &\iff f \text{ is Hermitian} \\ &\iff \forall v \in V\,[f(v,v) \in \mathbb{R}] \\ &\iff \forall v \in V\,[\langle T(v), v\rangle \in \mathbb{R}] \end{aligned}$$

∎

Forms on Vector Spaces

(9.5) We can talk forms in a more general setting: without inner products around. Fix a finite dimensional vector space V throughout.

Def'n. Form on a Vector Space
A ***form*** $f : V \times V \to \mathbb{F}$ is a sesquilinear function (i.e. linear at the first argument and conjugate linear at the second argument).

That is, the defition is identical to what we gave for inner product spaces, but we do not fix an inner product around. We also have seen that, once we fix an inner product (a *special form* indeed, so to speak) $\langle \cdot, \cdot\rangle : V \times V \to \mathbb{F}$, then there is a natural isomorphism between the space of forms and the space of linear operators on V.

Def'n. Matrix of a Form
Let $f : V \times V \to \mathbb{F}$ be a form and let β be a basis for V. Then the ***matrix*** of f with respect to β, $[f]_\beta$, is such that

$$\left([f]_\beta\right)_{ij} = f(v_j, v_i),$$

where $(v_1, \ldots, v_n) = \beta$.

Proposition 9.4. *Let $\langle \cdot, \cdot\rangle : V \times V \to \mathbb{F}$ be an inner product and if β is an orthonormal basis for V with respect to $\langle \cdot, \cdot\rangle$, then*

$$[f]_\beta = [T_f]_\beta,$$

where $T_f : V \to V$ is the unique linear operator such that $f(v,u) = \langle T_f(v), u\rangle$ for all $v, u \in V$.

Proof. To compute $[T_f]_\beta$, fix $j \in \{1, \ldots, n\}$, and notice that

$$T_f(v_j) = \sum_{i=1}^{n} \langle T_f(v_j), v_i\rangle v_i,$$

which means

$$\left([T_f]_\beta\right)_{ij} = \langle T_f(v_j), v_i\rangle = f(v_j, v_i) = \left([f]_\beta\right)_{ij}.$$

∎

Theorem 9.5. Principal Axis Theorem

If $f : V \times V \to \mathbb{F}$ is Hermitian, then there exists a basis β for V such that $[f]_\beta$ is real and diagonal. In fact, when an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ is fixed, we can choose β to be orthonormal with respect to the fixed inner product.

Proof. Observe that the second part of this theorem is stronger, so we prove that only. Fix an inner product $\langle \cdot, \cdot \rangle : V \to V \to \mathbb{F}$ on V. Then we know that there exists a unique $T_f : V \to V$ such that $\langle T_f(v), u \rangle = f(v,u)$ for all $v, u \in V$. First note the following.

(a) Observe that, given $v, w \in V$,

$$\langle T_f(v), w \rangle = f(v,w) = \overline{f(w,v)} = \overline{\langle T_f(w), v \rangle} = \langle v, T_f(w) \rangle .$$

That is, T_f is self-adjoint whenever f is Hermitian.

So by (a), T_f is orthonormally diagonalizable, say with respect to β. Moreover T_f is self-adjoint, $[T_f]_\beta$ is Hermitian, which means its diagonal entries are real. But $[T_f]_\beta$ is a diagonal matrix, it is a real matrix. By Proposition 9.4, $[T_f]_\beta = [f]_\beta$ since β is orthonormal, so $[f]_\beta$ is real and diagonal, as desired. ∎

(EX 9.6) Prove the converse of Theorem 9.5.

Proposition 9.6.

Let $f : V \times V \to \mathbb{F}$ be a form and let $\beta = (v_1, \ldots, v_n)$ be a basis for V. Then for any $v, w \in V$,

$$f(v,w) = [w]_\beta^* [f]_\beta [v]_\beta .$$

Proof. Observe that, given $v = \sum_{i=1}^n a_i v_i, w = \sum_{j=1}^n b_j v_j \in V$,

$$\begin{aligned} f(v,w) &= f\left(\sum_{i=1}^n a_i v_i, \sum_{j=1}^n b_i v_j \right) \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i \overline{b_j} f(a_i, b_j) = \sum_{i=1}^n \sum_{j=1}^n \overline{b_j} \left([f]_\beta \right)_{ji} a_i = [w]_\beta^* [f]_\beta [v]_\beta , \end{aligned}$$

as required. ∎

(9.7) We ask the following question: which forms are inner products? Note that, given a form $f : V \times V \to \mathbb{F}$, f is an inner product if it is Hermitian and positive-definite. We can determine whether or not f is an inner product by looking aht the matrix $[f]_\beta$.

Theorem 9.7.

Let $f : V \times V \to \mathbb{F}$ be a form and let β be a basis for V. Then f is an inner product if and only if $[f]_\beta = P^* P$ for some invertible $P \in M_{n \times n}(\mathbb{F})$.

Proof. Write $\beta = (v_1, \ldots, v_n)$ and $A = [f]_\beta$.

- ($\implies$) Suppose that f is an inner product. Then there exists an orthonormal basis $\gamma = (w_1, \ldots, w_n)$ for the inner product space (V, f). That is, $f(w_i, w_j) = \delta_{ij}$ (δ_{ij} is the Kronecker delta). Now let $Q \in M_{n \times n}(\mathbb{F})$ be such that

$$Q = \begin{bmatrix} [w_1]_\beta & \cdots & [w_n]_\beta \end{bmatrix}$$

which is a change of basis matrix, so Q is invertible. Now note that

$$\begin{aligned} (Q^* A Q)_{ij} &= [w_j]_\beta^* A [w_i]_\beta \\ &= f(w_i, w_j) && \text{by Proposition 9.6} \\ &= \delta_{ij}, \end{aligned}$$

which means Q^*AQ is the identity matrix. Thus we have

$$A = (Q^*)^{-1} I Q^{-1} = \left(Q^{-1}\right)^* Q^{-1},$$

so $P = Q^{-1}$ works.

- ($\Longleftarrow$) Conversely, suppose A is of the form $A = P^*P$ for some invertible $P \in M_{n\times n}(\mathbb{F})$. Fix $v, w \in V$ and notice that

$$\begin{aligned}
\overline{f(v,w)} &= \overline{[w]_\beta^* A [v]_\beta} && \text{by Proposition 9.6} \\
&= \overline{[w]_\beta^* P^* P [v]_\beta} \\
&= [w]_\beta^T P^T \overline{P} \overline{[v]_\beta} \\
&= \overline{[v]_\beta}^T \overline{P}^T P [w]_\beta \\
&= [v]_\beta^* P^* P [w]_\beta \\
&= f(w,v), && \text{by Proposition 9.6}
\end{aligned}$$

so f is Hermitian. Now assume that $v \neq 0$. Then

$$\begin{aligned}
f(v,v) &= [v]_\beta^* A [v]_\beta && \text{by Proposition 9.6} \\
&= [v]_\beta^* P^* P [v]_\beta \\
&= \left(P[v]_\beta\right)^* \left(P[v]_\beta\right) \\
&= \left\langle P[v]_\beta, P[v]_\beta \right\rangle \\
&> 0, && \text{since } v \neq 0 \text{ and } P \text{ is invertible}
\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product on $\mathbb{F}^{n\times 1}$. ∎