MATB24

Linear Algebra II

Lecture Notes

documented by Eric Wu

UNIVERSITY OF
TORONTO
SCARBOROUGH

Instructor(s):	**Thomas Kielstra**
Email:	thomas.kielstra@utoronto.ca
Office:	IA 4122
Textbook:	Hoffman, K. M., Kunze, R. (1971). *Linear Algebra.*

Contents

1 Fields and Ideals

1.1 Axioms of a Field

Axiom 1.1.1 (Field)**.** A *field*, denoted $\mathbb{F}$, is a set which has two binary operators

$$+_{\mathbb{F}} \text{ (addition) and } \times_{\mathbb{F}} \text{ (multiplication)}$$

such that if $a, b \in \mathbb{F}$, then

(a) $a +_{\mathbb{F}} b \in \mathbb{F}$ (closed under addition)

(b) $a \times_{\mathbb{F}} b \in \mathbb{F}$ (closed under multiplication)

The set must satisfy the following rules:

1. $\exists\, 0_{\mathbb{F}}$ such that $0_{\mathbb{F}} +_{\mathbb{F}} a = a$ for any $a \in \mathbb{F}$ (the additive identity exists)
2. $\forall\, a \in \mathbb{F},\ \exists - a \in \mathbb{F}$ such that $(-a) +_{\mathbb{F}} a = 0_{\mathbb{F}}$ (the additive inverse exists)
3. $\forall\, a, b, c \in \mathbb{F},\ a +_{\mathbb{F}} (b +_{\mathbb{F}} c) = (a +_{\mathbb{F}} b) +_{\mathbb{F}} c$ (Associative with addition)
4. $\forall\, a, b \in \mathbb{F},\ a +_{\mathbb{F}} b = b +_{\mathbb{F}} a$ (Commutative with addition)
5. $\exists\, 1_{\mathbb{F}}$ such that $1_{\mathbb{F}} \times_{\mathbb{F}} a = a$ for all $a \in \mathbb{F}$ (the multiplicative identity exists)
6. $\forall\, a \in \mathbb{F} \setminus \{0_{\mathbb{F}}\},\ \exists\, a^{-1} \in \mathbb{F}$ such that $a^{-1} \times_{\mathbb{F}} a = 1_{\mathbb{F}}$ (the multiplicative inverse exists)
7. $\forall\, a, b, c \in \mathbb{F},\ a \times_{\mathbb{F}} (b \times_{\mathbb{F}} c) = (a \times_{\mathbb{F}} b) \times_{\mathbb{F}} c$ (Associative with multiplication)
8. $\forall\, a, b \in \mathbb{F},\ a \times_{\mathbb{F}} b = b \times_{\mathbb{F}} a$ (Commutative with multiplication)
9. $\forall\, a, b, c \in \mathbb{F},\ a \times_{\mathbb{F}} (b +_{\mathbb{F}} c) = a \times_{\mathbb{F}} b +_{\mathbb{F}} a \times_{\mathbb{F}} c$ (Distributes over addition)

Remarks 1.1.1.0.1. Two common fields are $\mathbb{C}, \mathbb{Q}$ and $\mathbb{R}$. Also, two common nonexinclude $\mathbb{N}$ and $\mathbb{Z}$. Since for naturals, $A_{\mathbb{F}}4.$ is not satisfied; for integers, $A_{\mathbb{F}}5.$ is not satisfied. Furthermore, the order of the axioms matters.

Definition 1.1.2 (modulo n)**.** Let n be an integer such that $n \geq 2 \wedge k \in \mathbb{Z}$. Then,

$$k = qn + r$$

where $0 \leq r < n \wedge q \in \mathbb{Z}$. Now we define modulo by an integer n,

$$k \bmod n = r.$$

I.e., we define the integer remainder of k divided by n.

Ex.

i. $5 \bmod 7 = 5$.

ii. $-5 \bmod 7 = 2 \because -5 = (-1)7 + 2$.

iii. $12 \bmod 4 = 0$.

Remarks 1.1.2.0.1 (mod and long division)**.** Let $a \in Z^+$ be arbitrary. Consider an integer divsor d, quotient q and remainder r such that

$$a = qd + r.$$

Then, $a \bmod d = r$.

Ex. Compute $47 \bmod 3$ by long division.

Solution. Note that $\begin{array}{r} 15 \\ 3\overline{)47} \\ \underline{3} \\ 17 \\ \underline{15} \\ 2 \end{array}$. Thus, $47 \bmod 3 = 2$. ■

Remarks 1.1.2.0.2 (Extended long division and mod for negative numbers)**.** Let k be a negative integer. To find q and r s.t. $k = qn + r \wedge 0 \leq r < n$ we run the followings algorithm:

1. Apply long division to $|k|/n$ s.t. Q is the quotient and R is the remainder.

2. If $R = 0$, then

   (a) $q = -Q$

   (b) $r = R$

3. If $R \neq 0$, then

   (a) $q = -(Q + 1)$

   (b) $r = n - R$.

Ex. Compute $-56 \bmod 3$.

Solution. Consider $\begin{array}{r} 18 \\ 3\,\overline{)56} \\ \underline{3} \\ 26 \\ \underline{24} \\ 2 \end{array}$. By remark 1.1.4.0.2,

$$-56/3 = -(18+1)3 + 1 \implies -56 \bmod 3 = 1.$$

Or, note that

$$R = 2 \implies r = n - R = 3 - 2 = 1 \implies -56 \bmod 3 = 1.$$

■

Definition 1.1.3 (Congruent Modulo)**.** Let $n \geq 2$ be a fixed integer. Then, two integers m_1 and m_2 are *congruent modulo,* denoted

$$m_1 \cong m_2 \pmod{n} \iff m_1 - m_2 = kn \text{ s.t. } k \in \mathbb{Z}.$$

The integer n is called the *modulus* of the congruence.

Remarks 1.1.3.0.1 (congruence and remainder)**.** Let $a, b \in \mathbb{Z}$. Then,

$$a \cong b \pmod{n} \iff \text{ the remainder of } a/n = b/n.$$

Proposition 1.1.3.1 (Properties of congruent modulo - equivalent class)**.**

1. $A \cong A$ (Reflexivity)
2. $A \cong B \implies B \cong A$ (Symmetry)
3. $A \cong B \wedge B \cong C \implies A \cong C$ (Transitivity)

Definition 1.1.4 (Integers Modular $n\,(\mathbb{Z}_n)$)**.** Integers modulo $n, \mathbb{Z}_n$ is the set $\{0, \ldots, n-1\}$ with the two operators $+_{\bmod n} := \oplus$ and $\times_{\text{mod}_n} := \odot$. Let $a, b \in \mathbb{Z}_n$, then

$$a \oplus b = (a+b) \bmod n, \quad a \odot b = (ab) \bmod n.$$

Proposition 1.1.4.1 (properties of modular arithmetic)**.** Let $a, b \in \mathbb{Z}$.

1. $$\begin{aligned} ((a \bmod n) + (b \bmod n)) \bmod n &= (a + (b \bmod n)) \bmod n \\ &= ((a \bmod n) + b) \bmod n \\ &= (a + b) \bmod n. \end{aligned}$$

2. $$\begin{aligned} ((a \bmod n) \times (b \bmod n)) \bmod n &= (a \times (b \bmod n)) \bmod n \\ &= ((a \bmod n) \times b) \bmod n \\ &= (a \times b) \bmod n. \end{aligned}$$

Proof. Trivial, the idea is to define

$$a = q_1 n + r_1; b = q_2 n + r_2, r_1 + r_2 = q_3 n + r_3; r_1 r_2 = q_4 n + r_4.$$

And then apply definition repeatedly with algebraic rearangement of n with some $\mathbb{Z}$ coefficient. In particular, arithmetic equality $= r_3$ and multiplicative equality $= r_4$. □

Theorem 1.1.4.2 (Properties of a Field)**.** Let $\mathbb{F}$ be a field with $\boxplus_{\mathbb{F}}$ and $\boxdot_{\mathbb{F}}$. Then, the followings hold:

1. The additive identity $0_{\mathbb{F}}$ is unique.
2. The multiplicative identity $1_{\mathbb{F}}$ is unique.
3. For any $a \in \mathbb{F}$, the additive inverse $-a$ is unique.
4. For any $a \in \mathbb{F}$ such that $a \neq 0$, the multiplicative inverse a^{-1} is unique.
5. Addition is cancellational; For any $a, b, c \in \mathbb{F}$ if $a +_{\mathbb{F}} b = a +_{\mathbb{F}} c$, then $b = c$.
6. Given $a \neq 0_{\mathbb{F}}$, multiplication is cancellational;
   For any $a, b, c \in \mathbb{F}$ if $a \times_{\mathbb{F}} b = a \times_{\mathbb{F}} c$, then $b = c$.
7. Addition is distributive over Multiplication.
   $(a + b)c = ac + bc$.
8. If $a, b \in \mathbb{F}$ and $a * b = 0$, then $a = 0$ or $b = 0$.

Ex. Show that $\mathbb{Z}_6$ is not a field.

Proof. Note that $3 \times_{\text{mod}\,6} 5 = 3$ and $3 \times_{\text{mod}\,6} 1 = 3$ thus the multiplicative identity is *not* unique. Thus, $\mathbb{Z}_6$ is not a field. □

Ex. Show that $\mathbb{Z}_p$ is a field iff p is a prime.

Proof. First we show that p is not a prime $\implies$ $\mathbb{Z}_p$ is not a field. Sps p is not a prime. Then, p is composite. Hence $p = ab$ s.t. $a, b \in \mathbb{Z}_p - \{0, 1\}$. □

Definition 1.1.5 (Subfield)**.** Let $S \subseteq \mathbb{F}$. Then, S is a subfield if S is a field.

Theorem 1.1.5.1 (Field Closure)**.** Let $S \subseteq \mathbb{F}$. Then, S is a subfield iff

1. S contains at least 2 elements.
2. $\forall a, b \in S, a + (-b) \in S$. closed under additive inverse
3. $\forall a, b \in S, b \neq 0 \implies a \times b^{-1} \in S$ closed under multiplicative inverse.

1.2 Rings, Commutative Rings, Integral Domains, Principal Ideal Domains

Axiom 1.2.1 (Ring)**.** A set $\mathcal{R}$ is a ring with additive ($+_{\mathcal{R}}$) and multiplicative ($\times_{\mathcal{R}}$) operators such that

(a) If $a, b \in \mathcal{R}$, then $a +_{\mathcal{R}} b \in \mathcal{R}$ (Closed under addition)

(b) If $a, b \in \mathcal{R}$, then $a \times_{\mathcal{R}} b \in \mathcal{R}$ (Closed under multiplication)

1. $\exists\, 0_{\mathcal{R}}$ such that $0_{\mathcal{R}} +_{\mathcal{R}} a = a$ for all $a \in \mathcal{R}$ (The additive identity exists)
2. $\forall\, a \in \mathcal{R}$, $-a \in \mathcal{R}$ exists such that $(-a) +_{\mathcal{R}} a = 0_{\mathcal{R}}$ (The additive inverse exists)
3. $\forall\, a, b, c \in \mathcal{R},\ a +_{\mathcal{R}} (b +_{\mathcal{R}} c) = (a +_{\mathcal{R}} b) +_{\mathcal{R}} c$ (Associative with addition)
4. $\forall\, a, b \in \mathcal{R},\ a +_{\mathcal{R}} b = b +_{\mathcal{R}} a$ (Commutative with addition)
5. $\exists\, 1_{\mathcal{R}}$ such that $1_{\mathcal{R}} \times_{\mathcal{R}} a = a$ for any $a \in \mathcal{R}$ (The multiplicative identity exists)
6. $\forall\, a, b, c \in \mathcal{R},\ a \times_{\mathcal{R}} (b \times_{\mathcal{R}} c) = (a \times_{\mathcal{R}} b) \times_{\mathcal{R}} c$ (Associative with multiplication)
7. $\forall\, a, b, c \in \mathcal{R},\ a \times_{\mathcal{R}} (b +_{\mathcal{R}} c) = (a \times_{\mathcal{R}} b) +_{\mathcal{R}} (a \times_{\mathcal{R}} c)$ (Distributes over addition)

Definition 1.2.2 (Commutative ring)**.** Let $(\mathcal{R}, \times, +)$ be a ring. If the multiplication operator is commutative, then $(\mathcal{R}, \times, +)$ is a *commutative ring.*

Definition 1.2.3 (Integral Domain (non-zero commutative ring))**.** Let $(\mathcal{R}, \times, +)$ be a ring. If the multiplication operator is non-zero, then $(\mathcal{R}, \times, +)$ is an *integral domain.* Note: non-zero operator means that $a \times b = 0 \implies a = 0 \vee b = 0$. This condition is equivalent to multiplication being cancellational.

Definition 1.2.4 (Principal Ideal Domain)**.** Let $(\mathcal{R}, \times, +)$ be an *integral domain.* If every ideal of $\mathcal{R}$ is a principal ideal, then $(\mathcal{R}, \times, +)$ is a *principal ideal domain.*

Definition 1.2.5 (Field)**.** Let $(\mathcal{R}, \times, +)$ be a principal ideal domain. If

$$\forall a \in \mathcal{R}, \exists a^{-1} \text{ s.t. } a \cdot a^{-1} = 1,$$

then $(\mathcal{R}, \times, +)$ is a *field.*

1.3 Rings, Additive Subgroups, Ideals, Principal Ideals

Definition 1.3.1 (Additive Subgroup)**.** Let $(\mathcal{R}, \times, +)$ be a ring. Let $S \subseteq \mathcal{R}$. Then, $(S, +)$ is an *additive subgroup* if

1. $\forall a, b \in S, a + b \in S$
2. $0_{\mathcal{R}} \in S$
3. $\forall a \in S, \exists -a \in S$ s.t. $a +_{\mathcal{R}} (-a) = 0$
4. $\forall a, b, c \in S, a + (b + c) = (a + b) + c.$

Definition 1.3.2 (Ideals)**.** Let $(\mathcal{R}, \times, +)$ be a ring. Then a subgroup $(I, +)$ is a *left ideal* if

$$\forall r \in \mathcal{R} \text{ and } x \in I, rx \in I.$$

Else, subgroup $(I, +)$ is a *right ideal* if

$$\forall r \in \mathcal{R} \text{ and } x \in I, xr \in I.$$

Remarks 1.3.2.0.1. Since we are dealing with commutative rings, we will refer to two sided ideals as ideals.

Ex. Show that the even numbers form an ideal within the integers.

Solution. Let $x \in \{\text{even numbers}\}$ and $k \in \mathbb{Z}$. Then, $x \in 2h$ s.t. $h \in \mathbb{Z}$. Thus $kx = k(2h) = 2(kh) \in \{\text{even numbers}\}$. ■

Definition 1.3.3 (Polynomials over a field $\mathbb{F}$)**.** The set of polynomials over a field $\mathbb{F}, (\mathbb{F}[x])$ is the set of polynomials with coefficients from $\mathbb{F}$. If $p(x) \in \mathbb{F}[x]$, then

$$p(x) = \sum_{i=0}^{n} a_i x^i \text{ s.t. } a_i \in \mathbb{F}, \forall i.$$

Definition 1.3.4 (Generator for Ideal and principal ideal)**.** Let $(\mathcal{R}, \times, +)$ be a ring and $I \subseteq \mathcal{R}$ be an ideal. Then, $G \subseteq I$ is a *generator* for I, if

$$\forall i \in I, \exists g_1, \dots, g_n \in G \text{ and } r_1, \dots, r_n \in \mathcal{R} \text{ s.t. } i = \sum_{i=1}^{n} r_i g_i.$$

If $\exists g \in I$ s.t. $G = \{g\}$ is a generator, I is a *principal ideal*

Remarks 1.3.4.0.1 (We can equivalently define generator for ideal as follows)**.** Let $(\mathcal{R}, \times, +)$ be a ring and $I \subseteq \mathcal{R}$ be an ideal. The $G \subseteq I$ is a generation if

$$\forall S \subseteq I, S \text{ is an ideal and } S \neq \{0\} \implies G \subseteq S.$$

Ex. Consider the set $S \subseteq \mathbb{Z}_7\left[x\right]$ s.t. if $f \in S$, then $f(4) = 0$. Show that S is an ideal and that $(x+3)$ generates S.

Theorem 1.3.4.1 (Fundemental Remainder Theorem)**.** Let $p(x)$ in $\mathbb{F}[x]$ be a nonconstant polynomial such that $p(a) = 0$. Then there exists $q(x)$ in $\mathbb{F}[x]$ such that

$$p(x) = (x + (-a))(q(x))$$

Proof. Suppose $f \in \mathbb{Z}_7\left[x\right]$ and $g \in S$. Then $g(4) = 0$. Consider

$$(fg)(4) = f(4)g(4) = f(4)0 = 0.$$

Thus $(fg) \in S$. ? Let $g \in \mathbb{Z}_7\left[x\right]$ s.t. $g(x) = 0, \forall x$. Thus, $g(4) = 0$. Thus, $g \in S$. ?
Now we show $\{(x+3)\}$ is a generator. Let $f \in S$. Then, $f(4) = 0$.
By fundamental remainder theorem,

$$\begin{aligned} \exists G \in \mathbb{Z}_7\left[x\right] \text{ s.t. } f(x) &= (x + (-4)) \cdot g(x) \\ &= (x+3) \cdot g(x) && \text{as 3 is the additive inverse of 4 in } \mathbb{Z}_7. \end{aligned}$$

□

2 Vector Spaces

Axiom 2.0.1 (Vector Space)**.** A vector space is a set V with two binary operations:

$$\boxplus : V \times V \to V \quad \boxdot : \mathbb{F} \times V \to V$$

called vector addition and vector multiplication such that the following axioms hold:

$A_{\mathbb{V}}1.$ $\forall a, b \in V, a + b \in V$ closed under addition.

$A_{\mathbb{V}}2.$ $\forall c \in \mathbb{F}$ and $a \in V, ca \in V$ closed under multiplication.

$A_{\mathbb{V}}3.$ $\boxplus$ is *associative.*

$A_{\mathbb{V}}4.$ $\boxplus$ is *commutative.*

$A_{\mathbb{V}}5.$ There is an *additive element* $\vec{0} \in V$ s.t.

$$\vec{0} \boxplus \vec{x} = \vec{x}$$

for all $\vec{x} \in V$. We call $\vec{0}$ the *zero vector* of V.

$A_{\mathbb{V}}6.$ For each $\vec{x} \in V$, there is an additive inverse $\vec{x}' \in V$ such that

$$\vec{x} \boxplus \vec{x}' = \vec{0}.$$

$A_{\mathbb{V}}7.$ For all $\vec{x}$ and $\vec{y}$ and $c \in \mathbb{F}$, scaling by c *distributes over* addition:

$$c \boxdot (\vec{x} \boxplus \vec{y}) = c\vec{x} \boxplus c\vec{y}.$$

$A_{\mathbb{V}}8.$ For all $\vec{x} \in V$ and $c, d \in \mathbb{F}$, the *field addition* in $\mathbb{F}$
distributes scalar multiplication in V:

$$(c + d) \boxdot \vec{x} = (c \boxdot \vec{x}) \boxplus (d \boxdot \vec{x}).$$

$A_{\mathbb{V}}9.$ For all $\vec{x} \in V$ and $c, d \in \mathbb{F}, \boxdot$ associates with scalar multiplication on $\mathbb{F}:$

$$(cd) \boxdot \vec{x} = c \boxdot (d \boxdot \vec{x}).$$

$A_{\mathbb{V}}10.$ For all $\vec{x} \in V$, we have $1 \boxdot \vec{x} = \vec{x}$, the 1 is the *multiplicative identity* from $\mathbb{F}$.

Definition 2.0.2 (Subspaces)**.** Let $(V, +, \times)$ be a vector space. Then $U \subseteq V$ is a subspace if $(U, +, \times)$ is a vector space.

Theorem 2.0.2.1. Let V be a vector space. Then a subset U of V is a subspace if and only if U satisfies

1. $U \neq \emptyset$.
2. $\forall u, v \in U, u + v \in U$
3. $\forall c \in \mathbb{F}$ and $u \in U, cu \in U$.

Proposition 2.0.2.2 (Subspace as Restriction)**.** Let U be a non-empty subset of V. Then the set U is a subspace of V over the field $\mathbb{F}$ if

1. $a + b \in U, \forall a, b \in U$
2. $ca \in U, \forall a \in \mathbb{F}, a \in U$

This is what is meant when we say that addition and scalar multiplication is *restricted* from V to U

Proposition 2.0.2.3. If U is a subspace of V and W is a subspace of U, then W is a subspace of V.

Definition 2.0.3 (Sums, Intersection and Unions)**.** Let $S_1, \dots, S_n$ be a collection of sets. Then,

1. $S_1 \cup S_2 \cup \cdots \cup S_n = \bigcup_{i=1}^n S_i = \{x \text{ s.t. } \exists i \in \{1, \dots, n\}, x \in S_i\}$
2. $S_1 \cap S_2 \cup \cdots \cap S_n = \bigcap_{i=1}^n S_i = \{x \text{ s.t. } \forall i, x \in S_i\}$
3. $S_1 + S_2 + \cdots + S_n = \sum_{i=1}^n S_i = \{x = \sum_{i=1}^n a_i \text{ s.t. } a_i \in S_i\}$

Remarks 2.0.3.0.1. When dealing with sets we are used to looking at the *union* and *intersections* of the sets. For vector spaces we will look at the *sum* and *intersection* of subspaces.

2.1 Linear Comb., Spans, L.I., Bases

Definition 2.1.1 (Linear Comb.)**.** An element $\vec{v} \in V$ is a *linear combination of* $v_i, \forall i \in [1, n] \cap \mathbb{N}$ if there exists $a_1, \ldots, a_n \in \mathbb{F}$ st

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

Definition 2.1.2 (Span)**.** Let V be a vector space and S be a subset of V. The *span* of a set S is the set of all vectors in V that can be made using a *linear combination* of vectors in S.

$$\text{Span}(S) := \{\text{all possible linear combinations of } S\} = \left\{ v \in V \text{ s.t. } v = \sum_{i=1}^{n} a_i v_i \right\}.$$

Note, we use $\text{Span}_{\mathbb{F}}$ to emphasize the field on which the vector space is defined.

Proposition 2.1.2.1. Let V be a vector space and $v_1, \cdots, v_n \in V$. Then the span$\{v_1, \cdots, v_n\}$ is a subspace of V and is contained in all the subspaces of V that contain $v_1, \cdots, v_n$.

Definition 2.1.3 (Generator set/ Spanning Sets)**.** Let V be a vector space. If the span$(v_1, \cdots, v_k) = V$, we say that the set $\{v_1, \cdots, v_k\}$ generates the vector space V. The set $\{v_1, \cdots, v_k\}$ is a generating set (also known as a spanning set) of V.

Remeber that $\{v_1, \ldots, v_k\}$ is a set of vectors; whereas $\text{Span}(\{v_1, \ldots, v_k\})$ is the set of all possible L.C. of $\{v_1, \ldots, v_k\}$. *They are different.*

Proposition 2.1.3.1. Let V be a vector space, W_1 and W_2 be subspaces of V. Let S_1 be a generating set of W_1 and S_2 be a generating set of S_2. Then $S_1 \cup S_2$ is a generating set of $W_1 + W_2$.

Definition 2.1.4 (linearly independent)**.** Let V be a vector space. Then a set of vector $v_1, \ldots, v_k \in V$ are *linearly independent* iff

$$\forall a \in \mathbb{F}, \forall v \in V, \sum_{i=1}^{n} a_i v_i = 0$$

This is equivalent to saying that for any $v_i \notin \text{Span}(\{v_1, \ldots, v_k\}) \forall i$, if $\exists i$ s.t. $v_i \in \text{Span}(\{v_1, \ldots, v_k\})$ then the vectors $v_1, \ldots, v_k$ are *linearly dependent.*

Definition 2.1.5 (basis)**.** A set S is a basis of V if S is a set of *linearly independent vectors that generates V.*

Proposition 2.1.5.1. Suppose $\{v_1, \dots, v_n\}$ is a basis for V. Then, for arbitrary $v \in V, !$ scalars $a_1, \dots, a_k \in \mathbb{F}$ s.t.

$$v = a_1 v_n + \cdots + a_n v_n.$$

Definition 2.1.6 (dimension)**.** Let V be a vector space such that $\{v_1, \dots, v_n\}$ is a basis for V. Then, the dimension of V,

$$\dim(V) = |\{v_1, \dots, v_n\}| = n.$$

Definition 2.1.7 (cardinality of a set)**.** Let S be a set. Then $|S|$ is the *cardinality* of the set S s.t.

$$|S| = \text{ the number of elements in } S.$$

Definition 2.1.8 (Coordinate Vectors)**.** Let V be a vector space and $\beta = \{v_1, \dots, v_n\}$ be a basis for V, the scalars $a_1, \dots, a_n$ that satisfies

$$v = a_1 v_1 + \cdots + a_n v_n$$

are called the *coordinate of* v in this basis β. The coordinate vector is denoted

$$[v]_\beta = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{F}^{1\times n}.$$

Ex. The following vectors form a basis for the vector space $\mathbb{F}^{n\times 1}$.

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

Proof. Let $v \in \mathbb{F}^n$. Then $v = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a_1 e_1 + \cdots + a_n e_n$.

Thus, $v \in \text{Span}\left(\left\{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}\right\}\right)$. #excercise: show independence. □

Proposition 2.1.8.1. Suppose V is a finite vector space. Then,

- V has a basis which can be written and all bases of V has the same cardinality.
- Every generating set contains a basis
- For all linearly independent set, there is a basis such that the LI set is a subset of the basis
- All subspaces have a basis.

Definition 2.1.9 (Dimension II.)**.** The *cardinality* of the basis for a vector space is the *dimension* of the vector space, denoted $\dim(V)$.

Proposition 2.1.9.1. If U_1 and U_2 are finite dimensional subspaces of V, then

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

Proof. Outline: 1. create a basis for $U_1 \cap U_2(\alpha)$. 2. Extend said basis to create a basis for $U_1(\beta_1), U_2(\beta_2)$. 3. $\beta_1 \cup \beta_2$ is a basis for $U_1 + U_2$ (But how?) □

Ex. Let $\alpha = (1,1,1,0)^T$, $\beta = (0,1,1,1)^T$, $\gamma = (1,0,0,1)^T$. We can see that $\alpha, \beta, \gamma \in \mathbb{R}^4$ and $\alpha, \beta, \gamma \in (\mathbb{Z}_2)^4$. Show that α, β, γ are linearly independent in $\mathbb{R}^4$, but linearly dependent in $(\mathbb{Z}_2)^4$.

Proof. Consider a linear combination of α, β, γ equal to zero.

$$x_1(1,1,1,0)^T + x_2(0,1,1,1)^T + x_3(1,0,0,1)^T = 0 \Longrightarrow \begin{cases} x_1 + x_3 = 0 \\ x_1 + x_2 = 0 \\ x_1 + x_2 = 0 \\ x_2 + x_3 = 0 \end{cases}$$

For, $\mathbb{R}^n$

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right]$$

For $(\mathbb{Z}_2)^4$

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right]$$

□

2.2 Direct Sum

Definition 2.2.1 (Direct Sum)**.** The sum $V_1 + \cdots + V_n$ is a *direct sum* if the only choices of vectors $v_1 \in V_1, v_2 \in V_2, \ldots, v_n \in V_n$ s.t.

$$v_1 + \cdots + v_n = 0$$

is $v_1 = \cdots = v_n = 0$. We denote the direct sum as

$$V_1 \oplus \cdots \oplus V_n$$

Proposition 2.2.1.1. Let V be a finite dimension vector space. Then $V = V_1 \oplus \cdots \oplus V_n$ iff

1. $V = V_1 + \ldots V_n$.
2. $\forall i, j \in \{1, \ldots, n\}$ s.t. $i \neq j, V_i \cap V_j = \{0\}$.
3. $\dim(V) = \dim(V_1) + \cdots + \dim(V_n)$

Definition 2.2.2 (Ordered Basis)**.** A basis listed in a specific order is called an *ordered basis.*

Proposition 2.2.2.1. Suppose that $V = V_1 \oplus \cdots \oplus$ is finite dimensional and $\beta_i = \{v_{i1}, \ldots, v_{in_i}\}$ is a basis forall V_i. Then, the *ordered set*

$$S = \{v_{11}, \ldots, v_{1n_1}, v_{21}, \ldots, v_{2n_2}, \ldots, v_{n1}, \ldots, v_{nn_n}\}$$

obtained by *concatenating* $\beta_1, \ldots, \beta_n$ is a basis of V.

3 Linear Transformation

Definition 3.0.1. Let V and W be vector spaces defined over the field $\mathbb{F}$. Then, $T : V \to W$ is a *linear transformation* if T is a function such that

1. $T(x + y) = T(x) + T(y), \forall x, y \in V$
2. $T(cx) = cT(x), \forall c \in \mathbb{F}, x \in V$.

Definition 3.0.2 ($\mathcal{L}(V, W)$)**.** The set $\mathcal{L}(V, W)$ is the set of all linear transformations from V to W.

Definition 3.0.3 (linear operator)**.** A L.T., $T : V \to V$ is called a *linear operator.*

Definition 3.0.4 ($\mathcal{L}(V)$)**.** The set $\mathcal{L}(V)$ is the set of all linear operators of V.

Ex. Suppose $T : \mathbb{R} \to \mathbb{R}$ and T is a linear transformation. What can T be?

Solution. First observe that $T(cx) = cT(x); T(x) = xT(1) = kx$ s.t. $k \in \mathbb{R}$.
Thus, $T(x) = kx, k \in \mathbb{R}.T(x) = kx + 0$, i.e., linear function s.t. $y-$intercept equals to 0.

Remarks 3.0.4.0.1. Linear Function are a *very restricted* set of function. But they have *very nice* properties.

■

Ex. Let $V = W = \mathbb{R}$. Find the linear operations $T : V \to W$.

Solution. $T(x) = xT(1) = T(1)x$. Then, $T(x)$ is a line through the origin. ■

Exercise. Find eigenvalue of T, for $T \in \mathcal{L}(\mathbb{R})$.

Solution. $T(1)$. ■

Ex. Quantum Mechanics Ex.

In quantum analysis, vector $\alpha \in V$ are called *quantum bit* or *bit* for short. Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The matrix A is known as the *bit flip* operation. Let $\alpha = \begin{pmatrix} x \\ y \end{pmatrix} \in V$.

Define $T_A(\alpha) = A\alpha, i.e.,$

$$T_A \begin{bmatrix} x \\ y \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$$

If we consider out standard basis $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, we can notice that

$$T_A(e_1) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = e_2 \text{ and } T_A(e_2) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = e_1.$$

Definition 3.0.5 (Identity Operator)**.** The function $I : V \to V$ defined by $I(\alpha) = \alpha, \forall \alpha \in V$ is called the *identity operator*

Definition 3.0.6 (Zero Operator)**.** The function $O : V \to V$ defined by $O(\alpha) = \vec{0}, \forall \alpha \in V$ is called the *zero operator*

Ex. Let V be the space of all continuous functions $f : \mathbb{R} \to \mathbb{R}$. Let $T : V \to V$ such that

$$(Tf)(x) = \int_a^x f(t)dt.$$

Then, T is a linear operator called *integration operator.*

Proof. Let $f, g \in V, c \in \mathbb{R}$. WTS $T(cf + g) = cT(f) + T(g)$. Let $x \in \mathbb{R}$ be arbitrary. Then,

$$\begin{aligned} brT(f+g)(x) &= T(f+g)(f) \\ &= \int_a^x (f+g)(t)dt \\ &= \int_a^x f(t) + g(t)dt \\ &= \int_a^x f(x)dt + \int_a^x g(t)dt \\ &= (Tf)(x) + (Tg)(x) \qquad \text{proof of scaling left as exercise.} \end{aligned}$$

□

Ex. Let A be a fixed $m \times n$ matrix with entries in a field $\mathbb{F}$. Let $T_A : \mathbb{F}^{n\times 1} \to \mathbb{F}^{m\times 1}$ s.t.

$$T_A(x) = Ax,$$

where A is a real $m \times n$ matrix and $X \in \mathbb{F}^{n\times 1}$. Show that T_A is a linear transformation.

Proof. Let $c \in \mathbb{F}$ and $X, Y \in \mathbb{F}^{n\times 1}$.

$$\begin{aligned} T_A(X+Y) &= A(X+Y) \\ &= AX + AY \\ &= T_A(x) + T_A(Y). \end{aligned}$$

COnfirm if $n \times 1$ is correct. □

Ex. Rotations and Reflections in $\mathbb{R}^2$.

1. $T :=$ reflection in a line through the origin.

$$T(x,y) = (y,x).$$

   Check T_A, where $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then,

$$T_A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}.$$

2. Let $0 \leq \theta < 2\pi . T :=$ a counterclockwise rotation through the angle θ.

# fill up the notes!!

Remarks 3.0.6.0.1 (A is a orthogonal matrix)**.** Let $A^T :=$ transpose of A. Then,

$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \implies A^T = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Then, $A^T A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Exercise. Let V be the space of complex polynomials $f : \mathbb{C} \to \mathbb{C}$ given by

$$f(z) = a_0 + a_1 z + \cdots + a_n z^n, \quad \text{where } a_i \in \mathbb{C}, \forall i.$$

Let $Df : \mathbb{C} \to \mathbb{C}$ be the polynomial

$$(Df)(z) = a_1 + 2a_2 z + \cdots + na_n z^{n-1}$$

Then $D : V \to V$ is a linear operator called the *differentiation operator*.

Show that $(Df)(x)$ is a linear operator. **Exercise.**
Let $\alpha \in \mathbb{C}^n$ be a non-zero vector. Then

$$P_\alpha(v) = \frac{\langle v, \alpha \rangle}{\langle \alpha, \alpha \rangle} \alpha$$

is called the projection of v onto α. Show that $P_\alpha(v)$ is a linear transformation.
# fill up notes!!!!

Proposition 3.0.6.1 (Linear transformations "preserve" the additive identity)**.** If T is a linear transformation, then $T(0_v) = 0_w$.

Proof. Observe that

$$\begin{aligned}
0_v &= 0_v + 0_v \\
&\implies T(0_v) = T(0_v + 0_v) \\
&\implies T(0_v) = T(0_v) + T(0_v) && \text{as } T \text{ is linear} \\
&\implies T(0_v) + a = T(0_v) + T(0_v) + a && \text{for } a := \text{the additive inverse of } T(0_v) \\
&\implies T(0_v) + a = T(0_v) + (T(0_v) + a) \\
&\implies 0_w = T(0_v) + 0_w \\
&\implies 0_w = T(0_v)
\end{aligned}$$

□

Exercise. Similar to fields and vector spaces, linear transformations have a lot of nice properties. **Prove the following:**

1. The set $\mathcal{L}(V, W)$ forms a vector space.
2. Let $v \in V$ and $T \in \mathcal{L}(V, W)$. Then $-T(v) = T(-v) = T((-1)v) = (-1)T(v)$.
3. Let $S \in \mathcal{L}(V, W)$ and $T \in \mathcal{L}(U, V)$. Then $TS := T \circ S$ is in $\mathcal{L}(U, W)$.

Theorem 3.0.6.2 (ordered basis)**.** Let $\{\alpha_1, \dots, \alpha_n\}$ be an *ordered basis* for V, and let $\{\beta_1, \dots, \beta_n\}$ be vectors in W. Then *exactly one* linear transformation $T : V \to W$ **exists** such that:

$$T(\alpha_j) = \beta_j \quad \text{for } j = 1, 2, \dots, n.$$

This theorem is also useful in constructing linear transformations in the following manner:
Suppose you have an ordered basis $\{\alpha_1, \dots, \alpha_n\}$ and you need a linear transformation T such that $T\alpha_j = \beta_j$, for $j = 1, 2, \dots, n$. By applying this theorem, we can produce such a T.

Ex. Let us, once again, consider our rotation matrix, using our *standard basis*:

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Recall that $A = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$.

Solution. $T_A(e_1) = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} = \cos(\theta)e_1 + \sin(\theta)e_2$. And $T_A(e_2) = \cdots = -\sin(\theta)e_1 + \cos(\theta)e_2$. Note that

$$A = \left[[T_A(e_1)]_{\{e_1,e_2\}} \mid [T_A(e_2)]_{\{e_1,e_2\}} \right].$$

■

Definition 3.0.7 (nullspace, nullity, range, rank)**.** Let $T : V \to W$ be a linear transformation.

- The *nullspace* of T is the set $\{v \in V \mid T(v) = 0_W\}$.
- The *nullity* of T is defined as $\text{nullity}(T) := \dim(\text{nullspace}(T))$.
- The *range* of T, denoted $\text{range}(T)$, is the set $\{w \in W \mid w = T(v) \text{ for some } v \in V\}$.
- The *rank* of T is defined as $\text{rank}(T) := \dim(\text{range}(T))$.

Exercise. Show that nummspace of T is a *subspace* of V, while the range of T is a subspace of W.

Theorem 3.0.7.1 (Rank/Nulity Theorem)**.** For $T : V \to W$,

$$\text{rank}(T) + \text{nullity}(T) = \dim(V)$$

Exercise. Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be a swap operator.

$$T(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_n) = (x_1, \dots, x_{i-1}, x_j, x_{i+1}, \dots, x_{j-1}, x_i, x_{j+1}, \dots, x_n)$$

Show that T is a linear operator. Compute the *rank*(T) and *nullity*(T) and verify the rank-nullity theorem.

Remarks 3.0.7.1.1 (Permutation Notation.)**.** In abstract algebra, and later in the course, we often construct permutations using the following notation:

(ijk) means that the ith entry moves to the jth entry, the jth entry moves to the kth entry, and the kth entry moves to the ith entry.

Using this notation, *(ij)* denotes the swap operator T from the previous ex.

This notation is also sometimes referred to as *cycle decomposition*.

3.1 Algebra of Linear Transformations

Theorem 3.1.0.1. Let V and W be vector spaces. Suppose $S, T : V \to W$ are *linear transformations*. Any *linear combination* of S and T is also a linear transformation mapping V to W. That is,

$$\forall a, b \in \mathbb{F}, aS + bT : V \to W.$$

Let $v \in V$, then

$$(aS + bT)(v) = aS(v) + bT(v).$$

Remark: thus there are many different *linear transformation* from a vector space V to a vector space W.

Definition 3.1.1. The set $\mathcal{L}(V, W)$ is the set of *all linear transformations* from V to W. $\mathcal{L}(V, W)$ is itself a *vector space*. The set of *linear operators* that can act on the vector space V are denoted $\mathcal{L}(V) := \mathcal{L}(V, V)$. The *elements* of $\mathcal{L}(v)$ can be *multiplied* by composition: Let $S, T \in \mathcal{L}(V)$. Then,

$$ST \in \mathcal{L}(V) \text{ s.t. } ST(V) = S \circ T(v) = S(T(v)).$$

Ex. Let $A, B \in \mathbb{C}^{m\times m}$. Then, $\forall X \in \mathbb{C}^{n\times 1}$, $T_B T_A = T_{BA}$.
Let $x \in \mathbb{C}^{n\times 1}$. Then,

$$\begin{aligned} T_B T_A(x) &= T_B(T_A(x)) \\ &= T_B(Ax) \\ &= B(Ax) \\ &= (BA)x \\ &= T_{BA}(X) \end{aligned}$$

Recall that $AB \neq BA$ in general. Thus we conclude that

$$T_A T_B \neq T_B T_A.$$

The multiplication of linear operator (element of $\mathcal{L}(V)$) is *not commutative.*

Ex $m = n = 2$. The standard basis for $\mathbb{F}^{2\times 2}$, Let $x \in \mathbb{F}^{2\times 2}$. Then, $X = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.
Thus, $x \in \text{Span} \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$.
Note that $\forall x \in$ such span, $x = 0 \iff a, b, c, d = 0$. Thus, this set is a basis.
Suggested Exercise: Standard Basis for $\mathbb{F}^{m\times n}$

$\forall\, 1 \leq p \leq m$ and $1 \leq q \leq n$, let $E^{p,q}$ be the matrix with a 1 in the entry $e_{p,q}$ and 0's in all other entries.

Show that the following set is a basis for $\mathbb{F}^{m\times n}$:

$$\{E^{p,q} \mid 1 \leq p \leq m,\, 1 \leq q \leq n\}$$

Theorem 3.1.1.1 ($\mathcal{L}(V, W)$ dimension)**.** Let $\dim(V) = n, \dim(W) = m$ s.t. $n, m < \infty$, then

$$\dim(\mathcal{L}(V, W)) = mn.$$

Proof. Let $\beta = \{\alpha_1, \ldots, \alpha_n\}$ and $\beta' = \{\beta_1, \ldots, \beta_m\}$ be *ordered bases* for V and W respectively. For each pair (p, q) with $1 \leq p \leq m, 1 \leq q \leq n$, we define a map $E^{pq} \in \mathcal{L}(V, W)$ by

$$E^{pq}(\alpha_i) = \begin{cases} \beta_p, \text{if } i = q \\ 0, \text{if } i \neq q \end{cases}$$

Then,

$$E^{pq}(\alpha_i) = \gamma_{pq}\beta_j, \text{ where } \gamma_{pq} = \begin{cases} 1, i = q \\ 0, i \neq q. \end{cases}$$

From which it follows that

$$\{E^{pq} | 1 \leq p \leq m, 1 \leq q \leq n\} \text{ forms a basis for } \mathcal{L}(V, W).$$

Thus,

$$T(v) = T(\sum_{i=1}^{n} a_i\alpha_i) = \sum_{i=1}^{n} a_i T(\alpha_i).$$

Then,

$$\forall i, T(\alpha_i) = \sum_{j=1}^{m} b_j\beta_j = \sum_{i=1}^{m} b_j E^{ij}(\alpha_i) \implies T(v) = \sum_{i=1}^{n} a_i(\sum_{j=1}^{m} b_j\beta_j)$$

□

Definition 3.1.2 (Invertability)**.** $T \in \mathcal{L}(V, W)$ is invertable if and only if

1. T is one-to-one, i.e., $T\alpha = T\beta \implies \alpha = \beta$.
2. T is onto. $\forall w \in W, \exists v \in V$ s.t. $T(v) = w$.

Proposition 3.1.2.1 (equal dimensional invertability)**.** Let $T \in \mathcal{L}(V, W)$. If $\dim(V) = \dim(W) < \infty$, then T is *invertable* iff T is *one-to-one* or T is *onto.*

Remarks 3.1.2.1.1. This does not hold for infinite dimensional spaces. Consider the following counter ex. Let $A = \sum_{i=1}^{\infty} E^{i+1,i}$, $B = \sum_{i=1}^{\infty} E^{i,i+1}$. Then

$$BA = \sum_{i=1}^{\infty} E^{i,i}, \quad \text{whereas} \quad AB = \sum_{i=2}^{\infty} E^{i,i}.$$

Use the definition of matrix multiplication to verify this counter ex. Note that the left inverse exists for A but *no* right inverse exists. . . .

3.2 Isomorphisms and Coordinates

Definition 3.2.1 (Isomorphisms)**.** An invertible map $T \in \mathcal{L}(V, W)$ is also called an *isomorphism*. If such a map exists, we say V is *isomorphic* to W or V and W are isomorphic. Intuitively, isomorphic vector spaces are "the same up to relabeling".

Definition 3.2.2 (Standard basis for $\mathbb{F}^n$)**.** The set $\{e_1, \ldots, e_n\}$ is the "standard basis" for $\mathbb{F}^n$, where

$$\begin{aligned} e_1 &= (1, 0, \ldots, 0) \\ e_2 &= (0, 1, 0, \ldots, 0) \\ &\ \ \vdots \\ e_n &= (0, \ldots, 0, 1). \end{aligned}$$

Theorem 3.2.2.1. If V is a $\mathbb{F}$ vector space and $\dim(V) = n$, then V is isomorphic to $\mathbb{F}^n$.

Proof. Suppose $\{\alpha_1, \ldots, \alpha_n\}$ is an *ordered basis* for V. Then define $T \in \mathcal{L}(V, \mathbb{F}^n)$ by $T\alpha_i = e_i, \forall i$, where $\{e_1, \ldots, e_n\}$ is the standard basis for $\mathbb{F}^n$. Let $v \in V$. Then,

$$\begin{aligned} T(v) &= T(c_1\alpha_1 + \cdots + c_n\alpha_n) \\ &= c_1T(\alpha_1) + \cdots + c_nT(\alpha_n) \\ &= c_1e_1 + \cdots + c_ne_n \\ &= (c_1, \ldots, c_n). \end{aligned}$$

□

Exercise. Show that T from the previous exis an invertible linear transformation. Suppose V and W are both finite dimensional $\mathbb{F}$ vector spaces such that $\dim(V) = \dim(W)$. Show that V and W are isomorphic.

3.3 Matrix Representations of Linear Transformations

Definition 3.3.1 (Coordinate Vectors)**.** Let $\beta = \{\alpha_1, \ldots, \alpha_n\}$ be an *ordered basis* for V and a vector $\alpha \in V$, we let $[\alpha]_\beta$ denote the *coordinates* of α *relative* to β.

As $\alpha = c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n$, where $c_i \in \mathbb{F}$ are uniquely determined, then

$$[\alpha]_\beta = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

Definition 3.3.2 (Matrix representation)**.** Let V and W be *vector spaces*, with $\dim(V) = n$ and $\dim(W) = m$. Let β be an *ordered basis* for V and β' be an *ordered basis* for W. Then $\forall T \in \mathcal{L}(V, W)$, there is a unique $m \times n$ matrix A such that

$$[T(\alpha)]_{\beta'} = A[\alpha]_\beta, \quad \forall \alpha \in V$$

The matrix A is called the *matrix* of T *relative* to the *ordered bases* β, β'. Note that in the case where $w = v$ and $\beta = \beta'$, we use the notation $[T]_\beta$.

Proof. Let $A := (A_{ij}) \in \mathbb{F}^{m\times n}$, where the scalars A_{ij} are obtained by

$$T(\alpha_j) = \sum_{i=1}^{n} A_{ij}\beta_i, \forall j \in \{1, \ldots, n\},$$

where $\beta = \{\alpha_1, \ldots, \alpha_n\}$ and $\beta' = \{\beta_1, \ldots, \beta_m\}$, i.e.,

$$[T\alpha_j]_\beta' = \begin{bmatrix} A_{1j} \\ \vdots \\ A_{mj} \end{bmatrix}.$$

Now, let $\alpha = \sum_{j=1}^{n} c_j\alpha_j$.

Then,

$$\begin{aligned} T(\alpha) &= T(\sum_{j=1}^{n} c_j \alpha_j) \\ &= \sum_{j=1}^{n} c_j T(\alpha_j) \\ &= \sum_{i=1}^{n} c_j (\sum_{i=1}^{m} A_i j B_i) \end{aligned}$$

Thus,

$$[T(\alpha)]'_\beta = \begin{bmatrix} \sum_{j=1}^{n} A_{1j} c_j \\ \vdots \\ \sum_{j=1}^{n} A_{mj} c_j \end{bmatrix} = A\,[\alpha]_\beta\,.$$

□

Remarks 3.3.2.0.1. Note that the j^{th} column of A is given by the coordinate, $[T(\alpha_j)]_{\beta'}$

Ex. Let $T : \mathbb{R}^2 \to \mathbb{R}$ be defined by $T(x_1, x_2) = (0, x_2)$. Then T is a (*linear operator*) on $\mathbb{R}^2$. Let's explore how the operator T acts on the standard basis: $\beta = \{e_1, e_2\}$.

$$Te_1 = T(1,0) = (1,0) = 1e_1 + 0e_2 \implies [Te_1]_{e_1,e_2} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Similarly, $Te_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Note

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ x_2 \end{bmatrix}$$

So

$$[\alpha]_\beta = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

And

$$[T(\alpha)]_\beta = \begin{bmatrix} 0 \\ x_2 \end{bmatrix}.$$

# insert three ex.

Definition 3.3.3 (Matrix Representation)**.** Given a *linear operator* $T : V \to V$ and a *ordered basis* $\beta = \{\alpha_1, \ldots, \alpha_n\}$ for V, we can get the *matrix representation* $[T]_\beta$ for T relative to β, by:

$$[T]_\beta = \Big[[T\alpha_1]_\beta, \ldots, [T\alpha_n]_\beta\Big]$$

Question: Given $T \in \mathcal{L}(V)$, and order bases β, β', how are $[T]_\beta$ and $[T]'_\beta$ related.

Definition 3.3.4 (Change of Basis)**.** Let $\beta = \{\alpha_1, ..., \alpha_n\}$ and $\beta' = \{\alpha'_1, ..., \alpha'_n\}$ be an *ordered basis* for V. Suppose $T \in \mathcal{L}(V)$. Let $P = [P_1, \ldots, P_n]$ be the $n \times n$ matrix with columns $P_j = \left[\alpha'_j\right]_\beta \quad \forall 1 \leq j \leq n$. Then

$$[T]_{\beta'} = P^{-1}[T]_\beta P$$

Definition 3.3.5. The matrix P is the *change of basis* matrix from β' to β.

$$P^{-1} = [P'_1, \ldots, P'_n] \text{ where} P_j = \left[\alpha'_j\right]_\beta$$

Ex. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(x_1, x_2) = (0, x_2)$. Let $\beta = \{e_1, e_2\}$. Then

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

Let $\beta' = \{\alpha_1, \alpha_2\}$, where $\alpha_1 = (1, 1)$ and $\alpha_2 = (2, 1)$. Solve for $[T]_{\beta'}$ using a change of basis matrix.

Solution. Note that

$$\alpha_1 = (1,1) \implies [\alpha_1]_\beta = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\alpha_2 = (2,1) \implies [\alpha_2]_\beta = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Thus,

$$P = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}.$$

Note that by Gaussian Elimination, we can get

$$P^{-1} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}.$$

Thus,

$$[T]'_\beta = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ -1 & -1 \end{bmatrix}.$$

■

Definition 3.3.6 (Similarity)**.** Let $A, B \in \mathbb{F}^{n\times n}$. Then B is similar to A if an *invertable* matrix $P \in \mathbb{F}^{n\times n}$ exists such that $A = PBP^{-1}$

Ex. Let $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ be matrix *representation* in the standard basis, so

$$\begin{cases} Xe_1 = e_2 \\ Xe_2 = e_1 \end{cases} \qquad \begin{cases} Ze_1 = e_1 \\ Ze_2 = -e_2 \end{cases} .$$

Show that X and Z are similar.

Solution. Let $P = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$. Then,

$$P^{-1}ZP = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

■

Remarks 3.3.6.0.1 (Similar Matrices Represent the same linear transformations in different bases.)**.** Let $\beta = \{e_1, e_2\}$ and $\beta' = \{\alpha_1, \alpha_2\}$, where

$$\alpha_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), \quad \alpha_2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right).$$

Note that $X\alpha_1 = \alpha_1$ and $X\alpha_1 = \alpha_1$. In this ex, P is the *change of basis matrix* from β' to β, and

$$X = P^{-1}ZP.$$

This means that X *and* Z are matrices that represent the *same transformation*, but in *different bases.*

Theorem 3.3.6.1 (Properties of Determinant)**.** Let $M \in M_{n\times n}(\mathbb{R})$. Suppose that M' is obtained from M by row operation.

Scale a row. If $M \xrightarrow{(\lambda R_i \to R_i)} M'$ then $\det(M') = \lambda \det(M)$.

Exchange R_i and R_j. If $M \xrightarrow{R_i \leftrightarrow R_j} M'$ then $\det(M') = -\det(M)$.

Add a multiple of a row. If $M \xrightarrow{(R_j + \lambda R_i \to R_j)} M'$ then $\det(M') = \det(M)$.

Note: These properties of det are sometimes called "row operation invariance". This is not quite right, because the determinant does change when applying row operations. However, the determinant changes in simple and predictable ways.

Definition 3.3.7 (Submatrices - Minors)**.** The $i-j-$minor of $M \in M_{n\times n}(\mathbb{R})$ is the matrix M_{ij} obtained by deleting row i and column j from M.

Theorem 3.3.7.1 (Recursive Formula for det)**.** If $M = [m_{ij}] \in M_{n\times n}(\mathbb{R})$ and $1 \leq i \leq n$ then we define:

$$\det(M) = \sum_{j=1}^{n} (-1)^{i+j} m_{ij} \det(M_{ij}) = \sum_{i=1}^{n} (-1)^{i+j} m_{ij} \det(M_{ij})$$

3.4 Linear Functionals

Definition 3.4.1 (Linear Functionals)**.** Let V be a vector space over $\mathbb{F}$. The *linear transformations*, $\mathcal{L}(V, \mathbb{F})$ are called *linear functionals.*

Ex.(Trace) Let $A = A_{i,j} \in \mathbb{F}^{n\times n}$ and define the *trace* of A by

$$\operatorname{trace}(A) = \sum_{i=1}^{n} A_{ii}.$$

Proof. As $\operatorname{trace}(A) \in \mathbb{F}, \mathbb{F}^{n\times n} \to \mathbb{F}$. Let $c \in \mathbb{F}$ and $A, B \in \mathbb{F}^{n\times n}$. Then,

$$\begin{aligned}
\operatorname{trace}(cA + B) &= \sum_{i=1}^{n} (cA + B)_{ii} && \text{by definition of trace} \\
&= \sum_{i=1}^{n} cA_{ii} + B_{ii} && \text{by definition of matrix add/scaling} \\
&= c\sum_{i=1}^{n} A_{ii} + \sum_{i=1}^{n} B_{ii} && \text{as } \mathbb{F} \text{ is a field} \\
&= c\operatorname{trace}(A) + \operatorname{trace}(B) && \text{definition of trace}
\end{aligned}$$

Thus, $\operatorname{trace} \in \mathcal{L}(\mathbb{F}^{n\times n}, \mathbb{F})$, i.e., trace is a linear functional. □

Ex. (Evaluation maps.)
Let V be the space of *polynomial.* Let $t \in \mathbb{F}$. Show that

$$L_t(p) = p(t), \forall p \in V.$$

defines a *linear functional* $L_t \in \mathcal{L}(V, \mathbb{F})$.

Proof. As $p(t) \in \mathbb{F}, \forall p \in V$, $L_t(p) : V \to \mathbb{F}$. Let $c \in \mathbb{F}$ and $f, g \in V$. Then,

$$\begin{aligned}
L_t(cf + g) &= (cf + g)(t) \\
&= cf(t) + g(t) \\
&= cL_t(f) + L_t(g)
\end{aligned}$$

Thus, $L_t \in \mathcal{L}(V, \mathbb{F})$. □

Remarks 3.4.1.0.1. Such a *linear function* is often called aa *evaluation map* because it maps V to the field it is defined over by evaluating the functions at a given point.

Ex. (Definite integration.) Let $a, b \in \mathbb{R}$ and $V = C[a, b]$ (the set of *continuous functions* $f : [a, b] \to \mathbb{R}$). Define

$$L(f) = \int_a^b f(t)\, dt$$

Then L is a *linear functional* on V.

Proof. Let $f, G \in C[a, b]$ and $c \in \mathbb{R}$. The output of a definite integral is a real number, by calculus. Thus, $L : C[a, b] \to \mathbb{R}$. Then,

$$\begin{aligned} L(cf + g) &= \int_a^b (cf + g)(t)dt \\ &= \int_a^b cf(t) + g(t)dt \\ &= c\int_a^b f(t)dt + \int_a^b g(t)dt \\ &= cL(f) + L(g). \end{aligned}$$

Thus, $L \in \mathcal{L}(c\,[a, b]\,, \mathbb{R})$. □

Definition 3.4.2 (Standard Inner Product for $\mathbb{C}^n$)**.** Let $V = \mathbb{C}^n$ and fix $\alpha \in \mathbb{C}^n$. Let *inner product* $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ be the *standard inner product*;

$$\text{given } x = (x_1, \dots, x_n); y = (y_1, \dots, d_n) \to \langle x, y\rangle = \sum_{i=1}^{n} x_i \bar{y}_i$$

where $\bar{y}_i$ is the complex conjugate of y_i. Define $f_\alpha : \mathbb{C}^n \to \mathbb{C}$ by $f_\alpha(\beta) = \langle \beta, \alpha\rangle$ for all $\beta \in \mathbb{C}^n$.

Theorem 3.4.2.1.

$$f_\alpha \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}).$$

Proof. as $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}, f_\alpha : \mathbb{C}^n \to \mathbb{C}$ by definition. Let $c \in \mathbb{C}$ and $x, y \in \mathbb{C}^n$. Then

$$\begin{aligned} f_\alpha &= \langle cx + y, \alpha\rangle \\ &= \sum_{i=1}^{n} (cx + y)_i(\alpha_i^*) \\ &= \sum_{i=1}^{n} cx_i\alpha_i^* + y_i\alpha_i^* \\ &= c\sum_{i=1}^{n} x_i\alpha_i^* + \sum_{i=1}^{n} y_i\alpha_i^* \\ &= c\langle x, \alpha\rangle + \langle y, \alpha\rangle \\ &= cf_\alpha(x) + f_\alpha(y). \end{aligned}$$

Thus, $f_\alpha \in \mathcal{L}(\mathbb{C}^n, \mathbb{C})$. □

Definition 3.4.3 (Standard Inner Product for $\mathbb{R}^n$)**.** Let $V = \mathbb{R}^n$ and fix $\alpha \in \mathbb{R}^n$. Let *inner product* $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be the *standard inner product* ("dot product");

Define $T : \mathbb{R}^n \to \mathbb{R}$ by $T(x) = \langle x, \alpha \rangle$ for all $x \in \mathbb{R}^n$.

Ex. Let $\alpha = (i, 1) \in \mathbb{C}^2, \beta = (b_1, b_2)$. Determine the explicit formula for $f_\alpha(\beta)$. Note that

$$f_\alpha(b_1, b_2) = \langle (b_1, b_2), (i, 1) \rangle = b_i(i)^* + b_2(1)^* = b_1(-i) + b_2.$$

Thus, $f_\alpha(b_1, b_2) = b_2 - b_1 i$.

Theorem 3.4.3.1 (Riesz Representation Theorem)**.** *Every* linear functional on $\mathbb{C}^n$ is of the form $T(x) = \langle x, \alpha \rangle$ for some $\alpha \in \mathbb{C}^n$, i.e., if $f \in \mathcal{L}(\mathbb{C}^n, \mathbb{C})$, then

$$\exists \alpha \text{ s.t. } f(v) = \langle v, \alpha \rangle, \forall v \in \mathbb{C}^n.$$

Definition 3.4.4 (Annihilator)**.** Let S be a *subset* of the *vector space* V over $\mathbb{F}$. The *annihilator* of S is the set

$$S^o = \{T \in \mathcal{L}(V, \mathbb{F}) \,|\, T(v) = 0, \forall v \in S.\}$$

Note that S^o is a subspace of $\mathcal{L}(V, \mathbb{F})$. Also note that when $\mathbb{F} = \mathbb{R}, \mathbb{C}$ we replace T with f_α.

Definition 3.4.5 (Orthogonal Subspace)**.** Let S be a nonempty subset of the vector space $\mathbb{R}^n$ (or $\mathbb{C}^n$) and $\langle \cdot, \cdot \rangle$ be the corresponding standard inner product, then

$$S^\perp = \{x \in V \mid \langle x, y \rangle = 0, \forall\, y \in S\}.$$

Ex. Let $W = \operatorname{span}\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\} \subset \mathbb{R}^5$, where

$$\begin{cases} \alpha_1 = (2, -2, 3, 4, -1) \\ \alpha_2 = (-1, 1, 2, 5, 2) \\ \alpha_3 = (0, 0, -1, -2, 3) \\ \alpha_4 = (1, -1, 2, 3, 0) \end{cases}$$

Solve for $W^\perp$.

Solution. Let $x \in W^\perp$. Then, $\langle x, w\rangle = 0, \forall w \in w^\perp$. As $w \in w^\perp$,

$$\begin{aligned} w &= x_1\alpha_1 + \cdots + x_4\alpha_4 \\ &\implies \langle x|x_1\alpha_1 + \cdots + x_4\alpha_4\rangle = 0 \\ &\iff \sum_{i=1}^{n} x_i\langle x|\alpha_i\rangle = 0 \end{aligned}$$

If $\langle x|\alpha_i\rangle = 0, \forall i$, then $\langle x|w\rangle = 0$. (Converse left as exercise).
As $x \in W^\perp$, $\langle x|\alpha_i\rangle = 0, \forall i$. This means x is a solution to

$$\begin{cases} \langle x|\alpha_1\rangle = 0 \\ \vdots \\ \langle x|\alpha_4\rangle = 0 \end{cases}$$

Let $x = (v_1, \ldots, v_5)$. Then, $(v_1, \ldots, v_5)$ s a solution to

$$\begin{cases} 2v_1 - 2v_2 + 3v_3 + 4v_4 - v_5 &= 0 \\ -v_1 + v_2 + 2v_3 + 5v_4 + 2v_5 &= 0 \\ -3v_3 - 2v_4 + 3v_5 &= 0 \\ v_1 - v_2 + 2v_3 + 3v_4 &= 0 \end{cases}$$

Consider $W = \operatorname{Span}(\{v_1, \ldots, v_n\}) \subseteq \mathbb{C}^n$. Then,
$x \in W^\perp \iff x$ is a solution to

$$\begin{cases} \langle x|v_n\rangle = 0 \\ \langle x|v_n\rangle = 0 \\ \vdots \\ \langle x|v_n\rangle = 0. \end{cases}$$

■

3.5 Error Correcting (Hamming Codes) and XOR Encryption

In classical computers all our information is stored in bits. These bits are either on *1* or off *0*. We can represent a string of n bits as a vector $(x_1, x_2, ..., x_{\mathrm{n}}) \in (\mathbb{Z}_2)^n$.

In many situations, we would like to pass this information from one place to another. For ex, a DVD player reads from a DVD, your Laptop receives data from the Wi-Fi router, and so on. These situations can be represented as follows:

Alice has a piece of information and would like to pass it to Bob. She does so using the following process:

$$Alice \xrightarrow{\text{endcode}} \cdot \xrightarrow{\text{sending}} \cdot \xrightarrow{\text{super}} Bob.$$

Naive Solution. Let us first consider a straightforward error-correcting code to get a feel for error-correcting codes.

The code works as follows. Say Alice would like to send 0, 1, 0, she would instead send each bit in triplets. That is, she would send: *000, 111, 000*.

Then, if Bob receives *010, 101, 000*, Bob would know that at least one error has occurred since numbers do not occur in triplets. Using a "majority vote" strategy for each triplet, Bob would be able to fix the incorrect string to regain the original message *010*.

As we can see, there are two benefits to this method:

1. Bob knows stuff happened.

2. Bob can fix it.

However, note the drawbacks:

1. Two errors in one triplet can lead to incorrect deductions

2. It requires one to send a lot of excess information

Consider Hamming codes which were developed in an attempt to deal with the issue:

1. Alice encodes to an error-resistant code

2. Alice sends the code

3. Bob receives the code

4. Bob decodes the message

Definition 3.5.1 (Hamming Distance)**.** Let v_1 and v_2 be two strings (over $\mathbb{Z}_2^n$.) The *Hamming Distance* between v_1 and $v_2(d(v_1, v_2))$ is the numbers of bits as we travel along the string (vector) that are different.

In our ex, the 2nd and 5th entries of $v_1 = 0, 0, 0, 1, 1, 1, 0, 0, 0$ and $v_2 = 0, 1, 0, 1, 0, 1, 0, 0, 0$ are different, so their Hamming distance is 2.

It turns out that we can use vector addition to calculate the Hamming distance between vectors in $(\mathbb{Z}_2)^n$.

If we add two vectors using $\mathbb{Z}_2$ addition component-wise, then the entries of the resulting vector that are equal to 1 are exactly the entries that are different. Thus, the Hamming distance is the number of 1s in the summed string.

For ex,

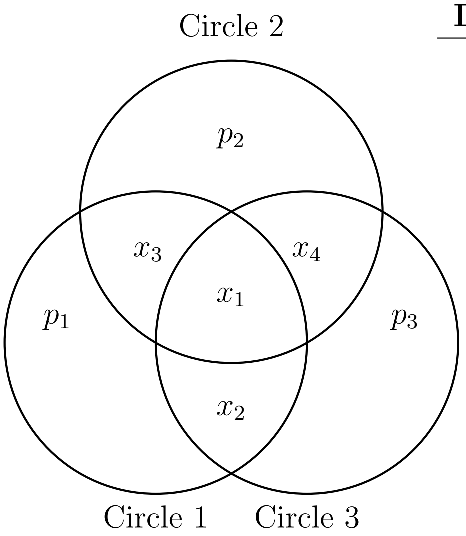
$$v_1 + v_2 = 0, 1, 0, 0, 1, 0, 0, 0, 0$$

So the Hamming distance between v_1 and v_2 is 2.

Definition 3.5.2 (Erro correcting code)**.** An *E-error correcting code* is a code designed to direct and fix a total of e errors.
If Alice sends Bob a message using an *e-error* correcting code, then the message can be decoded if there are $\leq e$ errors.

In proceeding section, we explore a special type of Hamming code, the *Hamming(7,4)* code. This is a 1-error correcting code.

Definition 3.5.3 (Hamming Codes)**.** Let the parity of a string of bits v be $d(v, 0)$,i.e., the number of 1 in the string. Note that the sum of the the bits over $\mathbb{Z}_2$ in a string is 0 iff the string has even parity, otherwise 1 with odd parity.
In illustration of Humming$(7, 4)$ code with 4 bits, (x_1, x_2, x_3, x_4) :



Location	**label**
x_1	$L_1 = (1, 1, 1)$
x_2	$L_2 = (1, 0, 1)$
x_3	$L_3 = (1, 1, 0)$
x_4	$L_4 = (0, 1, 1)$
p_1	$L_5 = (1, 0, 0)$
p_2	$L_6 = (0, 1, 0)$
p_3	$L_7 = (0, 0, 1)$

Where for any $L_i = (l_1, l_2, l_3), l_j = 1 \iff L_i$ is in $p_j, \forall j \in [1, 3] \cap \mathbb{N}$ and $i \in [1, 4] \cap \mathbb{N}$.
Also note that for any

$$p_i = (\rho_1, \rho_2, \rho_3), \rho_i = 1 \iff \text{the sum of all entries in the } i^{th} \text{ circle} = 1.$$

Ecoding the error resistant vector.
Consider $G \in (\mathbb{Z}_2)^{7\times 4}$ s.t.

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ | & | & | & | \\ L_1^T & L_2^T & L_3^T & L_4^T \\ | & | & | & | \end{bmatrix}.$$

Then, we have

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ | & | & | & | \\ L_1^T & L_2^T & L_3^T & L_4^T \\ | & | & | & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ p_1 \\ p_2 \\ p_3 \end{bmatrix}$$

Recieving the message. Suppose Bob receives the message $m = (x_1', \ldots, p_3')$. Further suppose that Bob uses the same VennDiagram as Alice. Let

$$H = \begin{bmatrix} L_1^T & L_2^T & L_3^T & L_4^T & L_5^T & L_6^T & L_7^T & \end{bmatrix}.$$

Then, $Hm = (c_1, c_2, c_3)^T$. In particular, shall any c_i be non-zero, an error has occurred—because Alice had ensured even parity within each circle. Assume only one error occurs. Then, a quick check shows that the error occurs precisely at the location $L_i = (c_1, c_2, c_3)$.

1. If $i \in [1, 4] \cap \mathbb{N}$. The vector $(x_1, \ldots, x_4)^T + e_i$ would then be the corrected message where e_i is a standard basis.
2. if $i > 4$, an error occurred in one of the parity bits, thus the message is error-free and irrelevant.

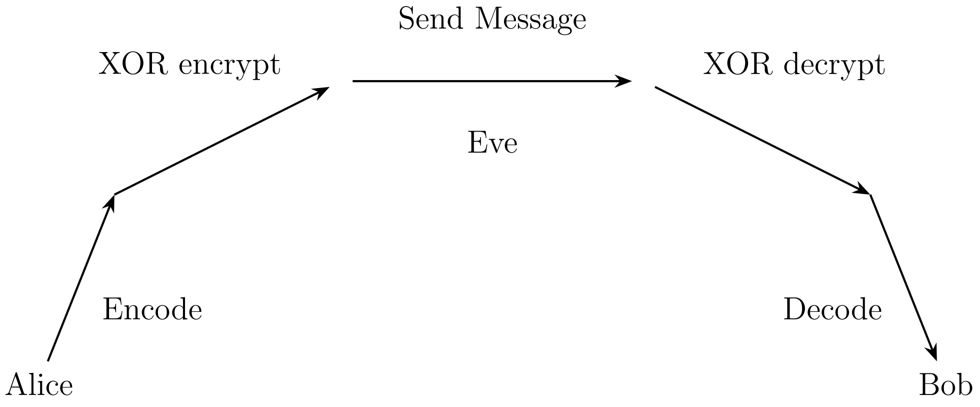
Note that for $k-$tuple, we have 2^k binary numbers. In particular, with k parity bits and $(0, 0, 0)$, we in total have $2^k - k - 1$ locations to store information.

3.6 XOR Encryption

Theorem 3.6.0.1. Let $x, y \in (\mathbb{Z}_2)^n$ s.t. $y \neq 0$. Then,

$$x + y \neq x, \quad x + y + y = x.$$

Motivation: after Eve receives message, she can add the *secret vector* y then recover the code encrypted by Alice by adding y.



4 Size and Distance

4.1 Metrics and Absolute Values for Integral Domains

Remarks 4.1.0.0.1. Thus far, we have explored many different vector space notions. We now discuss the notions of "length" and "distance" for vectors.

In point set topology, the *distance* between objects is defined using a *metric*, while the *length* of an object is defined using a *norm*.

Definition 4.1.1 (A Metric)**.** Let M be a set. Let $d : M \times M \to \mathbb{R}$ be a function. The function d is a metric if

1. $d(x, y) \geq 0, \forall x, y \in M$. non-negativity
2. $d(x, y) = d(y, x)$. commutativity/symmetric
3. $d(x, y) = 0 \iff x = y$. zero distance is equivalent to equality
4. $d(x, z) \leq d(x, y) + d(y, z)$. Δ inequality

Ex. Euclidean metric.
Let $z_1, z_2 \in \mathbb{C}$. Let $z_1 = a + bi$ and $z_2 = c + di$. Let

$$d(z_1, z_2) = \sqrt{(a-c)^2 + (b-d)^2}.$$

show that d is a metric.

Solution. First we show that $d(x, y) \geq 0$.
Let $x, y \in \mathbb{C}$ be arbitrary such that $x = a + bi$ and $y = c + di$. Then,

$$\begin{aligned} d(x,y) &= \sqrt{(a-c)^2 + (b-d)^2} \\ &\quad \text{note that } (a-c)^2 \text{ and } (b-d)^2 \geq 0 \implies (a-c)^2(b-d)^2 \geq 0 \\ &\quad \text{thus, } \sqrt{(a-c)^2 + (b-d)^2} \text{ is defined for all } x, y \in \mathbb{C} \\ &\quad \implies d(x,y) \geq 0, \forall x, y \in \mathbb{C} && \text{as range}(\sqrt{\cdot}) = [0, \infty). \end{aligned}$$

Then, we show that $d(x, y) = d(y, x)$. Note that

$$\begin{aligned} d(x,y) &= \sqrt{(a-c)^2 + (b-d)^2} \\ &= \sqrt{a^2 - 2ac + c^2 + b^2 - 2bd + d^2} \\ &= \sqrt{c^2 - 2ac + a^2 + d^2 - 2bd + b^2} \\ &= \sqrt{(c-a)^2 + (d-b)^2} \\ &= d(y,x) && \text{by definition} \end{aligned}$$

Now we show $d(x, y) = 0 \iff x = y$. Consider

$$\begin{aligned} d(x,y) = 0 = \sqrt{(a-c)^2 + (b-d)^2} \\ &\iff a - c = 0 \text{ and } b - d = 0 && \text{since } (\cdot)^2 \geq 0, \forall \cdot \\ &\iff a = c \text{ and } b = d \\ &\iff x = y && \text{by definition.} \end{aligned}$$

Exercise, prove Δ inequality. ■

Ex. Consider $\mathbb{Z}_p$, where $p \neq 2$ and p is prime. Let

$$d(a,b) = \min\left\{a +_{\text{mod } p} (-b),\ b +_{\text{mod } p} (-a)\right\}.$$

Show that d is a metric.

Solution. First we show non-negative. As elements of $\mathbb{Z}_p$ are non-negative,

$$a +_{\text{mod}\, p} (-b) \geq 0 \text{ and } b +_{\text{mod}\, p} (-a) \geq 0.$$

Then, we show symmetry. Let $a, b \in \mathbb{Z}_p$. Then,

$$\begin{aligned} d(a,b) &= \min\left\{a +_{\text{mod } p} (-b),\ b +_{\text{mod } p} (-a)\right\} \\ &= \min\left\{\ b +_{\text{mod } p} (-a), a +_{\text{mod } p} (-b)\right\} \\ &\qquad \text{by set extensionality axiom,i.e., set identity despite order} \\ &= d(b,a) && \text{by definition.} \end{aligned}$$

Next, we show that $d(a,b) = 0 \iff a = b$.
As $\mathbb{Z}_p$ is a field, additive inverse is unique.

$$\begin{aligned} a +_{\text{mod } p} (-b) = 0 &\iff -b = -a \\ &\iff a = b \end{aligned}$$

And that

$$\begin{aligned} b +_{\text{mod } p} (-a) = 0 &\iff -a = -b \\ &\iff a = b \end{aligned}$$

Lastly, we show Δ inequality. # left as an exercise. ■

Definition 4.1.2 (Absolute Value–valuation,magnitude,norm)**.** Let D be an integral domain or field. The function $|\cdot| : D \to \mathbb{R}$ is an absolute value function if

1. $|x| \geq 0, \forall x \in D$ non-negativity
2. $|x| = 0 \iff x = 0$
3. $|xy| = |x||y|, \forall x, y \in D$ multiplication
4. $|x+y| \leq |x| + |y|, \forall x, y \in D.$ Δ Inequality

Remarks 4.1.2.0.1. Oftentimes in math, there is a "natural" choice for something. For ex, the "natural" choice for a basis is the standard basis. Similarly, there are sets for which we have a "natural" absolute value. For ex,

- For real numbers: $|x| = \sqrt{x^2}$
- For complex numbers: $|a+bi| = \sqrt{a^2+b^2}$ modulus
- For Finite Field: The Trivial Absolute Value
- For Polynomials: There are plenty, but they are more obscure

Ex. The modulus and trivial absolute value.

Show that the modulus of a complex number forms an absolute value.

Trivial Absolute Value:

$$f(x) = \begin{cases} 1, & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

Show that f is an absolute value for all fields. **Ex.** The degree of a polynomial is not an absolute value!

Let f be a polynomial. The deg function is not an absolute value (as it fails multiplication and Δ inequality.)

Remarks 4.1.2.0.2 (Norms can mean different things for different algebraic structures.)**.** Some authors define a norm to be a function $f : D \to \mathbb{R}$ such that

1. $f(x) \geq 0$
2. **If** $f(x) \neq 0$ **and** $f(y) \neq 0$, **then** $f(x) \leq f(x)f(y)$.

This definition of a norm does not agree with our definition of norms: such functions are not necessarily absolute value functions. We will not use this nomenclature here.

Theorem 4.1.2.1. The absolute value of $\hat{1}$ is $1 (\in \mathbb{R})$

Proposition 4.1.2.2.

$$\hat{1}^{-1} = 1, \quad -0 = 0.$$

Proof. Let a be the multiplicative inverse for 1. Then,

$$\begin{aligned} 1 \cdot a &= 1 \\ &\implies |1 \cdot a| = |1| \\ &\implies |1||a| = |1| \\ &\implies |1||a| - |1| = 0 \\ &\implies |1|(|a| - 1) = 0. \end{aligned}$$

As $1 \neq 0, |1| \neq 0$. Thus,

$$\begin{aligned} |a| - 1 &= 0 \\ &\implies |a| = 1 \end{aligned}$$

as the multiplicative inverse of 1 is 1,

$$|\hat{a}| = 1.$$

□

Theorem 4.1.2.3 (uniqueness of trivial absolute value for finite fields.)**.** The trivial absolute value is the only absolute value for finite fields

Proof. Let $y \in \mathbb{F}$ be nonzero. As $\mathbb{F}$ is a finite field, the set $\{y, y^1, y^2, \ldots\}$ is finite. It follows that there exists $0 < a < b < \infty$ such that $y^a = y^b$, and so there exists $0 < k < \infty$ such that $y^k = 1$. Because $|\cdot|$ is a norm,

$$|y|^k = |y^k| = |1| = 1$$

so,

$$|y| = \begin{cases} \pm 1, & \text{if } k \text{ is even,} \\ 1, & \text{if } k \text{ is odd} \end{cases}$$

As $|\cdot|$ is an absolute value it is nonnegative, thus $|y| = 1$. It follows that $|\cdot|$ is the trivial absolute value. □

Ex. Let D be an integral domain. Let $a \in D$. Then, $|a| = |-a|$.

Remarks 4.1.2.3.1. Given an absolute value, we can define a metric.

Theorem 4.1.2.4 (Absolute values functions induce a metric)**.** Let D be an integral domain (or field). Let

$$|\cdot| : D \to \mathbb{R}$$

be an absolute value. Let $d(x, y) = |x - y|$, for all $x, y \in D$. Then d is a metric for D.

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

$$d(x, y) = |x - y| \geq 0 \qquad\qquad \text{as } |\cdot| : D \to [0, \infty)$$

Also,

$$\begin{aligned} d(x, y) &= |x - y| \\ &= |x + (-y)| \\ &= |x + (-1)y| \\ &= |(-1)y + x| \\ &= |-1|\,|(-1)y + x| \\ &= |(-1)((-1)y + x)| \\ &= |y - x| \\ &= d(y, x). \end{aligned}$$

Now, note that

$$\begin{aligned} d(x, y) = 0 &\iff |x - y| = 0 \\ &\iff x - y = 0 \\ &\iff x = y. \end{aligned}$$

Lastly,

$$\begin{aligned} d(x, z) &= |x + 0 - z| \\ &= |x - y + y - z| \\ &\leq |x - y| + |y - z| \\ &= d(x, y) + d(y, z). \end{aligned}$$

□

Ex. Show that the Euclidean distance is a metric by showing that it is the metric induced by the modulus function for complex numbers and the standard absolute value function for real numbers. That is to say, show

1. $|(a + bi) - (c + di)| = \sqrt{(a - c)^2 + (b - d)^2}$
2. $|a - b| = \sqrt{(a - b)^2}$

Remarks 4.1.2.4.1. Not all metrics are induced by an absolute value function. Consider $\mathbb{Z}_p$ s.t. $p \neq 2$. Let

$$d(a,b) = \min\left\{a - b(\pmod p), b - a(\pmod p)\right\}.$$

Show that d is not induced by the trivial absolute value.

4.2 Metrics and Norms for Vector Spaces

Definition 4.2.1 (A metric for vector spaces)**.** Let B be a vector space. Let $d : V \times V \to \mathbb{R}$ be a function. The function d is a metric, if:

1. $d(v, w) \geq 0, \forall v, w \in V$. non-negativity
2. $d(v, w) = d(w, v)$ commutativity/symmetric
3. $d(v, w) = 0 \iff v = w$. zero distance for equality
4. $d(u, w) \leq d(u, v) + d(v, w)$ Δ inequality

Ex. Euclidean Metric **Let** $x, y \in \mathbb{R}^2$**.** Let $f(x, y) = \sqrt{\sum_{i=1}^{2}(x_i - y_i)^2}$. **Show that the function** $f(x, y)$ **is a metric.**

Theorem 4.2.1.1 (1-metric –Taxi/Manhattan Metric)**.** Let $x, y \in \mathbb{F}^n$. and $d(x, y)$ be a metric for $\mathbb{F}$. Let

$$f(x, y) = \sum_{i=1}^{n} d(x_i, y_i).$$

Then, f is a metric.

Exercise. Let $\mathbb{F}$ be a field and d be a metric. Show that the following definitions of $f(x, y)$ define a metric on $\mathbb{F}^n$:

1. **Euclidean Metric (2-metric)**: $f(x, y) = \sqrt{\sum_{i=1}^{n} d(x_i, y_i)^2}$
2. L^∞ **or Chebyshev distance**: $f(x, y) = \max\{d(x_i, y_i)\}_{i=1}^{n}$

Remarks 4.2.1.1.1. Not all metrics are “nice” for vector spaces. When working with vector spaces, we typically would like our metric further satisfy two specific properties to be considered “nice”. Conveying why such metrics are “nice” is the goal of the rest of this section. Explicitly, we will show that a metric is “nice” iff a norm induces the metric.

1. Scaling our vectors by λ, scales our metric.
   - Let $\lambda \in \mathbb{F}$, then $d(\lambda x, \lambda y) = |\lambda| d(x, y)$.
2. Translation invariance
   - Let $z \in V$, then $d(x + z, y + z) = d(x, y)$.

Definition 4.2.2 (Vector Norm)**. A norm (vector norm).** Let V be a set. Let $\| \cdot \| : V \to \mathbb{R}$ be a function and $| \cdot | : \mathbb{F} \to \mathbb{R}$ be an absolute value. The function $\| \cdot \| : V \to \mathbb{R}$ is a norm if:

1. $\|x\| \geq 0$ for all $x \in V$. (Non-negative)
2. $\|x\| = 0$ if and only if $x = 0$. (zero if and only if zero)
3. $\|\lambda x\| = |\lambda| \|x\|$, for all $\lambda \in \mathbb{F}$ and $x \in V$. (distributes with scaling)
4. $\|x + y\| \leq \|x\| + \|y\|$, for all $x, y \in V$. (Triangle inequality)

Exercise. Let $\| \cdot \|$ be a vector space with absolute value $|\cdot|$.
Show that $\|a\| = \| - a\|, \forall a \in V$.
Ex. Euclidean Norm.
Let $x \in \mathbb{R}^2$. Let $\|x\| = \sqrt{x_1^2 + x_2^2}$. Show that $\| \cdot \|$ is a norm with the standard absolute value for real numbers.

Proof. Let $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}^2$. Then $x = (x_1, x_2)$. Thus,

$$\begin{aligned} \|\lambda x\| &= \|\lambda \left(x_1, x_2\right)\| \\ &= \|(\lambda x_1, \lambda x_2)\| \\ &= \sqrt{\left(\lambda x_1\right)^2 + (\lambda x_2)^2} \\ &= |\lambda| \, \|x\| \, . \end{aligned}$$

Also,

$$\|x + y\| \leq \|x\| + \|y\| \qquad \text{see week five lecture notes; note the iff with disjunction}$$

□

Remarks 4.2.2.0.1. We have shown that in $\mathbb{R}^2$ we can define a metric using the norm in the following manner:

$$d(x, y) = \|x - y\|_2.$$

The next theorem shows that this procedure of inducing a metric from a norm works generally.
See suggested exercises from lecture.

Theorem 4.2.2.1 (All vector norms induce a metric)**.** Let $\|\cdot\| : V \to \mathbb{R}$ be a norm with absolute value $|\cdot|$. Then $d(x, y) = \|x - y\|$ is a metric for V.

Proof. Let $x, y \in V$. Then, $d(x, y) = \|x - y\| \geq 0$. Further,

$$d(x, y) = \|x, y\| = \|(-1)(-1)(x - y)\| = |-1| \, \|-1(x - y)\| = 1 \, \|y - x\| = d(y, x).$$

Note that

$$d(x, y) = 0 \iff \|x - y\| = 0 \iff x - y = 0 \iff x = y.$$

Finally, for $z \in V$.

$$\begin{aligned} d(x, z) &= \|x - z\| = \|(x - y) + (y - z)\| \\ &\leq \|x - y\| + \|y - z\| \\ &= d(x, y) + d(y, z) \end{aligned}$$

□

Exercises: Let $\|\cdot\| : V \to \mathbb{R}$ be a norm. Let $d(x, y) = \|x - y\|$. Show that the following two properties hold:

- Let $\lambda \in \mathbb{F}$, then $d(\lambda x, \lambda y) = |\lambda| d(x, y)$. The function $|\lambda|$ depends on the field used to define the vector space.
- Let $z \in V$, then $d(x + z, y + z) = d(x, y)$.

Remarks 4.2.2.1.1. If V is a vector space with a norm $\|\cdot\|$, then we can create a metric for the vector space, by letting $d(x, y) = \|x - y\|$.

It is natural to ask if the converse holds. That is, given a vector space V and a metric d, can we create a norm by letting $\|x\| = d(x, 0)$? It turns out that the answer is yes provided that our metric is "nice."

Theorem 4.2.2.2 (Nice Metrics Induce a Norm)**.** Let V be a vector space and d be a metric such that

- Let $\lambda \in \mathbb{F}$, then $d(\lambda x, \lambda y) = |\lambda| d(x, y)$. The function $|\lambda|$ depends on the field used to define the vector space.
- Let $z \in V$, then $d(x + z, y + z) = d(x, y)$.

Show that $\|x\| = d(x, 0)$ is a norm.

Proof. let $x \in V$. Then,

$$\|x\| = d(x, 0) \geq 0.$$

Also,

$$\|x\| = 0 \iff d(x, 0) = 0 \iff x = 0.$$

Further,

$$\|\lambda x\| = d(\lambda x, 0) = |\lambda|\, d(x, 0) = |\lambda|\, \|x\|\,.$$

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

$$\begin{aligned}\|x + y\| &= d(x + y, 0) = d(x + y - y, 0 - y) \\ &= d(x, -y) \\ &\leq d(x, 0) + d(0, -y) && \text{triangle inequality} \\ &= d(x, 0) + d(0 + y, -y + y) && \text{assumption} \\ &= \|x\| + \|y\|\,.\end{aligned}$$

□

Remarks 4.2.2.2.1. We have just shown that given a "nice" metric d we can define a norm by setting $\|x\| = d(x, 0)$.

Now, observe that the metric d can alternatively be interpreted as the metric induced by this exact norm. Combining this observation with the results we have shown thus far, we see that a metric is "nice" iff the metric is induced by a norm!

4.3 Operator Norms and Matrix Norms

Definition 4.3.1 (operator norm)**.** Let V and W be both be real or both be complex vector spaces. Let $T : V \to W$ be a linear transformation. Let $\|\cdot\|_V$ be a norm for V and $\|\cdot\|_W$ be a norm for W. Then the *operator norm* of T is defined as

$$\|T\|_{op} = \sup\{\|T(v)\|_W : \|v\|_V \leq 1\}.$$

Definition 4.3.2 (matrix norm)**.** Let A be a real or complex $n \times m$ matrix. Let $\|\cdot\|_V$ be a norm for $\mathbb{F}^m$ and $\|\cdot\|_W$ be a norm for $\mathbb{F}^n$. Then the *matrix norm* of A is defined as

$$\|A\|_{op} = \sup\{\|Av\|_W : \|v\|_V \leq 1\}.$$

Exercise: Show that the Operator norm and Matrix norm are norms. That is to say:

1. $\|T\| \geq 0$
2. $\|T\| = 0$ if and only if $T = 0$
3. $\|aT\| = |a| \cdot \|T\|$
4. $\|S + T\| \leq \|S\| + \|T\|$

and

1. $\|A\| \geq 0$
2. $\|A\| = 0$ if and only if $T = 0$
3. $\|aA\| = |a| \cdot \|A\|$
4. $\|A + B\| \leq \|A\| + \|B\|$

5 Inner Product Spaces

5.1 Hermitian Inner Product

Definition 5.1.1 (Inner Product)**.** Let V be a vector space over $\mathbb{F}, (\mathbb{C} \vee \mathbb{R})$. An *Inner Product* on V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ such that $\forall \alpha, \beta, \gamma \in V$ and $\forall c \in \mathbb{F}$

1. $\langle \alpha + \beta | \gamma \rangle = \langle \alpha, \gamma \rangle + \langle \beta, \gamma \rangle$
2. $\langle c\alpha, \beta \rangle = c \langle \alpha, \beta \rangle$
3. $\langle \beta, \alpha \rangle = \overline{\langle \alpha, \beta \rangle}$
4. $\langle \alpha, \alpha \rangle \geq 0$ and $\langle \alpha, \alpha \rangle = 0 \iff \alpha = 0$.

Exercise. Prove that conjugate linear in the second argument:

$$\langle \alpha, c\beta + \gamma \rangle = \overline{c} \langle \alpha, \beta \rangle + \langle \alpha, \gamma \rangle .$$

Ex. Standard Inner Product.
Let $\alpha = (x_1, \cdots, x_n)$, $\beta = (y_1, \cdots, y_n)$.

$$\text{Over } \mathbb{C} \quad \Rightarrow \quad \alpha \cdot \beta = \sum_{j=1}^{n} x_j \overline{y_j}$$

$$\text{Over } \mathbb{R} \quad \Rightarrow \quad \alpha \cdot \beta = \sum_{j=1}^{n} x_j y_j$$

Proof. Let $\alpha, \beta \in \mathbb{C}^n$. Then,

$$\langle \alpha + \beta, \gamma \rangle = \langle \alpha, \gamma \rangle + \langle \beta, \gamma \rangle .$$

Let $\alpha, \beta, \gamma \in \mathbb{C}^n$ s.t. the i^{th} entry of x is x_i. Then,

$$\begin{aligned} \langle \alpha + \beta, \gamma \rangle &= \langle (\alpha_1, \ldots, \alpha_n) + (\beta_1, \ldots, \beta_n), (\gamma_1, \ldots, \gamma_n) \rangle \\ &= \sum_{i=1}^{n} (\alpha_i + y_i) \cdot \overline{z_i} \\ &= \sum_{i=1}^{n} x_i \overline{z_i} + \sum_{i=1}^{n} y_i \overline{z_i} \\ &= \langle x, z \rangle + \langle y, z \rangle . \end{aligned}$$

The rest left as an exercise. □

Exercise. For each of the following functions, show that they are inner products:

1. Let $\alpha = (x_1, x_2)$, $\beta = (y_1, y_2) \in \mathbb{R}^2$, define
$$\langle \alpha | \beta \rangle = x_1y_1 - x_2y_1 - x_1y_2 + 4x_2y_2.$$

2. Given $A := (a_{j,k}) \in \mathbb{C}^{n\times n}$, the $A^* := (\overline{a_{k,j}})$ is the *complex transpose* of A. We can define an inner product on $\mathbb{C}^{n\times n}$ by:
$$\langle A|B\rangle = \operatorname{trace}(AB^*) = \sum_{j=1}^{n}(AB^*)_{jj} = \sum_{j=1}^{n}\sum_{k=1}^{n} a_{j,k}\overline{b_{k,j}}.$$

3. Let V be the space of continuous functions $f : [0,1] \to \mathbb{C}$. Define an inner product of V by:
$$\langle f|g\rangle = \int_0^1 f(x)\overline{g(x)}\,dx.$$

Theorem 5.1.1.1 (Basic properties of inner products)**.** Let $\langle \cdot \mid \cdot \rangle$ be an inner product. Let x, y, z be arbitrary vectors and a, b be arbitrary scalars. Then:

1. $\langle x \mid 0\rangle = \langle 0 \mid x\rangle = 0$
2. $\langle x \mid x\rangle = 0$ if and only if $x = 0$
3. $\langle x \mid ay + bz\rangle = \overline{a}\langle x \mid y\rangle + \overline{b}\langle x \mid z\rangle$
4. $\langle x + y \mid x + y\rangle = \langle x \mid x\rangle + 2\operatorname{Re}(\langle x \mid y\rangle) + \langle y \mid y\rangle$

Theorem 5.1.1.2 (Inner Products Induce a Norm)**.** Let $\langle\cdot|\cdot\rangle$ be an inner product on V, we can define a *norm*
$$\|v\| = \sqrt{\langle v|v\rangle}.$$

Remarks 5.1.1.2.1 (Are all norm's induced by inner products)**.** The 1-norm and L^∞ norm are not induced by an inner product. Similar to how not all metrics are induced by norms, not all norms are induced by inner products. Luckily, there was—with determining if a metric could be induced by a norm—a nice way to determine if a norm is induced by an inner product.

Theorem 5.1.1.3 (Parallelorgram Law)**.** Let $\|\cdot\|$ be a norm for a vector space V. Then, $\|\cdot\|$ is induced by an inner product if and only if the norm satisfies the *Parallelogram Law*:
$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

Proof. Leave as an exercise. □

Exercise. Let $\langle \cdot \mid \cdot \rangle$ be an inner product and let $\| \cdot \|$ be defined as follows:

$$\|u\| = \sqrt{\langle u \mid u \rangle} \quad \forall u \in V.$$

Show that $\| \cdot \|$ satisfies

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

Definition 5.1.2 (Polarization identity)**.** Let V be a vector space and $\| \cdot \|$ be a norm that satisfies the parallelogram law. Then the *polarization identity* recovers the inner product from the norm:

Real domain:

$$\langle \alpha \mid \beta \rangle = \frac{1}{4}\|\alpha+\beta\|^2 - \frac{1}{4}\|\alpha-\beta\|^2$$

Complex case domain:

$$\langle \alpha \mid \beta \rangle = \frac{1}{4}\|\alpha+\beta\|^2 - \frac{1}{4}\|\alpha-\beta\|^2 + \frac{i}{4}\|\alpha+i\beta\|^2 - \frac{i}{4}\|\alpha-i\beta\|^2$$

Exercise. Consider $V = \mathbb{C}^n$. Let $x = (x_1, \ldots, x_n) \in V$. Let $\| \cdot \|$ be defined as follows:

$$\|x\| = \sqrt{\sum_{i=1}^{n} |x_i|^2},$$

where $| \cdot |$ is the modulus of the complex number.

1. Show that $| \cdot |$ is an absolute value function.

2. Show that $\| \cdot \|$ is a norm.

3. Define the metric induced by $\|x\|$.

4. Show that $\| \cdot \|$ satisfies the Parallelogram Law.

5. Define the inner product induced by $\|x\|$.

Definition 5.1.3 (Matrix Representation of Inner Products)**.** Let V be a vector space and let $\beta = \{\alpha_1, \cdots, \alpha_n\}$ be an ordered basis for V. Suppose $\langle \cdot \mid \cdot \rangle$ is an inner product on V. Then the *matrix* of the *inner product* in β is

$$G_{jk} := (\langle \alpha_k, \alpha_j \rangle) \in \mathbb{C}^{n \times n} \vee \mathbb{R}^{n \times n}.$$

Ex. The matrix G. Let V be a vector space and let $\beta = \{\alpha_1, \cdots, \alpha_n\}$ be an ordered basis for V. Suppose $\langle \cdot \mid \cdot \rangle$ is an inner product on V. Let

$$G_{j,k} = \langle \alpha_k \mid \alpha_j \rangle \in \mathbb{C}^{n\times n}.$$

Let $x, y \in V$, show that

$$\langle x \mid y \rangle = [y]_\beta^* G [x]_\beta.$$

Proof. As $x, y \in V$. $x = \sum_{i=1}^n x_i \alpha_i$ $y = \sum_{i=1}^n y_j \alpha_j$. Note that we have

$$\begin{aligned}
\langle x, y\rangle &= \left\langle \sum_{i=1}^n x_i\alpha_i, \sum_{i=1}^n y_j\alpha_j \right\rangle && \text{by definition of } x, y \\
&= \sum_{i=1}^n x_i \left\langle \alpha_i, \sum_{i=1}^n y_j\alpha_j \right\rangle && \text{as } \langle\rangle \text{ is linear in the first entry} \\
&= \sum_{i=1}^n x_i \overline{\left\langle \sum_{i=1}^n y_j\alpha_j, \alpha_i \right\rangle} && \text{conjugate symmetry} \\
&= \sum_{i=1}^n x_i \left(\sum_{j=1}^n \overline{y}_j \cdot \overline{\langle \alpha_j, \alpha_i\rangle} \right) \\
&= \sum_{i=1}^n x_i \left(\sum_{j=1}^n \overline{y}_j \cdot \langle \alpha_i, \alpha_j\rangle \right) \\
&= \sum_{j=1}^n \sum_{i=1}^n x_i \overline{y}_j \cdot \langle \alpha_i, \alpha_j \rangle
\end{aligned}$$

On the other hand,

$$\begin{aligned}
[y]_\beta^* \, G \, [x]_\beta &= \begin{bmatrix} \overline{y_1} & \dots & \overline{y_n} \end{bmatrix} G \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} \overline{y_1} & \dots & \overline{y_n} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n \langle \alpha_i, \alpha_1\rangle x_i \\ \sum_{i=1}^n \langle \alpha_i, \alpha_2\rangle x_i \\ \vdots \\ \sum_{i=1}^n \langle \alpha_i, \alpha_n\rangle x_i \end{bmatrix} \\
&= \sum_{j=1}^n \overline{y_j} \left(\sum_{i=1}^n \langle \alpha_i, \alpha_j \rangle x_i \right) \\
&= \sum_{j=1}^n \sum_{i=1}^n x_i \overline{y}_j \cdot \langle \alpha_i, \alpha_j\rangle \\
&= \langle x, y \rangle && \text{as needed.}
\end{aligned}$$

Exercise. Show that matrix G for the standard inner product, using the standard basis, is the identity matrix. **Exercise.**

1. Let β be a finite basis for a real vector space V and $\langle \cdot, \cdot \rangle$ be an inner product. Show that the matrix representation is a symmetric matrix.

2. Let β be a finite basis for a complex vector space V and $\langle \cdot, \cdot \rangle$ be an inner product. Show that the matrix representation is a Hermitian matrix (we will learn what this means shortly).

Ex. Let $\langle \cdot | \cdot \rangle$ be the standard inner product for $\mathbb{R}^n$. Define

$$A = \begin{bmatrix} \langle e_1, e_1 \rangle & \langle e_2, e_1 \rangle \\ \langle e_1, e_2 \rangle & \langle e_2, e_2 \rangle \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then, A is the matrix representation of the inner product.

5.2 Quadratic Forms

Definition 5.2.1 (Quadratic Forms)**.** Let $x \in \mathbb{R}^n$. A function $f(x_1, \ldots, x_n) = \sum_{i=1}^n \sum_{j=1}^n a_{i,j} x_i x_j$. Let $x \in \mathbb{C}^n$. A function $f(x_1, \ldots, x_n) = \sum_{i=1}^n \sum_{j=1}^n a_{i,j} \overline{x_i} x_j$.

Theorem 5.2.1.1 (Matrix Representation of quadratic forms)**.** Let $x \in \mathbb{R}^n$ and f be a quadratic form $f(x) = \sum_{i=1}^n \sum_{j=1}^n a_{i,j} x_i x_j$. Let $A = [a_{ij}]$. Then,

$$x^T A x = f(x).$$

Theorem 5.2.1.2 (Complex case)**.** Let $x \in \mathbb{R}^n$ and f be a quadratic form $f(x) = \sum_{i=1}^n \sum_{j=1}^n a_{i,j} \overline{x_i} x_j$. Let $A = [a_{ij}]$. Then,

$$\overline{x^T} A x = f(x).$$

5.3 Inner Product Spaces

An *inner product space* is a vector space, V, with specified *inner product* on V.

Theorem 5.3.0.1 (Inner Product Spaces)**.** Let $(V, \langle \cdot \rangle)$ be an inner product space. Let $\|\cdot\|$ be the norm induced by the inner product. Then, $\forall \alpha, \beta \in V$ and $c \in \mathbb{F}$,

1. $\|c\alpha\| = |c| \|\alpha\|$.
2. $\|\alpha\| > 0, \forall \alpha \neq 0$.
3. $|\langle \alpha | \beta \rangle| \leq \|\alpha\| * \|\beta\|$. Cauchy-Schwartz inequality
4. $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$. Triangle Inequality

Ex. By Cauchy-Schwartz inequality, we have

1. (Dot product)

$$\left| \sum_{k=1}^n x_k \overline{y_k} \right| \leq \left(\sum_{k=1}^n |x_k|^2 \right)^{1/2} \left(\sum_{k=1}^n |y_k|^2 \right)^{1/2}.$$

2.

$$|x_1 y_1 - x_2 y_1 - x_1 y_2 + 4 x_2 y_2| \leq \left((x_1 - x_2)^2 + 3x_2^2 \right)^{1/2} \left((y_1 - y_2)^2 + 3y_2^2 \right)^{1/2}.$$

3.

$$|\mathrm{tr}(AB^*)| \leq (\mathrm{tr}(A^* A))^{1/2} (\mathrm{tr}(B^* B))^{1/2}.$$

4.

$$\left| \int_0^1 f(t) \overline{g(t)} \, dt \right| \leq \left(\int_0^1 |f(t)|^2 \, dt \right)^{1/2} \left(\int_0^1 |g(t)|^2 \, dt \right)^{1/2}.$$

Proof. Of Cauchy-Schwartz Inequality. We prove by cases.
Case 1: $\alpha = a\beta$, for some $a\mathbb{F}$. Then,

$$\begin{aligned}
|\langle \alpha, \beta \rangle| &= |a \langle \beta, \beta \rangle| \\
&= |a| \langle \beta, \beta \rangle \\
&= \sqrt{|a|^2 \langle \beta \beta \rangle^2} \\
&= \sqrt{|a|^2 \langle \beta, \beta \rangle \langle \beta \beta \rangle} \\
&= \sqrt{a\overline{a} \langle \beta, \beta \rangle \langle \beta, \beta \rangle} \\
&= \sqrt{\langle a\beta, a\beta \rangle \langle \beta, \beta \rangle} \\
&= \sqrt{\langle \alpha, \alpha \rangle \langle \beta, \beta \rangle} \\
&= \sqrt{\langle \alpha, \alpha \rangle} \sqrt{\langle \beta, \beta \rangle} \\
&= \|\alpha\| \cdot \|\beta\| .
\end{aligned}$$

Case 2: $\alpha \neq a\beta, \forall a \in \mathbb{F}$ Note that thus $\alpha - a\beta \neq 0, \forall a$. By Properties of inner product,

$$\begin{aligned}
&\langle \alpha - a\beta, \alpha - a\beta \rangle > 0 \\
&\qquad \implies \langle \alpha, \alpha \rangle - a \langle \beta, \alpha \rangle - \overline{a} \langle \alpha, \beta \rangle + a\overline{a} \langle \beta, \beta \rangle > 0
\end{aligned}$$

From which it is established for all a. Consider $a = \frac{\langle \alpha, \beta \rangle}{\langle \beta, \beta \rangle}$. Then, substitution yields:

$$\begin{aligned}
&\langle \alpha \mid \alpha \rangle - \frac{\langle \alpha \mid \beta \rangle}{\langle \beta \mid \beta \rangle} \langle \beta \mid \alpha \rangle - \frac{\langle \beta \mid \alpha \rangle}{\langle \beta \mid \beta \rangle} \langle \alpha \mid \beta \rangle + \frac{\langle \alpha \mid \beta \rangle \langle \beta \mid \alpha \rangle}{\langle \beta \mid \beta \rangle} > 0 \\
\Rightarrow\quad &\langle \alpha \mid \alpha \rangle - \frac{\langle \alpha \mid \beta \rangle \langle \beta \mid \alpha \rangle}{\langle \beta \mid \beta \rangle} - \frac{\langle \beta \mid \alpha \rangle \langle \alpha \mid \beta \rangle}{\langle \beta \mid \beta \rangle} + \frac{\langle \alpha \mid \beta \rangle \langle \beta \mid \alpha \rangle}{\langle \beta \mid \beta \rangle} > 0 \\
\Rightarrow\quad &\langle \alpha \mid \alpha \rangle - \frac{\langle \alpha \mid \beta \rangle \langle \beta \mid \alpha \rangle}{\langle \beta \mid \beta \rangle} > 0 \\
\Rightarrow\quad &\langle \alpha \mid \alpha \rangle \langle \beta \mid \beta \rangle - \langle \alpha \mid \beta \rangle \langle \beta \mid \alpha \rangle > 0 \\
\Rightarrow\quad &\langle \alpha \mid \beta \rangle \langle \beta \mid \alpha \rangle < \langle \alpha \mid \alpha \rangle \langle \beta \mid \beta \rangle \\
\Rightarrow\quad &\langle \alpha \mid \beta \rangle \langle \beta \mid \alpha \rangle < \|\alpha\|^2 \|\beta\|^2 \\
\Rightarrow\quad &\sqrt{\langle \alpha \mid \beta \rangle \langle \beta \mid \alpha \rangle} < \|\alpha\| \|\beta\| \\
\Rightarrow\quad &|\langle \alpha \mid \beta \rangle| < \|\alpha\| \|\beta\|.
\end{aligned}$$

The idea is we pick the value of a a that minimizes the expression — and the resulting inequality still holds, and gives us the sharpest possible result. □

5.4 Orthogonality

Definition 5.4.1 (Orthogonality)**.** Let V be a vector space with a defined inner product. Then α is *orthogonal* to β if

$$\langle \alpha | \beta \rangle = 0.$$

Remarks 5.4.1.0.1. One can show that

$$\langle x, y \rangle = \|x\| \cdot \|y\| \cos(\theta) \text{ where theta is the angle between } x \text{ and } y$$

when $n = 2, 3$.

Definition 5.4.2 (Orthogonal and Orthonormal Sets)**.**

1. A set of vectors $S \subseteq V$ is *orthogonal* if

$$\langle \alpha, \beta \rangle = 0, \forall \alpha \neq \beta \in S.$$

2. $S \subseteq V$ is an *orthonormal set* if S is an *orthogonal set of unit vectors,* i.e.,

$$\|\alpha\| = 1, \forall \alpha \in S.$$

3. A basis that is orthonormal is called an *orthonormal basis.*

Ex.

1. Let $V = \mathbb{C}^{n \times n}$ with the trace inner product $\langle A \mid B \rangle = \mathrm{tr}(B^* A)$. The set $\{E^{p,q} : 1 \leq p, q \leq n\}$, where $E^{p,q}$ has zero entries everywhere except the p, qth entry, which is equal to 1.

2. The prototypical exof an orthonormal basis is the standard basis $\{e_1, \cdots, e_n\}$ for $\mathbb{R}^n$ and $\mathbb{C}^n$. While the standard basis is an orthonormal basis, with respect to the standard inner product, the standard basis does not form an orthonormal basis with respect to $\langle \alpha \mid \beta \rangle = x_1 y_1 - x_2 y_1 - x_1 y_2 + 4 x_2 y_2$

3. Let the vector space V be the set of continuous functions $f : [0, 1] \to \mathbb{C}$. For all $n \in \mathbb{N}$, let:

$$f_n(x) = \sqrt{2} \cos(2\pi n x), \quad g_n(x) = \sqrt{2} \sin(2\pi n x),$$

Prove that the set $\{1, f_1, g_1, f_2, g_2, \cdots\}$ is an infinite orthonormal set in V with respect to the inner product

$$\langle f \mid g \rangle = \int_0^1 f \overline{g}.$$

(This is often called the **Fourier Basis** for V.)

Proposition 5.4.2.1 (Orthogonal Sets are Linearly Independent)**.** Let V be an *inner product space.* An *orthogonal set* $S \subseteq V$ of non-zero vectors is *linearly independent.*

Proof. Let $\alpha_1, \ldots, \alpha_n$ be distinct vectors that is non-zero in S and let $\beta = \sum_i^m c_i\alpha_i$ Then,

$$\begin{aligned} \langle \beta | \alpha_k \rangle &= \langle c_1\alpha_1 + \cdots + c_m\alpha_m | \alpha_k \rangle \\ &= \sum_{j=1}^{m} c_j \langle \alpha_j | \alpha_k \rangle \\ &= c_k \langle \alpha_k | \alpha_k \rangle \\ &= c_k \|\alpha_k\|^2 \neq 0 \end{aligned}$$

Thus $c_k = \frac{\langle \beta | \alpha_k \rangle}{\|\alpha_k\|^2}, \forall 1 \leq k \leq m$. Note that as $\alpha_k \neq 0, \forall k$ this is valid. Note that if $\beta = 0$, then, $c_k = \frac{\langle 0 | \alpha_k \rangle}{\|\alpha_k\|^2} = 0$. Thus, $\{\alpha_1, \ldots, \alpha_m\}$ is linearly independent. □

Corollary 5.4.2.1.1 ($\beta \in S$)**.** If $\{\alpha_1, \ldots, \alpha_m\} = B$ is a set of *orthogonal non-zero vectors* that spans V, then $\forall \beta \in \text{Span}(B)$,

$$\beta = \sum_{k=1}^{m} \frac{\langle \beta | \alpha_k \rangle}{\|\alpha_k\|^2} \alpha_k.$$

Remarks 5.4.2.1.1 (Orthonormal Basis Defines Nice Coordinate Vectors)**.** Observe, if $\{\alpha_1, \cdots, \alpha_m\}$ is an **orthonormal basis** for V, then

$$\forall \beta \in V, \quad \beta = \sum_{k=1}^{n} \langle \beta \mid \alpha_k \rangle \alpha_k$$

This result tells us that if we have an orthonormal basis $a_1, \ldots, a_n$, we can similarly assign coordinates, except that this time the coordinates are: $\langle \beta, a_i \rangle$.

Thus, the standard basis and orthonormal basis are similar in the sense that they allow us to easily assign coordinates to vectors.

Ex. Let $\alpha_1 = \frac{1}{\sqrt{2}}(1,1)$, $\alpha_2 = \frac{1}{\sqrt{2}}(1,-1) \in \mathbb{R}^2$. Prove that $\{\alpha_1, \alpha_2\}$ is an orthonormal basis using the standard inner product.

Proof. **Prove Unit Length:**

$$\|\alpha_1\| = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}}[(1)(1) + (1)(1)] = 1$$

$$\|\alpha_2\| = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}}[(1)(1) + (-1)(-1)] = 1$$

Prove orthogonal:

$$\langle \alpha_1 \mid \alpha_2 \rangle = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}}[(1)(1) + (1)(-1)] = 0$$

Verifying the corollary for this ex: Let $\beta = (a,b)$, then

$$\langle \beta \mid \alpha_1 \rangle \alpha_1 + \langle \beta \mid \alpha_2 \rangle \alpha_2 = \frac{a+b}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) + \frac{a-b}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$$

$$= \frac{1}{2}[(a+b, a+b) + (a-b, a-b)] = (a,b) = \beta.$$

□

Ex. Compute the coordinate vector for $(1,3)$ using the basis

$$\beta = \left\{\frac{1}{\sqrt{2}}(1,1), \frac{1}{\sqrt{2}}(1,-1)\right\}.$$

Solution.

$$[(1,3)]_\beta = \begin{bmatrix} \left\langle (1,3) \mid \frac{1}{\sqrt{2}}(1,1) \right\rangle \\ \left\langle (1,3) \mid \frac{1}{\sqrt{2}}(1,-1) \right\rangle \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \langle (1,3) \mid (1,1) \rangle \\ \frac{1}{\sqrt{2}} \langle (1,3) \mid (1,-1) \rangle \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}}(1+3) \\ \frac{1}{\sqrt{2}}(1-3) \end{bmatrix} = \begin{bmatrix} \frac{4}{\sqrt{2}} \\ \frac{-2}{\sqrt{2}} \end{bmatrix}.$$

■

Theorem 5.4.2.2 (Matrix Representations of Linear Transformations with respect to Orthonormal Bases)**.** Let $T \in \mathcal{L}(V)$ and suppose $\beta = \{\alpha_1, \cdots, \alpha_n\}$ is an **ordered** **orthonormal basis** for V. Let

$$A = [T]_\beta = (A_{kj}),$$

where $1 \le k, j \le n$, then

$$A_{kj} = \langle T\alpha_j \mid \alpha_k \rangle, \quad \forall\, 1 \le j, k \le n$$

Ex. Let $T \in \mathcal{L}(\mathbb{R}^2)$ and

$$\beta = \left\{ \frac{1}{\sqrt{2}}(1,1), \frac{1}{\sqrt{2}}(1,-1) \right\}.$$

If $T(x,y) = (x+y, x-y)$, compute $[T]_\beta$.

$$[T]_\beta = \begin{bmatrix} \langle T\alpha_1 \mid \alpha_1 \rangle & \langle T\alpha_2 \mid \alpha_1 \rangle \\ \langle T\alpha_1 \mid \alpha_2 \rangle & \langle T\alpha_2 \mid \alpha_2 \rangle \end{bmatrix} = \begin{bmatrix} \langle (2/\sqrt{2}, 0) \mid (1/\sqrt{2}, 1/\sqrt{2}) \rangle & \langle (0, 2/\sqrt{2}) \mid (1/\sqrt{2}, 1/\sqrt{2}) \rangle \\ \langle (2/\sqrt{2}, 0) \mid (1/\sqrt{2}, -1/\sqrt{2}) \rangle & \langle (0, 2/\sqrt{2}) \mid (1/\sqrt{2}, -1/\sqrt{2}) \rangle \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 2 \\ 2 & -2 \end{bmatrix}.$$

5.5 Gram-Schmidt Process

Remarks 5.5.0.0.1. As we just learned, and we will continue to learn, life is generally much better when working with an orthonormal basis as opposed to a regular basis. Thus, it is of interest to develop a method to produce one from a basis which preserves the span and is orthonormal. Explicitly, what we want to achieve is the following:

Let $\{\beta_1, \cdots, \beta_m\}$ be **Linearly Independent** in V. Then we can construct a set of **orthonormal** vectors $\{\alpha_1, \cdots, \alpha_m\}$ in V that preserves the span, that is

$$\text{span}\{\alpha_1, \cdots, \alpha_k\} = \text{span}\{\beta_1, \cdots, \beta_k\}, \quad \forall k = 1, \cdots, \beta_m.$$

The Gram-Schmidt process accomplishes this goal.

Theorem 5.5.0.1 (Gram-Schmidt Procedure)**.** Let $\{\beta_1, \dots, \beta_j\}$ be a linearly independent set of vectors. Consider the set of vectors $\{\alpha_1, \dots, \alpha_j\}$ defined by the Gram-Schmidt procedure:

1. **Let** $\alpha_1 = \beta_1$.
2. **Let** $\alpha_2 = \beta_2 - \frac{\langle \beta_2 | \alpha_1 \rangle}{\|\alpha_1\|^2}\alpha_1$.

$\vdots$

j. **Let**

$$\alpha_j = \beta_j - \sum_{k=1}^{j-1} \frac{\langle \beta_j \mid \alpha_k \rangle}{\|\alpha_k\|^2}\alpha_k.$$

Then, the set $\{\alpha_1, \dots, \alpha_j\}$ is orthogonal and $\alpha_i \neq 0$. Note that the Gram-Schmidt process only produces an orthogonal set and not an orthonormal one.

Theorem 5.5.0.2 (Gram-Schmidt-Orthonormality)**.** Let V be a finite dimensional vector space such that $\dim(V) \geq 2$. Let $\beta = \left\{ \frac{\alpha_1}{\|\alpha_1\|}, \ldots, \frac{\alpha_1}{\|\alpha_n\|} \right\}$ be a basis constructed using the Gram-Schmidt process. Then, β is *Orthonormal.*

Proof. Firstly, note that $\|\alpha_i\| \neq 0$ so we can define β as follows.

$$\beta = \{\alpha_1/\|\alpha_1\|, \ldots, \alpha_j/\|\alpha_j\|\}.$$

First, we will prove that every vector is unit length.

$$\left\langle \frac{\alpha_i}{\|\alpha_i\|}, \frac{\alpha_i}{\|\alpha_i\|} \right\rangle = \frac{1}{\|\alpha_i\|^2}\langle \alpha_i, \alpha_i \rangle = \frac{1}{\langle \alpha_i, \alpha_i \rangle}\langle \alpha_i, \alpha_i \rangle = 1.$$

As we have proven every vector is unit length, we must show that

$$\left\langle \frac{\alpha_i}{\|\alpha_i\|}, \frac{\alpha_j}{\|\alpha_j\|} \right\rangle = 0 \quad \text{when } i \neq j.$$

Note that

$$\left\langle \frac{\alpha_i}{\|\alpha_i\|}, \frac{\alpha_j}{\|\alpha_j\|} \right\rangle = \frac{1}{\|\alpha_i\|\|\alpha_j\|}\langle \alpha_i, \alpha_j \rangle.$$

Thus,

$$\left\langle \frac{\alpha_i}{\|\alpha_i\|}, \frac{\alpha_j}{\|\alpha_j\|} \right\rangle = 0 \iff \langle \alpha_i, \alpha_j \rangle = 0.$$

To make the analysis easier, we will show that $\langle \alpha_i, \alpha_j \rangle = 0$ for all $i \neq j$. (This is going to be a proof via strong induction.)

Base case: $n = 2$**.**

$$\langle \alpha_2 \mid \alpha_1 \rangle = \langle \beta_2 \mid \alpha_1 \rangle - \frac{\langle \beta_2 \mid \alpha_1 \rangle}{\langle \alpha_1 \mid \alpha_1 \rangle}\langle \alpha_1 \mid \alpha_1 \rangle = \langle \beta_2 \mid \alpha_1 \rangle - \langle \beta_2 \mid \alpha_1 \rangle = 0.$$

$n + 1$ *case:*

Assume that $\langle \alpha_i, \alpha_j \rangle = 0$ for $i \neq j$, when $i, j \in \{1, \ldots, n\}$. (This is our use of strong induction.)

Consider $\langle \alpha_{n+1} \mid \alpha_j \rangle$**:**

$$\begin{aligned}\langle \alpha_{n+1} \mid \alpha_j \rangle &= \left\langle \beta_{n+1} - \sum_{k=1}^{n} \frac{\langle \beta_{n+1} \mid \alpha_k \rangle}{\langle \alpha_k \mid \alpha_k \rangle}\alpha_k \,\middle|\, \alpha_j \right\rangle \\ &= \langle \beta_{n+1} \mid \alpha_j \rangle - \left\langle \sum_{k=1}^{n} \frac{\langle \beta_{n+1} \mid \alpha_k \rangle}{\langle \alpha_k \mid \alpha_k \rangle}\alpha_k \,\middle|\, \alpha_j \right\rangle\end{aligned}$$

$$= \langle \beta_{n+1} \mid \alpha_j \rangle - \sum_{k=1}^{n} \frac{\langle \beta_{n+1} \mid \alpha_k \rangle}{\langle \alpha_k \mid \alpha_k \rangle} \langle \alpha_k \mid \alpha_j \rangle$$

$$= \langle \beta_{n+1} \mid \alpha_j \rangle - \sum_{k=1}^{j-1} \frac{\langle \beta_{n+1} \mid \alpha_k \rangle}{\langle \alpha_k \mid \alpha_k \rangle} \langle \alpha_k \mid \alpha_j \rangle - \frac{\langle \beta_{n+1} \mid \alpha_j \rangle}{\langle \alpha_j \mid \alpha_j \rangle} \langle \alpha_j \mid \alpha_j \rangle - \sum_{k=j+1}^{n} \frac{\langle \beta_{n+1} \mid \alpha_k \rangle}{\langle \alpha_k \mid \alpha_k \rangle} \langle \alpha_k \mid \alpha_j \rangle$$

$$= \langle \beta_{n+1} \mid \alpha_j \rangle - 0 - \langle \beta_{n+1} \mid \alpha_j \rangle - 0 = 0.$$

□

Corollary 5.5.0.2.1. Every finite-dimensional inner product space has an orthonormal basis.

5.6 Orthogonal Projections

Definition 5.6.1 (Best Approximation)**.** Let W be a subspace of an inner product space V. Let $\beta \in V$. The *best approximation* to β, by the vectors in V is the vector $\alpha \in W$ s.t.

$$\|\beta - \alpha\| \leq \|\beta - \gamma\| \iff d(\beta, \alpha) \leq d(\beta, \gamma) \iff \|\beta - \alpha\| = \min_{\gamma \in W} \|\beta - \gamma\|, \forall \gamma \in W.$$

Theorem 5.6.1.1. Let W be a subspace of an inner product space V and let $\beta \in V$. Then:

1. **Best approximations are characterized by an orthogonality relation:**

   The vector $x \in W$ is a best approximation to β by vectors in W if and only if $\beta - x$ is orthogonal to every vector in W.

2. **The best approximation exists and can be computed with a formula:**

   If W is finite-dimensional and $\{w_1, \ldots, w_n\}$ is any orthonormal basis for W, then the vector
   $$x = \sum_{i=1}^{n} \frac{\langle \beta \mid w_i \rangle}{\langle w_i | w_i \rangle} w_i$$
   is the (unique) best approximation to β by vectors in W.

3. **Best approximations are unique:**

   (a) If a best approximation to β by vectors in W exists, it is unique.

Proof. First let $\gamma \in V$, then $\beta - \gamma = (\gamma - \alpha) + (\alpha - \gamma)$, thus

$$\|\beta - \gamma\|^2 = \|\beta - \alpha\|^2 + 2\Re\left(\langle \beta - \alpha | \alpha - \gamma \rangle\right) + \|\alpha - \gamma\|^2.$$

$\Leftarrow$. Suppose $\beta - \alpha$ is orthogonal to every vector in W, that $\gamma \in W$ s.t. $\gamma \neq \alpha$. Then, since $\alpha - \gamma \in W$, it follows that as $\langle \beta - \alpha | \alpha - \gamma \rangle = 0$ and $\|\alpha - \gamma\| > 0$,

$$\|\beta - \gamma\|^2 = \|\beta - \gamma\|^2 + \|\alpha - \gamma\|^2 > \|\beta - \alpha\|^2 .$$

Thus, α is the best approximation to β by definition.
$\Rightarrow$. Suppose $\|\beta - \alpha\|^2 \leq \|\beta - \gamma\|^2 , \forall \gamma \in W$. Thus, by definition

$$2\Re\left(\langle \beta - \alpha | \gamma \rangle\right) + \|\alpha - \gamma\|^2 \geq 0, \forall \gamma \in W.$$

As $\alpha \in W \implies \alpha - \gamma \in W$ holds for all vectors in W. Consider such arbitrary $r \in W$,

$$2\Re \langle \beta - \alpha | r \rangle + \|r\|^2 \geq 0.$$

If $\gamma \in W$ and $\gamma \neq \alpha$, we may thus specify

$$r = -\frac{\langle \beta - \alpha | \alpha - \gamma \rangle}{\|\alpha - \gamma\|^2}(\alpha - \gamma).$$

From which it follows that

$$\begin{aligned}
&2\Re\left(\left\langle \beta - \alpha \,\middle|\, -\frac{\langle \beta - \alpha \mid \alpha - \gamma \rangle}{\|\alpha - \gamma\|^2}(\alpha - \gamma) \right\rangle\right) + \left\| \frac{\langle -\beta - \alpha \mid \alpha - \gamma \rangle}{\|\alpha - \gamma\|^2}(\alpha - \gamma) \right\|^2 \geq 0 \\
&\Rightarrow \; -2\Re\left(\frac{\overline{\langle \beta - \alpha \mid \alpha - \gamma \rangle}}{\|\alpha - \gamma\|^2} \langle \beta - \alpha \mid \alpha - \gamma \rangle \right) + \left\langle \frac{\langle \beta - \alpha \mid \alpha - \gamma \rangle}{\|\alpha - \gamma\|^2}(\alpha - \gamma) \,\middle|\, \frac{\langle \beta - \alpha \mid \alpha - \gamma \rangle}{\|\alpha - \gamma\|^2}(\alpha - \gamma) \right\rangle \geq 0 \\
&\Rightarrow \; -2 \cdot \frac{|\langle \beta - \alpha \mid \alpha - \gamma \rangle|^2}{\|\alpha - \gamma\|^2} + \frac{\langle \beta - \alpha \mid \alpha - \gamma \rangle}{\|\alpha - \gamma\|^2} \cdot \frac{\overline{\langle \beta - \alpha \mid \alpha - \gamma \rangle}}{\|\alpha - \gamma\|^2} \cdot \langle \alpha - \gamma \mid \alpha - \gamma \rangle \geq 0 \\
&\Rightarrow \; -2 \cdot \frac{|\langle \beta - \alpha \mid \alpha - \gamma \rangle|^2}{\|\alpha - \gamma\|^2} + \frac{|\langle \beta - \alpha \mid \alpha - \gamma \rangle|^2}{\|\alpha - \gamma\|^4} \cdot \|\alpha - \gamma\|^2 \geq 0 \\
&\Rightarrow \; -2 \cdot \frac{|\langle \beta - \alpha \mid \alpha - \gamma \rangle|^2}{\|\alpha - \gamma\|^2} + \frac{|\langle \beta - \alpha \mid \alpha - \gamma \rangle|^2}{\|\alpha - \gamma\|^2} \geq 0 \\
&\Rightarrow \; -\frac{|\langle \beta - \alpha \mid \alpha - \gamma \rangle|^2}{\|\alpha - \gamma\|^2} \geq 0
\end{aligned}$$

which is true only if

$$\langle \beta - \alpha | \alpha - \gamma \rangle = 0.$$

Therefore $\beta - \alpha$ is perpendicular to every vector in W.

Now we prove uniqueness. Suppose there is two best approximation a, a'. WTS: $a = a$, i.e., $\langle a - a' | a - a' \rangle = 0$. Thus,

$$\begin{aligned}\langle a - a' | a - a' \rangle &= \langle \beta - \beta + a - a' | a - a' \rangle \\ &= \langle \beta - a' | a - a' \rangle - \langle \beta - a | a - a \rangle \\ &= 0 - 0 = 0.\end{aligned}$$

Lastly, sps the W is finite-dimensional subspace of V. Then, by Gram-Schmidt W has an orthogonal basis, $\{\alpha_1, \ldots, \alpha_k\}$. Sps $\beta \notin W$, for otherwise the result is trivial. Then, the set $\{\alpha_1, \ldots, \alpha_k, \beta\}$ is linearly independent so we may apply Gram-Schmidt. Applying so yields that $\beta - \alpha$ is orthogonal to vectors in Span $\{\alpha_1, \ldots, \alpha_n\}$, i.e., $\forall w \in W$. If $\gamma \neq \alpha$ is in W, it follows that

$$\|\beta - \gamma\| > \|\beta - \alpha\| .$$

□

Definition 5.6.2 (Orthogonal Projection)**.** Let $\{\alpha_1, \ldots, \alpha_m\}$ be an orthonormal basis for W. The linear operator $E \in \mathcal{L}(V)$ defined by the best approximations is called the orthogonal projection of β onto W.

$$E(\beta) = \sum_{i=1}^{n} \langle \beta | \alpha_i \rangle \alpha_i.$$

Exercise. The orthogonal operator $E(\beta)$ has the following properties. Prove that

1. E is linear
2. E is a projection ($E^2 = E$)
3. range$(E) = W$.

Definition 5.6.3 (Orthogonal Compliment)**.** The *orthogonal complement* of a subspace $W \subseteq V$ is a subspace

$$W^{\perp} = \{v \in V \text{ s.t. } \langle v, w\rangle = 0, \forall w \in W\}.$$

Theorem 5.6.3.1 ($\beta \mapsto \beta - E(\beta)$)**.** Let V be an inner product space, W a finite-dimensional subspace of V, and E the orthogonal projection of V on W. The mapping

$$\beta \mapsto \beta - E(\beta)$$

i.e.,

$$(I - E)(\beta) = \beta - \sum_{i=1}^{n} \langle \beta | \alpha_i \rangle \alpha_i$$

is the orthogonal projection of V on $W^{\perp}$.

Proof. Let β be an arbitrary vector in V. Then, it follows from the properties of best approximations that $\beta - E(\beta) \in W^{\perp}$ and $\forall \gamma \in W^{\perp}$,

$$\beta - \gamma = E(\beta) + (\beta - E(\beta) - \gamma).$$

Since $E(\beta) \in W$, and $\beta - E(\beta) - \gamma \in W^{\perp}$, from the Pythagorean Theorem we get:

$$\|\beta - \gamma\|^2 = \|E(\beta)\|^2 + \|\beta - E(\beta) - \gamma\|^2 \geq \|\beta - (\beta - E(\beta))\|^2.$$

with strict inequality when $\gamma \neq \beta - E(\beta)$. Therefore, $\beta - E(\beta)$ is the best approximation to β by the vectors in $W^{\perp}$. □

It turns out that we are able to calculate the orthogonal projection of V onto $W^{\perp}$ using our function E:

$$I - E \text{ is the orthogonal projection of } V \text{ onto } W^{\perp}, \text{ and range}(I - E) = W^{\perp}.$$

□

Remarks 5.6.3.1.1. Let W be a plane through the origin in $\mathbb{R}^3$. Geometrically, one sees that $W^{\perp}$ is the line centred at the origin normal to the plane. In a similar fashion, the orthogonal complement to this line is the original plane, so in this case $(W^{\perp})^{\perp} = W$. It turns out that this is true more generally, as we shall see.

Ex.

$$\begin{aligned} &E(\alpha) = \alpha && \forall \alpha \in W && (I-E)(\alpha) = \alpha && \forall \alpha \in W^{\perp} \\ &E(\alpha) = 0 && \forall \alpha \in W^{\perp} && (I-E)(\alpha) = 0 && \forall \alpha \in W \end{aligned}$$

Theorem 5.6.3.2. Let W be a finite-dimensional subspace of a finite-dimensional inner product space V, and let E be the orthogonal projection of V on W. Then E is a projection of V onto W, $W^{\perp}$ is the nullspace of E, and

$$V = W \oplus W^{\perp}.$$

Proof left as an exercise:

- Step 1: Show that E is a projection.
- Step 2: Show that E is a linear operator.
- Step 3: Show $W^{\perp}$ is the nullspace of E.
- Step 4: Show that $W \oplus W^{\perp}$ is a direct sum.

Corollary 5.6.3.2.1. Let W be a finite-dimensional subspace of an inner product space V, and let E be the orthogonal projection of V onto W. Then $I - E$ is a projection of V onto $W^{\perp}$, and W is the nullspace of $I - E$.

Corollary 5.6.3.2.2 (Bessel's Inequality)**.** Let $\{\alpha_1, \cdots, \alpha_n\}$ be an orthogonal set of non-zero vectors in an inner product space V. If β is any vector in V, then

$$\sum_{i=1}^{n} \frac{|\langle \beta \mid \alpha_k \rangle|^2}{\|\alpha_k\|^2} \leq \|\beta\|^2$$

with equality if and only if

$$\beta = \sum_{i=1}^{n} \frac{\langle \beta \mid \alpha_k \rangle}{\|\alpha_k\|^2} \alpha_k.$$

Proof. Let

$$\gamma = \sum_{i=1}^{n} \frac{\langle \beta \mid \alpha_k \rangle}{\|\alpha_k\|^2} \alpha_k, \quad \text{and } \delta = \beta - \gamma. \quad \text{Then, } \beta = \gamma + \delta.$$

By definition of γ, $\langle \delta \mid \gamma \rangle = 0$.

Hence, by the Pythagorean Theorem,

$$\|\beta\|^2 = \|\gamma\|^2 + \|\delta\|^2.$$

By definition of γ,

$$\|\gamma\|^2 = \sum_{i=1}^{n} \frac{|\langle \beta \mid \alpha_k \rangle|^2}{\|\alpha_k\|^2} \leq \|\gamma\|^2 + \|\delta\|^2 = \|\beta\|^2,$$

with equality if and only if $\|\delta\|^2 = 0$.

The equality part follows immediately.

It now suffices to prove that

$$\|\gamma\|^2 = \sum_{i=1}^{n} \frac{|\langle \beta \mid \alpha_k \rangle|^2}{\|\alpha_k\|^2}.$$

The remainder of the proof is left as an exercise. □

Ex. Let $W = \operatorname{span}\{(1,-1)\}$ be a subspace of $\mathbb{R}^2$. Let $E : \mathbb{R}^2 \to \mathbb{R}^2$ be the orthogonal projection of $\mathbb{R}^2$ onto W.

1. Find a formula for $E(x_1, x_2)$.
2. $\beta = \{e_1, e_2\}$. Find $[E]_\beta$
3. Find $W^\perp$
4. Let $\beta' = \{(1,-1), (1,1)\}$, Solve for E.

Solution. 1. Consider

$$\begin{aligned} E(x_1, x_2) &= \sum_{k=1}^{1} \frac{\langle (x_1, x_2) | \alpha_i \rangle}{\|\alpha_i\|^2} \cdot \alpha_i \\ &= \left(\frac{1}{2}x_1 - \frac{1}{2}x_2, -\frac{1}{2}x + \frac{1}{2}x_2 \right). \end{aligned}$$

2. Note that

$$\begin{aligned} E(e_1) &= E(1,0) = (\frac{1}{2}, \frac{-1}{2}) \\ E(e_2) &= E(0,1) = (\frac{-1}{2}, \frac{1}{2}) \\ \implies [E]_\beta &= \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}. \end{aligned}$$

It is noteworthy that $[E]_\beta = [E]_\beta^2 = [E]_\beta^* = [E^*]_\beta$.
3. Observe that

$$\begin{aligned} v = (x_1, x_2) \in W^\perp &\iff (x_1, x_2) \text{ satisfies } \langle (x_1, x_2) \,|\, (1,-1) \rangle = 0 \\ &\iff (x_1, x_2) \text{ is a solution to } x_1 - x_2 = 0 \\ &\iff x_1 = x_2 \implies w^\perp = \operatorname{Span}(\{(1,1)\}). \end{aligned}$$

4. Note that

$$\begin{aligned} E(1,-1) = \left(\frac{1}{2} - \frac{-1}{2}, \frac{-1}{2} + \frac{-1}{2} \right) &= (1,-1) \\ E(1,1) &= (0,0) \\ \implies [E]_{\beta'} &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

■

5.7 The Least Squares Problem $V = \mathbb{R}^n$ s.t. $n < \infty$

Consider a data set

$$\{(1,2),(2,2),(3,4)\}.$$

Observe that there is not a line that passes through these three point. As one can attempt to solve for $y = bx + a$ s.t. $\begin{cases} 2 = b + a \\ 2 = 2b + a \\ 4 = 3b + a \end{cases} \implies \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$. Which is an inconsistent system of equations. Let e denote the difference between our estimation and the actual data, i.e., $y_i - \hat{y}_i$. And so we have

$$e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}.$$

Note that

$$\|e\|_2 = \sqrt{\sum_{i=1}^{n} |e_i|^2} \iff (\|e\|_2)^2 = \sum_{i=1}^{n} e_i^2.$$

This is what we call the *least squares* where $\|e\|$ is the least squares error of the approximation. The line that minimizes the least squares error of the approximation is called the *least square approximation line.*

Suppose we have a set of data

$$\{(x_1, y_2), \ldots, (x_n, y_n)\}.$$

Consider

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, x = \begin{bmatrix} a \\ b \end{bmatrix}, c = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

A quick computation shows that $e = c - Ax$. It follows that finding a line of best fit is equivalent to fining a vector

$$\vec{x} = (a, b) \in \mathbb{R}^2 \text{ s.t. } \|c - A\vec{x}\| \leq \|c - A\vec{x}\|, \forall x \in \mathbb{R}^2.$$

Definition 5.7.1 (Least Squares Solutions)**.** If A is an $m \times n$ matrix and $b \in \mathbb{R}^m$, a least squares solution of $Ax = b$ is a vector $\vec{x} \in \mathbb{R}^n$ such that

$$\|c - A\vec{x}\| \leq \|c - Ax\|, \forall x \in \mathbb{R}^n.$$

where $\|\cdot\|$ is induced by the standard inner production.

Theorem 5.7.1.1 (The Least Squares Theorem)**.** Let A be an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then $Ax = b$ always has at least one least squared solution $\bar{x}$. Moreover,

1. If x is a least square solution of $Ax = b$, then x is a solution of the system of equations $A^T Ax = A^T b$
2. A has linearly independent columns iff $A^T A$ is invertible. It follows from 1. that in this case, the least squares solution of $Ax = b$ is unique and is given by

$$\bar{x} = \left(A^T A\right)^{-1} A^T b$$

Proof. Suppose we have a set of data with $k - 1$ explanatory variables, n observations, and a dependent variable y:

$$\{(x_{1,i}, \ldots, x_{k-1,i}, y_i)\}_{i=1}^n .$$

Thus, the model is

$$Y_{\textit{population}} = X_{\textit{population}}\beta + \varepsilon,$$

where ε is the noise. Consider a multiple regression model estimation thus built:

$$\hat{Y} = \hat{X}\hat{\beta},$$

where $\hat{X} \in \mathbb{R}^{n \times k}$ which records each of the observation; explicitly,

$$X := \begin{bmatrix} 1 & x_{1,1} & \ldots & x_{k-1,1} \\ 1 & x_{1,2} & \ldots & x_{k-1,2} \\ 1 & \vdots & \vdots & \vdots \\ 1 & x_{1,n} & \ldots & x_{k-1,n} \end{bmatrix},$$

with $\hat{\beta} \in \mathbb{R}^k$ being the vector list of coefficient and intercept.

Let $Y := \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ be the vector recording the actual observed dependent variable.

Then, define the least square residual,

$$\|u\|_n = \left\| Y - \hat{Y} \right\|_n = \left\| Y - \hat{X}\hat{\beta} \right\|_n .$$

Clearly, the objective now is to minimize $\|u\|_n$.
Let $L_{\hat{X}} : \mathbb{R}^k \to \mathbb{R}^n$ be a linear transformation such that $\hat{\beta} \mapsto \hat{Y}$. Then, note that

$$\text{range}(L_{\hat{X}}) = \text{column}\left(\hat{X}\right) = \text{Span}\left\{ \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \dots, \begin{bmatrix} x_{k-1,1} \\ \vdots \\ x_{k-1,n} \end{bmatrix} \right\}.$$

Now to achieve the minimization, we want to find $\hat{X}\hat{\beta}$ satisfying

$$\left\| Y - \hat{X}\hat{\beta} \right\| \leq \|Y - \vec{z}\|, \forall \vec{z} \in \text{range}(L_{\hat{X}}) = \text{column}\left(\hat{X}\right).$$

By theorem, *best approximation exists and can be computed by its orthogonal projection*, the best approximation of Y onto column$(\hat{X})$ is thus the orthogonal projection of Y onto column $\left(\hat{X}\right)$,

$$\hat{X}\hat{\beta} = \text{proj}_{\text{column}\left(\hat{X}\right)}(Y) = E(Y).$$

Consider our $\hat{X}\hat{\beta}$. Note that

$$x_i^T\left(Y - \hat{X}\hat{\beta}\right) = \left\langle x_i^T | Y - \hat{X}\hat{\beta} \right\rangle = 0, \forall i$$

$$\text{as } x_i^T \in \text{column}\,\hat{X}, \text{ and } Y - \text{proj}_{\text{column}\left(\hat{X}\right)}(Y) = \text{perp}_{\text{column}\left(\hat{X}\right)}(Y).$$

From which it follows that

$$\begin{aligned} \hat{X}^T(Y - \hat{X}\hat{\beta}) &= 0 && \text{we stack all of } x_i^T \\ \implies \hat{X}^T Y - \hat{X}^T\hat{X}\hat{\beta} &= 0 && \text{associativity of matrix} \\ \implies \hat{X}^T Y &= \hat{X}^T\hat{X}\hat{\beta}. \end{aligned}$$

In particular, suppose $\hat{X}^T\hat{X}$ is invertible, i.e., *perfect-multicollinearity does not exists,* then

$$\hat{\beta} = \left(\hat{X}^T\hat{X}\right)^{-1}\hat{X}^T Y.$$

□

Thus, back to our ex,

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, b = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}.$$

To compute the line of best fit we solve

$$A^T A x = A^T.$$

That is,

$$\begin{aligned} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} \\ \implies \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} 8 \\ 18 \end{bmatrix} \\ \implies a &= \frac{2}{3} \text{ and } b = 1. \end{aligned}$$

Thus, we obtain $\hat{y} = x + \frac{2}{3}$.

6 Linear Operator and Adjoint Operators

6.1 Linear Functionals

Definition 6.1.1 (linear functionals and f_β)**.** A function f that maps a $\mathbb{F}$ vector space V to a scalar $c \in \mathbb{F}$ is called a *linear functional.*

Given any *inner product space* V and fixed vector β, f_β is the linear functional defined by:

$$f_\beta(\alpha) = \langle \alpha, \beta \rangle$$

Theorem 6.1.1.1 (Riesz Representation Theorem)**.** Let V be a *finite dimensional inner product space* and f be a *linear functional* on V. Then

$$\exists ! \beta \in V \text{ s.t. } f(\alpha) = \langle \alpha | \beta \rangle, \forall \alpha \in V.$$

Proof. Let $\{\alpha_1, \ldots, \alpha_n\}$ be an orthonormal basis of V. Let

$$\beta = \sum_{i=1}^{n} \overline{f(\alpha_i)} a_i$$

.

Claim. $f = f_\alpha$. It suffices to check $f_\beta\left(\alpha_k\right) = f\left(\alpha_k\right), \forall k$.

$$\begin{aligned} f_\beta\left(\alpha_k\right) &= \left\langle \alpha_k | \overline{f(\alpha_i)} a_i \right\rangle \\ &= \sum_{i=1}^{k-1} f(\alpha_k)\left\langle \alpha_k|\alpha_i\right\rangle + f(\alpha_k)\left\langle \alpha_k|\alpha_k\right\rangle + \sum_{i=k-1}^{n} f(\alpha_i)\left\langle \alpha_k|\alpha_i\right\rangle \\ &= f(\alpha_k). \end{aligned}$$

□

Now we prove uniqueness. Suppose $\exists \gamma, \beta$ s.t. $\langle \alpha|\beta\rangle = \langle \alpha|\gamma\rangle, \forall \alpha$. Then,

$$\begin{aligned} \langle \alpha|\beta\rangle - \langle \alpha|\gamma\rangle = 0 & \\ &\implies \langle \alpha|\beta - \gamma\rangle = 0 \\ &\implies \langle \beta - \gamma|\beta - \gamma\rangle = 0 \\ &\implies \beta - \gamma = 0 \\ &\implies \beta = \gamma. \end{aligned}$$

□

Remarks 6.1.1.1.1 (Necessity of Finite-Dimensionality)**.** It is possible for us to prove that the Riesz Representation Theorem is not true for infinite-diemensional vector spaces. Let the inner product be

$$\langle f|g\rangle = \int_0^1 (f\overline{g})(x)dx.$$

6.2 Adjoints

Theorem 6.2.0.1 (Adjoints exists for finite-dimensional vector spaces)**.** For any linear operator T on a finite-dimensional inner product space V,

$$\forall T \in \mathcal{L}V, \exists! T^* \in \mathcal{L}(V) \text{ s.t. } \forall \alpha, \beta \in V, \langle T\alpha|\beta\rangle = \langle \alpha|T^*\beta\rangle\,,$$

where T^* is the adjoint of T.

Proof. Let β be any vector in V with $T \in \mathcal{L}(V)$.
Let f be the function such that $\alpha \mapsto \langle T\alpha|\beta\rangle$, i.e., $f(\alpha) = \langle T\alpha|\beta\rangle$. Note that f is a linear functional. By *Riesz Representation Theorem*,

$$\exists! \beta' \text{ s.t. } \langle \alpha|\beta'\rangle = f(\alpha) = \langle T\alpha|\beta\rangle\,.$$

Which is true $\forall \alpha \in V$. Let $T^* : \beta \mapsto \beta'$. Then, $T^*\beta = \beta'$. Thus,

$$\langle \alpha|\beta'\rangle = \langle \alpha|T^*\beta\rangle = f\left(\alpha\right) = \langle T\alpha|\beta\rangle\,.$$

Now we prove linearity and uniqueness of T^*. If $\dim(V) = 0$, then all linear transformations are the same; thus suppose $\dim(V) > 0$. Then,

$$\begin{aligned}
\langle \alpha|T_1^*\beta\rangle &= \langle \alpha|T_2^*\beta\rangle \\
&\implies \langle \alpha|T_1^*\beta\rangle - \langle \alpha|T_2^*\beta\rangle = 0 \\
&\implies \langle \beta|T_1^*\beta - T_2^*\beta\rangle = 0, \forall \alpha, \beta \in V. \\
&\implies \langle T_1^*\beta - T_2^*\beta|T_1^*\beta - T_2^*\beta\rangle = 0, \forall \beta \in V \\
&\implies T_1^* = T_2^*.
\end{aligned}$$

Thus, T^* is unique. Now linearity is left as an exercise. □

Definition 6.2.1 (Adjoint)**.** Let T be a linear operator on an inner product space V. Then we say that T has an adjoint on V if there exists a linear operator T^* on V such that

$$\langle T\alpha \mid \beta\rangle = \langle \alpha \mid T^*\beta\rangle \quad \forall \alpha, \beta \in V.$$

(Note that if V is not finite dimensional, then T^* may not exist.)

Proposition 6.2.1.1 (Properties)**.**

1. If an adjoint exists, then it is unique
2. The adjoint depends on T and definition of $\langle \cdot|\cdot\rangle$.
3. If V is finite dimensional, then an adjoint always exists. Note that the converse does not hold.

Theorem 6.2.1.2 (Matrix Representations of linear transformations with respect to Orthonormal Bases)**.** Let $T \in \mathcal{L}(V)$ and suppose $\beta = \{\alpha_1, \ldots, \alpha_n\}$ is an *orthonormal* basis for V. Let $A = [T]_\beta = (A_{kj})$, where $1 \le k, j \le n$ then,

$$A_{kj} = \langle T\alpha_j | \alpha_k \rangle, \forall k, j \in [1, n] \cap \mathbb{N}.$$

Theorem 6.2.1.3 (Matrix Representation of Adjoints wrt Orthonormal Basis)**.** Let $V \in \mathcal{L}(V)$ and β be any ordered orthonormal basis $\{\alpha_1, \ldots, \alpha_n\}$. Let $[T]_\beta = (A_{kj})$ then

$$[T^*]_\beta = \left(\overline{A}_{jk}\right) = (\langle \alpha_j | T^* \alpha_k \rangle) = [T]^*_\beta.$$

(The conjugate transpose of A), i.e., the matrix representation for T^* given the ordered basis β is the *conjugate transpose* of the matrix representation for T using the same basis.

Proof. Let $A = [T]_\beta$ and $B = [T^*]_\beta$. Then,

$$A_{jk} = \langle T\alpha_k | \alpha_j \rangle \text{ and } B_{kj} = \langle T^* \alpha_j | \alpha_k \rangle$$

But

$$\overline{B_{kj}} = \langle \alpha_k | T^* \alpha_j \rangle = \langle T\alpha_k | \alpha_j \rangle = A_{jk}.$$

□

Ex. Consider $A = \begin{bmatrix} 2 & 1+i \\ i & 3-i \end{bmatrix}$. Compute A^*.

Solution.

$$A^* = \begin{bmatrix} \overline{2} & \overline{i} \\ \overline{1+i} & \overline{3-i} \end{bmatrix} = \begin{bmatrix} 2 & -i \\ 1-i & 3+i \end{bmatrix}.$$

■

Corollary 6.2.1.3.1 ($(T^*)^* = T$). Let V be a finite-dimensional vector space and $T \in \mathcal{L}(V)$, then

$$(T^*)^* = T.$$

Proof. As exercise. □

Remarks 6.2.1.3.1 (Orthonomral basis is required)**.** If β is not an orthonormal basis, then the matrix representation of T^* $[T^*]_\beta$ is not necessarily equal to the conjugate transpose of the matrix representation of T ,i.e., $\left([T]_\beta\right)^* \neq [T^*]_\beta$ necessarily.

Definition 6.2.2 (Self Adjoint - Hermition)**.** A *linear operator* $T \in \mathcal{L}(V)$ is *self adjoint* or *Hermition* if

$$T = T^* \left(A \in \mathbb{F}^{n \times n}; A = A^*\right).$$

The matrix representation of an inner product is hermitian matrix. Thus, its name hermitian inner product spaces.

6.3 Pauli Matrices

It is well known that *observables* (quantum properties that we can measure) in quantum mechanics are all represented using *Hermition matrices.* The *Pauli Matrices* are a special set of Hermitian matrices in $\mathbb{C}^{2\times 2}$ that are connected with the "angular momentum" relations in quantum mechanics

Definition 6.3.1 (Pauli Matrices)**.** There are four Puali Matrices

1. $I = (\sigma_0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
2. $X = (\sigma_1) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
3. $Y = (\sigma_2) = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}$
4. $Z = (\sigma_3) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

Proposition 6.3.1.1 (Properties of Pauli Matrices)**.** The Pauli Matrices have some very interesting properties.

$$XY = -iZ, \quad YZ = -iX, \quad ZX = -iY$$

$$YX = iZ, \quad ZY = iX, \quad XZ = iY$$

Also,

$$XX = YY = ZZ = I.$$

Thus,

$$\{\pm X, \pm Y, \pm Z, \pm I, \pm iX, \pm iY, \pm iZ, \pm iI\}$$

form a *group* with respect to Matrix Multiplication.

Remarks 6.3.1.1.1. Pauli Matrices can send a message that is resistant to probabilistic attack. As we shall see, Pauli Matrices are self adjoint and *unitary.*

Definition 6.3.2 (Preserving Inner Products and Vector Isomorphism)**.** Let V and W be inner product spaces over $\mathbb{F}$, $T \in \mathcal{L}(V, W)$, then T **preserves** inner **products** if

$$\langle T\alpha | T\beta \rangle_\beta = \langle \alpha | \beta \rangle, \forall \alpha, \beta \in V.$$

Note a vector isomorphism of V onto W is an *inner product isomorphism* T of V onto W, which preserves the inner products.

6.4 Unitary Operator

Definition 6.4.1 (Unitary Operator)**.** A **unitary operator** on an inner product space is an (inner product) vector isomorphism of the space to itself. That is,

$$T \in \mathcal{L}(V) \text{ s.t. } \langle T\alpha | T\beta \rangle = \langle \alpha | \beta \rangle, \forall \alpha, \beta \in V.$$

Definition 6.4.2 (Orthogonal and Orthonormal Matrices)**.**
A matrix $A \in \mathbb{R}^{n\times n}$ is **orthogonal** (real) if $AA^T = A^TA = I_n$.
A matrix $A \in \mathbb{C}^{n\times n}$ is **orthonormal** (complex) if $AA^* = A^*A = I_n$.

Remarks 6.4.2.0.1. Orthoonal and Orthonormal Matrices are unitary operators acting on $M_{n\times 1}(\mathbb{R})$ and $M_{n\times 1}(\mathbb{C})$ respectively.

Ex.

1. Identity Matrix/Operator : $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
2. *Pauli Matrices.*
3. $\operatorname{diag}\left(e^{i\theta_1}, \dots, e^{i\theta_n}\right)$, where $\theta_j \in [0, 2\pi)$.
4. Rotation Matrix.

Proof.

$$I(I)^* = II = I.$$

Now we show

$$\operatorname{diag}\left(e^{i\theta_1} \dots, e^{i\theta_n}\right) \cdot \left[\operatorname{diag}\left(e^{i\theta_1} \dots, e^{i\theta_n}\right)\right]^* = I$$

$$\begin{aligned} \text{The given } &= \operatorname{diag}\left(e^{i\theta_1} \dots, e^{i\theta_n}\right) \cdot \operatorname{diag}\left(e^{\overline{i\theta_1}} \dots, e^{\overline{i\theta_n}}\right) \\ &= \operatorname{diag}\left(e^{i\theta_1} \dots, e^{i\theta_n}\right) \cdot \operatorname{diag}\left(e^{-i\theta_1} \dots, e^{-i\theta_n}\right) \\ &= \operatorname{diag}\left(e^{i\theta_1}e^{-i\theta_1}, \dots, e^{i\theta_n}e^{-i\theta_n}\right) \\ &= \operatorname{diag}(1, \dots, 1) \\ &= I. \end{aligned}$$

□

Theorem 6.4.2.1 (Characterizations of Unitary Operators for Finite Dimensional Vector Spaces)**.**

1. U is unitary
2. $UU^* = I = U^*U$
3. U is inner-product preserving:
   $$\langle Ux, Uy \rangle = \langle x, y \rangle \quad \text{for all } x, y \in V$$
4. U is norm preserving:
   $$\|Ux\| = \|x\| \quad \text{for all } x \in V$$
5. If $\{\alpha_1, \ldots, \alpha_n\}$ is an *orthonormal basis* for V, then $\{U\alpha_n, \ldots, U\alpha_n\}$ is also an *orthonormal basis* for V.
6. If β is any *orthonormal basis* for V, then the *columns* of $[U]_\beta$ are *orthonormal* in $\mathbb{F}^n$ with the standard inner product.
7. If β is any *orthonormal basis* for V, then the *rows* of $[U]_\beta$ are *orthonormal* in $\mathbb{F}^n$ with the standard inner product.

Proof. 1. $\iff$ *2.* □

6.5 Quantum Application

A quantum system is represented using a complete inner product space, also known as a Hilbert Space. For finite dimensional systems, a finite inner product space is a complete inner product space. You will discuss this more in a real analysis class. Every state that the *quantum state* can be in is represented using a unit vector. **Ex.** The spin of an electron.

$$e_1 = (1,0) = \text{“}|0\rangle''\text{represents the spin up of an electron}$$
$$e_2 = (0,1) = \text{“}|1\rangle''\text{represents the spin down of an electron}$$

A system state could be

$$\alpha = \frac{1}{\sqrt{2}}\left(e_1 + e_2\right) = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right).$$

This is known as the quantum superposition of $|0\rangle$ and $|1\rangle$. Every quantum operation maps a state of a system to a new state. Let a quantum operation be represented using the following matrix:

$$U = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad U|0\rangle = u\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) = \alpha.$$

Since every state is represented using a *unit vector,* any *quantum operator* $U \in \mathcal{L}(V)$ is *norm preserving.* Therefore, it is a unitary operator. Since unitary operators satisfy the property that $UU^* = I$, we can state that “all quantum operations are reversible.

6.6 Normal Operator

Definition 6.6.1 (Normal Operators)**.** Let V be a finite-dimensional inner product space and $T \in \mathcal{L}(V)$. Then, T is normal if it commutes with it's adjoint:

$$TT^* = T^*T.$$

Theorem 6.6.1.1 (Basic Properties of Normal Operators)**.** Let V be a finite-dimensional inner product space and T be a normal operator on V.
Then

1. $\|T\alpha\| = \|T^*\alpha\|$
2. for any scalar c, $(T - cI)$ is also normal
3. $(T - cI)^* = T^* - \bar{c}I$
4. T has a eigenvector α with eigenvalue c if and only if T^* has characteristic vector α with characteristic value $\bar{c}$.

From 1,2,3,

$$\|(T - cI)\,\alpha\| = \|(T^* - \bar{c}I)\,\alpha\|\,.$$

It follows that

$$T\alpha = c\alpha \iff T^*\alpha = \bar{c}\alpha.$$

Definition 6.6.2. A Matrix $A \in \mathbb{C}^{n\times n}$ is *normal* if $AA^* = A^*A$.

Ex.

1. Identity
2. Hermition Matrices
3. Unitary Matrices

Theorem 6.6.2.1 (Uppe Triangular Matrices are Normal iff they are diagonal)**.** Let V be a finite-dimensional inner product space, $T \in \mathcal{L}(V)$, and β be an orthonormal basis for V. Suppose that the matrix $A = [T]_\beta$ is upper triangular, then T is normal iff A is a diagonal matrix.

Proof. Since β is an orthonormal basis, $A^* = [T^*]_\beta$. Suppose A is diagonal. Then, $A = \operatorname{diag}(a_{11}, \dots, a_{nn})$ and $A^* = \operatorname{diag}(\bar{a_{11}}, \dots, \bar{a_{nn}})$.
Thus, $[TT^*]_\beta = AA^* = \operatorname{diag}(a_{11}\bar{a_{11}}, \dots, a_{nn}\bar{a_n}n)$.
Similarly,

$$[T^*T] = \operatorname{diag}(a_{11}\bar{a_{11}}, \dots, a_{nn}\bar{a_n}n)\,.$$

Thus, $[TT^*]_\beta = [T^*T]_\beta$. Thus, $TT^* = T^*T$.
Now suppose T is normal and $A = [T]_\beta$. Then, $A^* = [T^*]_\beta$. If A is upper triangle and $\beta = \{v_1, \dots, v_n\}$, then

$$[T(v_1)]_\beta = \begin{bmatrix} a_{11} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \implies T(v_1) = a_{11} \implies T^*(v_1) = \bar{a_{11}}v_1.$$

From which it follows that

$$[T^*(v_1)]_\beta = \begin{bmatrix} \bar{a_{11}} \\ \vdots \\ 0 \end{bmatrix} \implies a_{1j} = 0, \forall j \in \{2, \ldots, n\}.$$

Repeating this argument forces all of the diagonal entries to 0. Thus, A is diagonal. □

Theorem 6.6.2.2 (Orthogonal Complements of Invariant Subspaces are Invariant under Adjoints)**.** Let V be an finite-dimensional inner product space ad that $T \in \mathcal{V}$. Sps W is a subspace of V which is invariant under T. Then, $W^\perp$ is invariant under T^*, i.e.,

$$(w \in W \implies T(w) \in W) \implies \left(w \in W^\perp \implies T^*(w) \in W^\perp\right).$$

Proof. Let $\alpha \in W$ and $\beta \in W^\perp$. Suppose $T\alpha \in W$. Then,

$$\langle T\alpha | \beta \rangle = 0.$$

Thus,

$$\langle \alpha | T^* \beta \rangle = 0 \text{ as } \langle T\alpha | \beta \rangle = \langle \alpha | T^* \beta \rangle.$$

Thus, $T^*\beta \in W^\perp$. □

Theorem 6.6.2.3 (Schur's Decomposition)**.** Let V be a finite-dimensional complex inner product space and let T be any linear operator on V. Then there is an orthonormal basis for V in which the matrix of T is *upper triangular.*

Proof. We prove by induction. Base case: let $\dim(V) = 1$. Then,

$$[T]_\beta = [a_{11}], \text{ which is upper triangular.}$$

$n+1$ case: V is an $n+1$ dimensional complex vector space. Assume that if W is a complex inner product space of dimension n, there exists an orthonormal basis such that the matrix of T is upper triangular. As T is an operator on a complex vector space, there exists a non-zero vector such that $Tv = cv$. Additionally, if $\alpha = \frac{v}{\|v\|}$, then, $T\alpha = c\alpha$. Let $W = \operatorname{Span}(\{\alpha\})$. Let $W^\perp = \operatorname{Span}(\{v_1, \ldots, v_n\})$. Let $\beta = \{\alpha, v_1, \ldots, v_n\}$. Then,

$$[T]_\beta = \begin{bmatrix} c & V & W \\ 0 & \hat{W} & \hat{W} \\ 0 & \hat{W} & \hat{W} \end{bmatrix}.$$

Let $S_{\hat{W}} : \operatorname{Span}(\{v_1, \ldots, v_n\}) \to \operatorname{Span}(\{v_1, \ldots, v_n\})$ be the linear transformation defined using $\hat{W}$ in basis

$$\hat{\beta} = \{v_1, \ldots, v_n\}.$$

By our inductive hypothesis, there exists a basis $\hat{\beta}' = \{\alpha_1, \dots, \alpha_n\}$ s.t. $[S_{\hat{W}}]_{\hat{\beta}'}$ is upper triangular. Let P be the change of basis matrix from $\hat{\beta} \to \hat{\beta}'$. Then,

$$[S_{\hat{W}}]_{\hat{\beta}} = P^{-1}[S_{\hat{W}}]P \implies [S_{\hat{W}}]_{\hat{\beta}'} = P[S_{\hat{W}}]_{\hat{\beta}}P^{-1} = P\hat{W}P^{-1}.$$

Notice $\beta' = \{\alpha, \alpha_1, \dots, \alpha_n\}$ is an orthonormal basis for V. Additionally, the matrix $\begin{bmatrix}1 & 0\\ 0 & P\end{bmatrix}$ is the change of basis matrix from $\beta \to \beta'$. It follows that $[T]_\beta = \begin{bmatrix}1 & 0\\ 0 & P\end{bmatrix}^{-1}[T]_{\beta'}\begin{bmatrix}1 & 0\\ 0 & P\end{bmatrix}$.
Thus,

$$[T]_{\beta'} = \begin{bmatrix}1 & 0\\ 0 & P\end{bmatrix}[T]_\beta\begin{bmatrix}1 & 0\\ 0 & P^{-1}\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & p\end{bmatrix}\begin{bmatrix}c & v\\ 0 & w\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & p^{-1}\end{bmatrix} = \begin{bmatrix}c & p^{-1}v\\ 0 & w\end{bmatrix}.$$

Thus, the matrix is upper triangular. □

Remarks 6.6.2.3.1. Shcur's Decomposition tells us that when V is a complex inner product space, there exists an orthonormal basis β s.t. T is upper triangular. We say that T *unitarly triangalizable.* This leads to the following definition.

Definition 6.6.3 (Unitarily Diagonalizable)**.** *Let V be a finite-dimensional vector space, $T \in L(V)$. T is unitarily diagonalizable if there exists an orthonormal basis such that the matrix representation of T with respect to this basis is diagonal.*

Definition 6.6.4 (Unitary triangularizable and unitary diagonalizable matrices)**.** A matrix A is *unitary upper triangularizable* if there exists a unitary matrix P such that

$$A = P^{-1}BP.$$

where B is upper triangular.

A matrix is *unitary* diagonalizable if there exists a unitary matrix P such that

$$A = P^{-1}DP,$$

where D is a diagonal matrix.

Theorem 6.6.4.1 (Spectral Theorem for Normal Operators)**.** Let V be a finite-dimensional complex inner product space. Let $T \in L(V)$.

The following are equivalent:

1. T is unitarily diagonalizable
2. T is normal

Proof. Exc: $1 \to 2$. By Schur's Deconposition, $\exists$ orthonormal basis s.t. $[T]_\beta$ is upper triangular. By our previous theorem, a matrix representation of a normal operator is triangular implies that it is diagonal. □

Definition 6.6.5 (Spectrum of a Matrix)**.** Given $A \in \mathbb{F}^{n\times n}$, the set of characteristic values (eigenvalues of A) is called the spectrum of A, denoted by

$$\sigma(A) := \{\lambda \in \mathbb{F} \text{ s.t. } \exists \alpha \neq 0, A\alpha = \lambda\alpha, \}.$$

This implies that $\sigma(A) \subseteq \mathbb{F}$, and $\sigma(A)$ has at most n values.

Remarks 6.6.5.0.1. Let A and B be similar matrices. Then,

$$1. \det(A) = \det(B), \quad 2. \operatorname{trace}(A) = \operatorname{trace}(B), \quad 3. \sigma(A) = \sigma(B).$$

Theorem 6.6.5.1 (Relationshop between the trance, the determinant and characteristic values)**.** Let A be $n \times n$ matrix with $\mathbb{C}$ and $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A, repeating the terms if there is algebraic multiplicity. Then,

$$1. \operatorname{trace}(A) = \sum_{i=1}^{n} \lambda_i, \quad 2. \det(A) = \prod_{i=1}^{n} \lambda_i.$$

Theorem 6.6.5.2 (The spectrum of Unitary Operators on a complex inner product lies on the unit circle)**.** Let V be a complex inner product space and $U \in \mathcal{L}(V)$ be unitary and $|\cdot|$ be the modulus of a complex number. Then,

$$\sigma(U) \subseteq \{\lambda \in \mathbb{C} \text{ s.t. } |\lambda| = 1\}.$$

Proof. If $Ux = \lambda x$, then, $|\lambda| = 1$. Note that

$$\begin{aligned}
\|x\| &= \|Ux\| \\
&= \|\lambda x\| \\
&= |\lambda|\,\|x\| \\
&\implies \|x\| - |\lambda|\,\|x\| = 0 \\
&\implies \|x\|(1 - |\lambda|) = 0 \\
&\implies |\lambda| = 1 && \text{since } \|x\| \neq 0.
\end{aligned}$$

□

Theorem 6.6.5.3 (Spectral Theorem for Unitary Operators)**.** Let $U \in \mathcal{L}(V)$ be unitary. Then an orthonormal basis β of unit vectors of V exists, such that

$$[U]_\beta = \operatorname{diag}(e^{i\theta})$$

where $e^{i\theta_1}, \ldots, e^{i\theta_n}$ are the characteristic values of U.

Theorem 6.6.5.4 (Unitary Similarity)**.** Let $U \in \mathbb{F}^{n\times n}$ be a unitary matrix. Then there exists a unitary matrix $V \in \mathbb{F}^{n\times n}$ such that

$$VUV^* = \operatorname{diag}(e^{i\theta})$$

where $e^{i\theta_1}, \ldots, e^{i\theta_n}$ are the characteristic values of U including multiplicities. We say that U is *unitarily similar* to the diagonal matrix.

Remarks 6.6.5.4.1. All quantum operators are unitarily similar to a diagonal matrix.

7 Polynomials

7.1 Polynomials

Definition 7.1.1 (Polynomials)**.** Let $\mathbb{F}[x] :=$ the set of polynomials with coefficients in $\mathbb{F}$. Let

$$f(x) = c_0 + c_1x^1 + \ldots + c_kx^k, \quad g(x) = a_0 + \ldots + a_nx^n \in \mathbb{F}[x].$$

We define the degree of f, denoted $\deg(f)$, to be the number k, where $c_k \neq 0$. If $\nexists k > 0$ such that $c_k \neq 0$, then $\deg(f) = 0$. We define $g(x) = f(x)$ if all the coefficients of f and g are equal (i.e. $a_i = c_i$ for all $0 \leq i \leq k$).

Furthermore, if $c_k = 1$, then we say f is .

Remarks 7.1.1.0.1 (Polynomials v.s. Polynomial functions)**.** It is important to note that our definitions of polynomials are purely formal, and are not to be confused with polynomial functions.

For ex, the polynomial function $f : \mathbb{R} \to \mathbb{R}$ defined by the rule $x \mapsto x^2$ is not the same as the polynomial $x^2 \in \mathbb{F}[x]$. The former is a function, whereas the latter is merely a symbol.

Definition 7.1.2 (Applying Polynomials to Operators: $f(T)$ and $f(A)$)**.** Let $f = c_0 + c_1x + \cdots + c_kx^k \in \mathbb{F}[x], T \in \mathcal{L}(V)$ and $A \in M_{n\times n}(\mathbb{F})$ then,

$$f(T) = c_0I + c_1T + \ldots c_kT^k = \sum_{i=0}^{k} c_iT^i$$
$$f(A) = c_0I + c_1A + \ldots c_kA^k = \sum_{i=0}^{k} c_iA^i.$$

Ex. $A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}$. $f(x) = x^2 - x + 2$. Then,

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

Ex. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T(x_1, x_2) = (x_2, 3x_1 - x_2)$. Find $f(T)$where $f(x) = x^2 - 1$.

Solution.

$$\begin{aligned} f(T)(x_1, x_1) &= \left(T^2 - I\right)(x_1, x_2) \\ &= (2x_1 - x_2, -3x_1 + x_2). \end{aligned}$$

■

7.2 Polynomial Ideals

We now discuss some factorization results on polynomials. As we shall see later, these results will be helpful in the quest of finding nice matrix representation. There are two main factorization results that we shall discuss: the *Division Algorithm for Polynomials and the Fundamental Theorem of Algebra*. Then, we will provide a new alternative definition for polynomial ideals.

Theorem 7.2.0.1 (Division Algorithm for polynomials)**.** If $f, d \in \mathbb{F}[x]$ and $d \neq 0$, then $\exists! q \in \mathbb{F}[x]$ s.t.

1. $f = dq + r$
2. either $r = 0 \vee \deg(r) < \deg(d)$.

If the remainder $r = 0$, we say

1. d divides f
2. f is a *multiple* of d
3. q is the quotient of f and d.

Corollary 7.2.0.1.1 (Fundamental Remainder Theorem)**.** Let $c \in \mathbb{F}$. Then $f(x)$ is divisible by $x - c$ if and only if $f(c) = 0$.

Theorem 7.2.0.2 (Division Algorithm for Integers)**.** If $f \in \mathbb{Z}_+ \cup \{0\}$ and $d \in \mathbb{Z}_+$, then a unique $q, r \in \mathbb{Z}_+ \cup \{0\}$ exists such that

1. $f = dq + r$
2. either $r = 0$ or $|r| < |d|$

If the remainder $r = 0$, we say

1. d **divides** f
2. f is a **multiple** of d
3. q is the **quotient** of f and d

Ex. Let $f = 16$ and $d = 3$. Solve for q and r s.t. $f = qd + r$.

Solution. Note that

$$16 = 5(3) + 1 \implies q = 5 \text{ and } r = 1.$$

■

Ex. Let $f(x) = x^2 + x + 1$ and $d(x) = x - 1$. Solve for $q(x)$ and $r(x)$ s.t. $f(x) = q(x)d(x) + r(x)$.

Solution. Noe that

$$x^2 + x + 1 = (x + 2)(x - 1) + 3.$$

■

Ex. $f(x) = x^4 - 2x^3 - 2x^2 - 2x - 3$ and $d(x) = x^3 + 6x^2 + 7x + 1$. Solve for $q(x)$ and $r(x)$ s.t. $f(x) = q(x)d(x) + r(x)$.

Solution. Note that

$$x^4 - 2x^3 - 2x^2 - 2x - 3 = (x - 8)(x^3 + 6x^2 + 7x + 1) + (39x^2 + 53x + 5).$$

■

Definition 7.2.1 (Irreducible and Prime Polynomials). A function $f \in \mathbb{F}[x]$ is *reducible* over $\mathbb{F}$ if two polynomials $g, h \in \mathbb{F}[x]$ exist such that

1. $\deg(g) \geq 1$ and $\deg(h) \geq 1$
2. $f = gh$

Otherwise f is *irreducible* over $\mathbb{F}$.

A polynomial $p \in \mathbb{F}[x]$ is a *prime polynomial* if it

1. $p \neq 0$
2. $\nexists f \in \mathbb{F}[x]$ such that $fp = 1$
3. If $p \mid ab$, then $p \mid a$ or $p \mid b$

Remarks 7.2.1.0.1 (Irreducble v.s. Prime). A non-scalar irreducible polynomial is a *prime polynomial.*

In higher-level algebra courses, you may explore the differences between being irreducible and prime, but we do not explore this here.

Theorem 7.2.1.1 (Fundamental Theorem of Arithmetic). If $n \in \mathbb{Z}$ then a unique set of positive primes exist such that

$$n = up_1^{n_1} \dots p_k^{n_k}$$

where $u = -1$ if n is negative and otherwise 1.

Theorem 7.2.1.2 (Prime Factorization of Polynomials). Let $f \in \mathbb{F}[x]$ be a non-scalar polynomial. Then there are unique distinct prime polynomials $p_1, \dots, p_n \in \mathbb{F}[x]$, positive integers $e_1, \dots, e_n$, and scalar a such that

$$f = ap_1^{e_1}p_2^{e_2} \cdots p_n^{e_n}.$$

Theorem 7.2.1.3 (Fundamental Theorem of Algebra). Every *non-zero polynomial in* $\mathbb{C}[x]$ factors completely; i.e.,

$$\exists c_1, \dots, c_k \in \mathbb{C} \text{ and } \exists r_1, \dots, r_k \in \mathbb{Z} \text{ such that}$$

$$f(x) = (x - c_1)^{r_1}(x - c_2)^{r_2} \cdots (x - c_k)^{r_k}$$

The roots of $f(x)$ are $c_1, \dots, c_k$ and the respective multiplicity is $r_1, \dots, r_k$.

Remarks 7.2.1.3.1 (Prime Factorization of Complex Polynomials). An immediate corollary of the fundamental theory of algebra is that the only prime polynomials in $\mathbb{C}\,[x]$ are the linear factors of the form $(x - c)$ which have degree 1. This is *not* necessarily true for *polynomial* over $\mathbb{R}$. To better understand this, consider the following ex.

Ex. Let $f(x) = x^3 - 1$. Find the prime factors of $f(x)$ over $\mathbb{R}$ and $\mathbb{C}$.

Solution. Over $\mathbb{R}$.

$$x^3 - 1 = (x-1)(x^2+x+1) \implies x = 1 \vee x = \frac{-1 \pm \sqrt{1^2 - 4}}{2} = \frac{-1}{2} + \frac{\sqrt{3}}{2} i.$$

But the latter is not in $\mathbb{R}$. Thus, the *primary decomposition* of $f(x)$ is

$$(x-1)\left(x^2+x+1\right).$$

Over $\mathbb{C}$.

$$x^3 - 1 = (x-1)\left(x - \left(\frac{-1}{2} + \frac{\sqrt{3}}{2} i\right)\right)\left(x - \left(\frac{-1}{2} - \frac{\sqrt{3}}{2} i\right)\right)$$

■

Theorem 7.2.1.4 (Alternate Characterizations of Polynomial Ideals)**.** Let $M \subset \mathbb{F}[x]$. The following are equivalent:

1. M is an ideal
2. M is a nonempty subset such that:
   (a) if $f \in \mathbb{F}[x]$ and $g \in M$, then, $fg = gf \in M$
   (b) if $f, g \in M$, then, $f + g \in M$.

Corollary 7.2.1.4.1 (Polynomial Ideals are Subspaces)**.** Let M be an ideal of $\mathbb{F}[x]$. Then, M is a subspace of $\mathbb{F}[x]$.

Proof. Let $c \in \mathbb{F}$ and $f, g \in M$. Note $c \in \mathbb{F}[x]$. Thus,

$$cf \in M \text{ and } cf + g \in M.$$

Thus, by the *characterization by closure theorem,* M is a subspace. □

Definition 7.2.2 (Finitely Generated Polynomial Ideal)**.** Let $D = \{d_1, \ldots, d_n\} \subseteq \mathbb{F}[x]$. Then, the *ideal generated* by D is the set

$$\mathbb{F}[x] D = \{f_1 d_1 + \cdots + f_n d_n \text{ s.t. } f_i \in \mathbb{F}[x]\}$$

Definition 7.2.3 (Principal Ideal)**.** If D contains only a single element $d \in \mathbb{F}[x]$ then the ideal M generated by d is called a *principal ideal,*

$$M = \mathbb{F}[x] d = \{fd \text{ s.t. } f \in \mathbb{F}[x]\}.$$

Ex. $d(x) = x^2$. The principal ideal $\mathbb{C}[x]\,d$ is the set of all polynomials that have x^2 as a factor.

$$\mathbb{C}[x]\,x^2 = \left\{x^2lx^2 + ix^3 + \dots\right\}.$$

Ex. Show that $\mathbb{F}[x]$ is a principal ideal. What is the generator.

Solution.

$$\mathbb{F}[x] = \mathbb{F}[x]\,1$$

Note that if $f \in \mathbb{F}[x]$, then $(f \cdot 1)(x) = f(x), \forall x$. Thus, $f \in \mathbb{F}[x] \cdot 1$. ■

Ex. Let $d_1(x) = x + 2$, $d_2(x) = x^2 + 8x + 16$. Let $M = \mathbb{F}[x]d_1 + \mathbb{F}[x]d_2$.
What is interesting about this exis that $M = \mathbb{F}[x]$.

$$\frac{1}{4}d_2(x) - \frac{1}{4}(x+6)d_1 = 1 \in M$$

$$\therefore \text{ the constant function } 1 \in M$$

$$\implies \mathbb{F}[x] = \mathbb{F}[x] \cdot 1 \subseteq M \subseteq \mathbb{F}[x]$$

$$\therefore M = \mathbb{F}[x]$$

M is the principal ideal generated by 1.

Theorem 7.2.3.1 ($\mathbb{F}[x]$ is a Principal Ideal Domain)**.** Let M be a non-zero ideal in $\mathbb{F}[x]$, then exists a *unique monic* polynomial $d \in \mathbb{F}[x]$ s.t. M is the principal ideal generated by d, i.e., $M = \mathbb{F}[x]\,d$.

Proof. Left as an exercise. □

Definition 7.2.4 (GCD of Polynomials)**.** Let $p_1, \dots, p_n \in \mathbb{F}[x]$. A **monic** polynomial $d \in \mathbb{F}[x]$ such that:

1. $d \in \mathbb{F}[x](\{p_1, \dots, p_n\}) = \mathbb{F}[x]p_1 + \dots + \mathbb{F}[x]p_n$

2. d divides each of the p_i

3. d is divisible by every polynomial which divides each of $p_1, \dots, p_n$

is called the **greatest common divisor** (gcd) of $p_1, \dots, p_n$.
We say $p_1, \dots, p_n$ are **relatively prime** if $\gcd(p_1, \dots, p_n) = 1$.

Remarks 7.2.4.0.1 (Problem with definition)**.** The careful reader may note some strange things about this definition:

1. How do we know d exists?
2. We referred to d as THE greatest common divisor despite not knowing that d is unique. Is d indeed unique?

We now make amends to these holes, showing that the greatest common divisors do indeed exist and that they are unique.

Theorem 7.2.4.1 (The GCD is the unique monic generator)**.** Let $p_1, \ldots, p_n, d \in \mathbb{F}[x]$. The following are equivalent:

1. d is the unique monic generator of $\mathbb{F}[x](\{p_1, \ldots, p_n\})$
2. d is a gcd of $p_1, \ldots, p_n$.

Definition 7.2.5 (Te Euclidean Algorithm)**.** Suppose you are given a and b. WLOF, assume $a > b$. The following step compute the gcd,

1. set $f = a, d = b$ then solve for p, r s.t. $f = qd + r$.
2. while $r \neq 0$ set $f = d, d = r$ and solve for q, r s.t. $f = qd + r$.
3. When $r = 0, \gcd(a, b) = d$ from the last iteration.

Ex. Find $\gcd(20, 15)$.

Solution.

$$\begin{aligned} 20 &= 1 \times 15 + 5 \\ &\implies 15 = 3 \times 5 + 0 \\ &\implies \gcd(20, 15) = 5. \end{aligned}$$

■

Definition 7.2.6 (Euclidean Algorithm for Polynomials)**.** Given f_1 and f_2. WLOG, let $\deg(f_1) \geq \deg(f_2)$.

1. Set $f = f_1$, $d = f_2$, then solve for q, r such that $f = qd + r$.
2. While $r \neq 0$, set $f = d$, $d = r$ and solve for q, r such that $f = qd + r$.
3. When $r = 0$, $\gcd(a, b) = a_k^{-1} d$ where q_k is the lead coefficient of d from your last iteration.

Ex. Let $d_1 = x^3 + 2x^2 - x - 2$ and $d_2 = x^2 + 2x + 1$. What is the $\gcd(d_1, d_2)$?

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

So we write:

$$d_1(x) = x \cdot d_2(x) + (-2)(x + 1) \Rightarrow \gcd(d_1, d_2) = \gcd(d_2, r)$$

Now divide:

$$\begin{aligned} \frac{x^2 + 2x + 1}{x + 1} &= \tfrac{x}{2} + \tfrac{1}{2} \\ x^2 + 2x + 1 &- (x^2 + 2x + 1) = 0 \end{aligned}$$

We conclude that $\gcd(d_1, d_2) = x + 1$.

The only property that *stops* polynomials from being a field is the lack of a *multiplicative inverse.* Note that this is the same property that *stops* the *integers* from having a *multiplicative inverse.* As we discussed in the first lecture, there were two ways we were able to create a *multiplicative inverse.*

1. *Rational Numbers*

2. *Modular Arithmetic*

Theorem 7.2.6.1. The set of rational functions,

$$\left\{ f\left(x\right) = \frac{h\left(x\right)}{g\left(x\right)} \text{ s.t. } h(x), g(x) \in \mathbb{F}\left[x\right] \text{ and } g\left(x\right) \neq 0 \right\}$$

is a field.

Proof. Exercise. □

In the first lecture we claimed this was true, but we shall now prove that it is true for integers in such a way that the same result holds for polynomials.

As we discussed in the first week of lectures, the integers can be used to create a field using modular arithmetic, but why? It is a result of it being a *principal ideal domain.* As polynomials are also *principal ideal domain*, we can also use a similar argument to prove that using modular arithmetic for polynomials also creates a field. Using the similar argument will hopefully make it easier to follow. The argument relies on the following theorem.

Theorem 7.2.6.2. Let $(\mathcal{R}, +, \times)$ be a principal ideal domain. Let $a, b \in \mathcal{R}$, then there exist a $c, d \in \mathcal{R}$ s.t.

$$ca + db = \gcd\left(a, b\right).$$

Proof. Let $a, b \in \mathcal{R}$. Let I be the ideal generated by ab. As I is an ideal, there exists a $c \in \mathbb{I}$ that generates I. Additionally, as I was generated by $\gcd(a, b)$. Thus, $\gcd(a, b) \in I$. Thus, $\exists c, d$ s.t. $ca + db \gcd(a, b)$. □

Theorem 7.2.6.3 (Creating a field from a Principal Ideal Domain)**.** Let $(\mathcal{R}, +, \times)$ be a principal ideal domain and $a \in \mathcal{R}$ be a prime element. Then, $\mathcal{R}$ mod a forms a field.

Proof. see lecture notes. □

7.3 Annihilating Polynomials

Remarks 7.3.0.0.1 (Defining Operators using Polynomials)**.** Let $T \in \mathcal{L}(V)$ where V is an $\mathbb{F}$vector space. Given a polynomial

$$p(x) = c_0 + \dots_{c\,k} x^k \in \mathbb{F}[x]$$

we can define the operator $p(T) = c_0 I + c_1 T + \cdots + c_k T^k$.

Definition 7.3.1 (Annihilator)**.** Let V be an $\mathbb{F}$ vector space, $T \in \mathcal{L}(V)$. A polynomial p annihilates T if $p(T) = 0$.

Similarly, a polynomial p annihilates a square matrix A if $p(A) = 0$. The annihilator of T, denoted $M(T)$, is the set of all polynomials in $\mathbb{F}[x]$ which annihilate T.

Proposition 7.3.1.1 (The set of annihilators is an ideal)**.** Let $T \in \mathcal{L}(V)$. Consider the set $Ann(T) = M(T) := \{f \in \mathbb{F}[x] \text{ s.t. } f(T) = 0\}$.

1. It is a principal ideal of $\mathbb{F}[x]$.

2. If V is a finite-dimensional vector space, then this ideal contains more than just the 0 map (i.e., it is nontrivial).

Proof. We show that M is an ideal of $\mathbb{F}[x]$. Consider $Z(x) = 0, \forall x \in V$. Notice $Z = f(T)$ s.t. $f(x) = 0$. Thus, $f \in M$. Let $f, g \in M$. Then, $f(T) = Z$ and $g(T) = Z$. Thus, $(f + g)(T) = f(T) + g(T) = Z + Z = Z$. Thus, $f + g \in M$. Let $f \in \mathbb{F}[x]\, g \in M$. Then, $g(T) = Z$. Thus, $(fg)(T) = f(T) \cdot g(T) = f(T) \cdot Z = Z$. Thus, $fg \in M$. Reacall that $\dim(\mathcal{L}(V, V)) = \dim V \times \dim V$. Thus $\left\{I, T, T^2, \dots, T^{n^2}\right\}$ must be L.D., that is $\exists$ an scalar such that $a_{n^2} T^{n^2} + a_{n^2-1} T^{n^2-1} + \cdots + a_0 I = Z$. Let $f(x) = a_{n^2} x^{n^2} + \cdots + a_1 x^1 + a_0$, then $f(T) = Z$. □

Proposition 7.3.1.2. Let $A \in \mathbb{F}^{n \times n}$. Consider the set $M = \{f \in \mathbb{F}[x] \text{ s.t. } f(A) = 0\}$,

1. M is a principal ideal of $\mathbb{F}[x]$

2. If V is a finite-dimensional vector space, then this ideal is always non-zero.

Definition 7.3.2 (Minimal Polynomial)**.** The *minimal polynomial* for T is the *unique monic* generator for the ideal

$$M(T) = \{f \in \mathbb{F}[x] \text{ s.t. } f(T) = 0\}$$

The *minimal polynomial* for a matrix $A \in \mathbb{F}^{n\times n}$ is the *unique monic* generator for the ideal

$$M(A) = \{f \in \mathbb{F}[x] \text{ s.t. } f(A) = 0\}$$

Remarks 7.3.2.0.1. The minimal polynomial can be used to determine whether a matrix or linear transformation is diagonalizable or triangularizable. This is the subject of the next section. For now, we will focus on basic properties of the minimal polynomial and how to compute it.

Theorem 7.3.2.1 (Alternate characterization of minimal polynomials)**.** Let V be an $\mathbb{F}$ vector space, $T \in \mathcal{L}(V), (A \in \mathbb{F}^{n\times n}), p \in \mathbb{F}[x]$. The following are equivalent

1. p is the minimal polynomial for $M(T)\,(M(A))$
2. p is a polynomial such that
   (a) p is monic
   (b) $p(T) = 0$
   (c) $f(T) = 0 \implies \deg(f) \geq \deg(p)$.

Ex. Show the characteristic polynomial of $A = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}$ annihilates A.

Solution. $f(x) = \det(xI - A)$ then $f(A) = Z$.
$f(x) = \det\left(\begin{bmatrix} x-1 & -2 \\ -3 & x \end{bmatrix}\right) = x^2 - x - 6$. Thus,

$$f(A) = A^2 - A - 6I = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

■

Theorem 7.3.2.2 (Cayler-Hamilton Theorem)**.** Let $T \in \mathcal{L}(V), \dim(V) < \infty$ and $f(X) = \det(xI - T)$ be the characteristic polynomial for T. Then,

$$f(T) = Z \text{ and } f \in M(T).$$

Similarly, if A is an $n \times n$ matrix with characteristic polynomial $f(x)$,then $f(A) = 0$.

Proof. It suffices to prove the matrix version as $[f(T)]_\beta = f([T]\beta)$. Let $f(x) = \det(xI - A)$. We know from the det section that A adj $(A) = \det(A) \cdot I \implies (xI - A)$ adj $(xI - A) = \det(xI - A) I = f(A)I \implies 0 = f(A) \implies f \in M(A)$. □

Corollary 7.3.2.2.1. Let p be the *monic polynomial* for T and f be the *characteristic polynomial.* Then, the minimal polynomial p divides the characteristic polynomial f.

Theorem 7.3.2.3. Let $T \in \mathcal{L}(V)$ and c be a characteristic value. Show that $(x - c)$ divides te minimal; polynomal.

Corollary 7.3.2.3.1. The minimal polynomial has the same roots as the characteristic polynomial

Theorem 7.3.2.4. The minimal polynomial has the same prime factors as the characteristic polynomial.

Ex. Suppose the characteristic polynomial of $A : \mathbb{C}^{10} \to \mathbb{C}^{10}$ is $f(x) = (x - 1)^8(x + 1)(x - 2)$.

Solution. $p = (x + 1), (x - 2, \ldots, (x - 1)^8(x + 1)(x - 2)$. We know

$$p = (x - 1)^r (x + 1)(x - 1) \text{ s.t. } r \in \mathbb{N} \cap [1, 8].$$

■

Ex. Let

$$A = \begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix}.$$

Compute the minimal polynomial of A. Note that $f(x) = (x + 1)^2(x - 3)$. We check

$$f(x) = (x + 1)(x - 3) = 0 \text{ or } p(x) = (x + 1)^2(x - 3).$$

7.4 Invariant Subspaces

Definition 7.4.1 (Invariant Subspaces)**.** Let $T \in \mathcal{V}$ and let W be a *subspace* of V. Then W is *invariant* under T if:

$$\forall \beta \in W, T(\beta) \in W, \text{ i.e., } T(W) \subseteq W.$$

Ex. Let $\beta = \{e_1, e_2\}\,; [T]_\beta = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Show that Span $(\{e_1\})$ is invariant under T.

Solution.

$$[T(e_1)]_\beta = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \implies T(ce_1) = c \cdot e_1 = ce_1.$$

If $v \in \text{Span}\{e_1\}$, then $v = ce_1$. Note that $T(ce_1) = ce_1 \in \text{Span}\{e_1\}$. Thus, $\text{Span}\{e_1\}$ is invariant. ■

Exercise. Let $W_1 = V$ and $W_2 = \{0\}$. Show that W_1 and W_2 are both invariant under T. **Ex.** A vector space without nontrivial invariant subspace. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ represented in the standard basis by $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Show the only invariant subspace is the trivial invariant subspace.

Proof. Note that

$$\det(xI - A) = \det\left\{\begin{bmatrix} x & 1 \\ -1 & x \end{bmatrix}\right\} = x^2 + 1.$$

Thus, as $\mathbb{R}$ is the field, $x^2 + 1$ does not factor, that is, there are *no* eigenvalue. Suppose for the sake of contradiction that there is an invariant subspace 1 dimensional $W = \text{Span}\{\alpha\}$ s.t. $\alpha \neq 0$. By definition of an invariant subspace, $T\alpha \in W \implies \exists c$ s.t. $T\alpha = c\alpha$, i.e., T has a characteristic value contradicting our observation that it has *no eigenvalue.* □

Remarks 7.4.1.0.1. This exis theoretically significant. It tells us that we cannot achieve "nice" matrix representations for vector spaces over an arbitrary field $\mathbb{F}$ because nontrivial invariant subspaces may not exist. (Recall the comment that "nice" matrix representations rely on the existence of nontrivial invariant subspaces.) Thus, we must choose a $\mathbb{F}$ better for this cause. As we shall see, $\mathbb{C}$ is the perfect choice because of the Fundamental Theorem of Algebra.

Ex. Consider the infinite-dimensional vector space $\mathbb{F}[x]$.
Let $D : \mathbb{F}[x] \to \mathbb{F}[x]$ be the differentiation operator.
Let W be the subspace of polynomials of degree less than or equal to n.

$$W = \left\{ f(x) = c_0 + c_1 x + \ldots + c_k x^k \,\middle|\, c_j \in \mathbb{F},\, k \leq n \right\}.$$

Then, W is invariant under D.

Theorem 7.4.1.1 (Commuting Operators Provide Invariant Subspaces)**.** Let $T, U \in \mathcal{L}(V)$ s.t. $TU = UT$, i.e., U and T are *commutative.* Let $W = \text{range}(U)$ and $N = \ker(U)$. Then, W and N are *invariant under T.*

Proof. Let $v \in W$. Then, $v = Ux$ s.t. $x \in W$. Consider $T(v)$. Note that

$$\begin{aligned} T(v) = T(Ux) &= TU(x) \\ &= UT(x) \\ &= U(T(x)) \in \text{range}(U) = W. \end{aligned}$$

Thus, $T(V) \in W$. Let $\alpha \in \ker(U)$. Then, $U\alpha = 0 \implies T(U\alpha) = 0 \implies u(T\alpha) = 0 \implies T\alpha \in \ker(U) = N$. □

Definition 7.4.2 (Restriction Operator)**.** Let W be *invariant* under the *operator T.* The *restriction operator $T_W : W \to W$* induced by T is the linear operator T_W defined by $T_W(w) = T(w)$, for all $w \in W$. Note that since the domain of T_W is W and the domain of T is V, it may be the case that $T_W \neq T$.

Theorem 7.4.2.1 (Invariant Subspaces give block diagonal matrix representations)**.** Let VL be a finite dimensional vector space. et W be an *invariant subspace* under the *linear operator T.* Then there exists a basis such that the matrix representation of T with respect to this basis has the following form:

$$[T]_\beta = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} \quad \text{where } \beta = \beta_1 \cup \beta_2.$$

(The 0 is very nice.)
(We shall explain what β_1, β_2 mean in a moment.)

Proof. Let $\beta_1 = \{\alpha_1, \ldots, \alpha_j\}$ be a basis for W. Then,

$$T\alpha_i = a_{1,i}v_1, \ldots, a_{j,i}v_j, \forall i.$$

Let $\beta_2 = \{\alpha_{j+1}, \ldots, \alpha_n\}$ be the set of vectors used to extend β_1 to create a basis for V. Then,

$$T\alpha_i = a_{1,i}v_1 + \cdots + a_{j,i}v_j + 0(v_{j+1} + \cdots + v_n), \forall i \in \{1, \ldots, j\}.$$

That is,

$$[T]_{\beta_1, \beta_2} = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}.$$

□

Corollary 7.4.2.1.1 (Characteristic and Minimal Polynomial of a Restriction Operators Divides those of the parent operator)**.** Let W be an invariant subspace for T. Then the characteristic polynomial (resp. minimal polynomial) for the restriction operator T_W divides the characteristic polynomial (resp. minimal polynomial) for T.

Proof. Here we prove only the characteristic polynomial result. As W is an invariant subspace, exists $\beta = \{\beta_1, \beta_2\}$ s.t.

$$[T]_{\beta_1,\beta_2} = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

Now show that

$$\det\left(xI - [T]_\beta\right) = \det\left(xI - B\right) \cdot \det\left(xI - D\right).$$

As $\det\left(xI - B\right)$ is the characteristic polynomial of T_W, it divides $\det\left(xI - [T]_\beta\right)$ which is the characteristic polynomial of T. □

Theorem 7.4.2.2. Let $T \in \mathcal{L}(V)$ and let $c_1, \cdots, c_n$ be the *distinct* characteristic values of T. Then there exists a basis β such that:

$$[T]_\beta = \left[\begin{array}{c|c} \operatorname{diag}(c_1 I_{d_1}, \cdots, c_k I_{d_k}) & C \\ \hline 0 & D \end{array}\right] \text{ where } d_i = \dim\left(\ker\left(T - cI\right)\right).$$

Proof. See lecture notes. □

Corollary 7.4.2.2.1 (Sum Characterization of Diagonalizable)**.** Let $T \in \mathcal{L}(V)$, $c_1, \ldots, c_k$ be the characteristic values and $W_i = \mathit{nullspace}(T - c_i I)$. The following are equivalent:

1. T is *diagonalizable*
2. $W_1 + \cdots + W_k = V$
3. $W_1 \oplus \cdots \oplus W_k = V$

Definition 7.4.3 (Conductor/Stuffer)**.** Let V be an $\mathbb{F}$ vector space and $T \in \mathcal{L}(V)$. Let W be an invariant subspace for T, α be a vector in V. The *T-conductor* of α into W is the set:

$$S_T(\alpha; W) := \{g \in \mathbb{F}[x] : g(T)\alpha \in W\}$$

(i.e., the set of all polynomials $g \in \mathbb{F}[x]$ such that $g(T)\alpha$ is in W)
In the case that $W = 0$, we call $S_T(\alpha; \{0\})$ the *T-Annihilator* of α.

Lemma 7.4.3.1 (Conductors Are Ideals)**.** Let W be an invariant subspace for T, then $S_T(\alpha; W)$ is a principal ideal in $\mathbb{F}[x]$.

Definition 7.4.4 (T-Conductor)**.** Since $S_T(\alpha; W)$ is a *principal idea,* it has a *monic* generator g. We call g the *T-conductor of* α onto W.

Remarks 7.4.4.0.1 (Computing the T-conductor using the minimal polynomial)**.** Note that as the *minimal polynomials* maps $\alpha \to 0 \in W$, the minimal polynomial of T is in $S_T(\alpha; W)$. Since the T-conductor generates the conductor, it follows that it divided the *minimal polynomial* of T. Thus, the T-conductor is made of a product of the factors of the minimal polynomial. This fact is useful in computing the T-conductor.

7.5 Triangularizability/Diagonalizability by Minimal Polynomial

Lemma 7.5.0.1. Let V be a finite-dimensional vector space over $\mathbb{F}$. Let $T \in \mathcal{L}(V)$ such that the minimal polynomial for T is a product of linear factors:

$$p(x) = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}.$$

Let W be a proper subspace of V invariant under T. Then there exists vector $\alpha \in V$ such that

1. $\alpha \notin W$
2. $(T - c_i I)\alpha \in W$. (α gets stuffed by $T - c_i I$.)

Proof. Let $\beta \notin W$. Let g be the T-Conductor of $S_T(\beta; W)$. As the g divides the minimal polynomial of T, there exists $r_1, \cdots, r_k$, where $r_i \in \{0, ..., d_i\}$ for all i, such that

$$g(x) = (x - c_1)^{r_1} \dots (x - c_k)^{r_k} \text{ s.t. at least one } c_i \neq 0.$$

Redefine

$$g = \prod_{i=1}^{n} (x - \lambda_i), \quad \text{where } n = \sum_{i=1}^{k} r_i$$

. Then consider the sequence $\{g_1, g_2, ..., g_n\}$, where

$$g_j = \prod_{i=1}^{j} (x - \lambda_i).$$

As $\beta \notin W$ and $g_n(T)\beta \in W$, there must exist an $m \in \{1, ..., n\}$ such that

$$g_{m-1}(T)\beta \notin W \text{ and } (T - \lambda_m I)\, g_{m-1}(T)\beta = g_m(T)\beta \in W.$$

Letting $\alpha = g_{m-1}(T)\beta$ gives our result as $\alpha \notin W$, but $(T - \lambda_m I)\alpha \in W$. □

Definition 7.5.1 (Upper Triangulable)**.** $T \in \mathcal{L}(V)$ is *triangulable* if there exists an *orghonomral basis* β for V s.t. $[T]_\beta$ is *upper or lower triangular.*

Remarks 7.5.1.0.1. *Triangular metrices* are connected with *invariant subspaces*:

Suppose $\beta = \{\alpha_1, \dots, \alpha_n\}$ and

$$[T]_\beta = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \vdots \end{bmatrix}$$

Now define W_i in the following way.

$$\begin{aligned} W_1 &= \operatorname{span}(\alpha_1) \\ W_2 &= \operatorname{span}(\alpha_1, \alpha_2) \\ &\vdots \\ W_{n-1} &= \operatorname{span}(\alpha_1, \dots, \alpha_{n-1}) \\ W_n &= \operatorname{span}(\alpha_1, \dots, \alpha_n) \end{aligned}$$

The set $W_1, \dots, W_n$ have a few very nice properties. (Here, we use *subset* to denote a proper subset.)

1. (Well-ordered) $W_1 \subset W_2 \subset \dots \subset W_n$
2. (T-invariant) $T(W_j) \subseteq W_i$ for all $1 \leq j \leq i \leq n$

If we show that a *basis* such that $W_1 \subset W_2 \subset \dots \subset W_n$ and W_i is T-invariant for all i, then T is triangulable.

To show that a *subspace* is *invariant*, we simply need to check a *basis* of the subspace is *invariant*.

Theorem 7.5.1.1 (Invariant Subspace Characterization of Triangularbility)**.** Let V be a $\mathbb{F}$ vector space, $T \in \mathcal{L}(V)$. The following are equivalent:

1. T is upper triangularizable.
2. There exists a basis $\{\alpha_1, \dots, \alpha_n\}$ for V such that if $W_i = \operatorname{span}(\alpha_1, \dots, \alpha_i)$, then:
   (a) (Well-ordered) $W_1 \subset W_2 \subset \dots \subset W_n$
   (b) (T-invariant) $T(W_j) \subseteq W_i$ for all $1 \leq j \leq i \leq n$

Theorem 7.5.1.2 (Characterizations of Triangularizability and Diagonalizability in terms of the minimal polynomial: IMPORTANT)**.** Let $T \in \mathcal{L}(V)$ and $\dim(V) = n < \infty$. Then

1. T is *triangularizable* if and only if the prime factorization of the *minimal polynomial* for T is a *product* of *linear factors*.
2. T is *diagonalizable* if and only if the prime factorization of the *minimal polynomial* for T is a *product* of *distinct linear factors*.

Proof. (1.) $\Rightarrow$

Suppose $[T]_\beta$ is *triangularizable*, then the characteristic polynomial for T is

$$f(x) = \det\left(xI - [T]_\beta\right) = (x - ([T]_\beta)_{11}) \dots (x - ([T]_\beta)_{nn})) = (x - c_1)^{d_k} \dots (x - c_k)^{d_k}$$

where c_i are *distinct* characteristic values of T. Therefore, the *minimal polynomial* for T has a similar form since it divides $F(x)$, by the Cayley-Hamilton Theorem:

$$p(x) = (x - c_1)^{r_1} \dots (x - c_k)^r_k \text{ s.t. } r_i \in \{0, \dots, d_i\} \text{ and } \exists r_i \neq 0.$$

(2.) $\Rightarrow$

If T is diagonal, then $p(x) = (x - c_1) \cdots (x - c_k)$ is the *minimal polynomial* for T.

$$p\left([T]_\beta\right)_{ij} = \begin{cases} p\left([T]_\beta\right)_{ii}, i = j \\ 0, i \neq j \end{cases} \implies p\left(\left([T]_\beta\right)_{ii}\right) = 0 \forall i$$
$$\implies p\left([T]_\beta\right) = 0.$$

(1.) $\Leftarrow$

Suppose the minimal polynomial is a product of linear factors. Apply the previous lemma to $W = \{0\}$ to get α_1. Then

$$\begin{aligned}(T - c_iI)_{\alpha_i} \in w &\implies (T - c_iI)_{\alpha_1} = 0 \\ &\implies T\alpha_1 = c_i\alpha_i \in \text{Span}\left(\{\alpha_1\}\right).\end{aligned}$$

Note that W_1 is T-invariant. Now apply the previous lemma to W_1 to get α_2. Then

$$(T - c_jI)_{\alpha_2} \in W_1 \implies (T - c_jI)_{\alpha_2} = k\alpha_1 \implies T\alpha_2 = c_j\alpha_2 + k\alpha_1 \in \text{Span}\left(\{\alpha_1, \alpha_2\}\right).$$

Let $W_2 = \{\alpha_1, \alpha_2\}$ Note that W_2 is T-invariant. Continue this way until you have $W_n = V$. By construction $W_i \subset W_{i+1}, \forall i \in \{1, \dots, n-1\}$ and $T\left(w_j\right) \subseteq w_i \forall 1 \leq j, i \leq n$. Thus, T is triangularizable.

(2.) $\Leftarrow$ diagonalizability (Not diagonalizable $\Rightarrow$ not distinct factors)

Suppose T is not diagonal. Suppose W is the subspace spanned by all the characteristic vectors of T. If $W = V$, then T is diagonalizable as we proved earlier, which contradicts our assumption.

Instead, suppose $W \neq V$.

By the previous lemma, there exists a vector $\alpha \notin W$ and a characteristic value c_i of T such that the vector

$$\beta = (T - c_iI)\alpha \in W.$$

Let $h(T)$ be some polynomial. We have shown that $h(T)\beta \in W$. Now $p = (x - c_i)q$, for some polynomial q. We show $q\left(c_i\right) = 0$. Pick h s.t. $q\left(T\right) - q\left(c_i\right) = \left(T - c_i\right) h\left(T\right)$, i.e.,$q - q\left(c_i\right) = \left(x - c_i\right) h$. As $p\left(T\right) \alpha = 0$ and $p\left(T\right) \alpha = \left(T - c_i\right) \left(q\left(T\right) \alpha\right)$,

$$q\left(T\right) \alpha \text{ is an eigenvector,}$$

thus $q\left(T\right) \alpha \in W$. As $q\left(T\right) \alpha = q\left(c_i\right) \alpha = h(T)\beta = h(T)\beta - q\left(T\right) \alpha \in W$. But $\alpha \notin W$. Thus, $q\left(c_i\right) = 0$. Thus, $\left(x - c_i\right) | q\left(x\right)$, i.e., $p\left(x\right)$ has the linear factor $\left(x - c_i\right)$ at least twice. □

Corollary 7.5.1.2.1. Every Operator over $\mathbb{C}$ vector space is triangularizable.

Corollary 7.5.1.2.2. Sps $T \in \mathcal{L}\left(V\right)$ and f is the characteristic polynomial of T. Then,

1. if f is a product of linear factors T is triangularizable.

2. if f is a product of distinct linear factors, then T is diagonalizable.

Ex. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be represented by $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Determine if T is triangularizable; if so, is it diagonalizable.

Solution. $\det\left(xI - A\right) = \det(\begin{bmatrix} x & 1 \\ -1 & x \end{bmatrix}) = x^2 + 1$. As the discriminant is negative, $x^2 + 1$ has no real roots. Thus, $x^2 + 1$ is prime over $\mathbb{R}$. Therefore the characreristic polynomial is *not* a porduct of linear factors. Thus, T is neither triangularizable nor diagonalizable. ■

Ex. What if $T : \mathbb{C}^2 \to \mathbb{C}^2$.

Solution. $\det\left(xI - A\right) = \det(\begin{bmatrix} x & 1 \\ -1 & x \end{bmatrix}) = x^2 + 1 = (x - i)(x + i)$. As the characteristic polynomial is a product of distinct linear factors, T is both triangularizable and diagonalizable. ■

Ex. Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 2 & -2 & 2 \\ 2 & -3 & 2 \end{bmatrix}$. Is A similar to a triangular matrix over $\mathbb{R}$?

Solution.

$$\det\left(\begin{bmatrix} x & 1 & 0 \\ 2 & x+2 & 2 \\ 2 & -3 & x-2 \end{bmatrix}\right) = x\left[\left(x + 2\right)\left(x - 2\right) + 6\right] + \left[-2(x - 2) - 4\right] = x^3.$$

Thus, it is triangularizable but *not* diagonalizable. ■

Remarks 7.5.1.2.1. We have shown that every operator over a $\mathbb{C}$ vector space admits a triangular matrix representation. Turns out another matrix representation is nice in the following sense.

Definition 7.5.2 (Block Diagonal Matrices)**.** A matrix A is a *block diagonal matrix* if

$$A = \operatorname{diag}(A_1, \dots, A_k) = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_k \end{bmatrix} \text{ s.t. } A_i \text{ is a square matrix.}$$

Note that this study will eventually lead to an improvement: the *Jordan Canonical Form*:

$$\begin{bmatrix} A_1^n & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_k^n \end{bmatrix}$$

7.6 Direct-Sum Decomposition and Projection

Definition 7.6.1 (Independent Subsets)**.** Let $W_1, \dots, W_k$ be subspaces of V. The subspaces $W_1, \dots, W_k$ are said to be *independent* if $\sum_{i=1}^k \alpha_i = 0$ s.t. $\alpha_i \in W_i \implies \alpha_i = 0, \forall i$.

Remarks 7.6.1.0.1 (Equivalence)**.** Note that it is equivalent to say

1. $\forall j \in \{2, \dots, k\}, W_j \cap \{W_1 + \cdots + W_{j-1}\} = 0$
2. If $\beta_1, \dots, \beta_k$ are ordered basis for $W_1, \dots, W_k$, then the concatenation, $\beta = \{\beta_1, \dots, \beta_k\}$, is a basis for $W_1 \oplus \cdots \oplus W_k$. The final sum is called the *direct sum* and $W = W_1 \oplus \cdots \oplus W_k$.

Ex. Let $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}^3$ be linearly independent. Let $W_1 = \operatorname{Span}(\{\alpha_1, \dots, \alpha_2\})$. Let $\alpha_3 \in V$ be a vector not in W_1. Let $W_2 = \operatorname{Span}(\{\alpha_3\})$. Then,

$$\mathbb{R}^3 = W_1 \oplus W_2.$$

Ex. If $\{\alpha_1, \dots, \alpha_n\}$ is any basis for V with $W_1 = \operatorname{Span}\{\alpha_1\}, W_2 = \operatorname{Span}\{\alpha_2\}, \dots, W_n = \operatorname{Span}\{\alpha_n\}$. Show that $V = W_1 \oplus \cdots \oplus W_n$.

Theorem 7.6.1.1. Let $T \in \mathcal{L}(V)$ with $c_1, \dots, c_n$ be the *distinct characteristic values for* T and $W_i = \ker(T - c_i I), j = 1, \dots, k$. Then, the set $\{W_1, \dots, W_n\}$ is linearly independent. T is *diagonalizable* $\iff V = W_1 \oplus \cdots \oplus W_n = W_1 + \cdots + W_n$.

Definition 7.6.2 (Generalized Projection)**.** A *projection on* V is an operator $E \in \mathcal{L}(V)$ s.t. $E^2 = E$. (A matrix A is a projection matrix if $A^2 = A$.)

Ex. Prove any projection matrix A has eigenvalues 0 and 1.

Theorem 7.6.2.1 (Properties of Projections)**.** There are four well-known properties of projections. Let $R = \operatorname{range}(E)$ and $N = \ker(E)$.

1. $\beta \in R \iff E(\beta) = \beta$.
2. $\forall \alpha \in V, E\alpha \in R$ and $(I - E)(\alpha) \in N$.
3. $V = R \oplus N (\forall \alpha \in V, \alpha = E\alpha + (I - E)\alpha)$.
4. E is diagonalizable. $\exists$a basis β s.t. $[E]_\beta = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$.

E is called the projection onto R *along* N.

Ex. Find a projection E that porjects $\mathbb{R}^2$ onto $R = \operatorname{Span}\{(1,2)\}$ along $N = \operatorname{Span}\{3, -1\}$. We can define E by $E(x_1, x_2) = E(a_1\alpha_1 + b\alpha_2)$ for $\beta = \{\alpha_1, \alpha_2\}$ which explicitly gives

$$E(x_1, x_2) = a(1,2) = \frac{x_1 + 3x_2}{7} = \left(\frac{x_1 + 3x_2}{7}, \frac{2x_2 + 6x_2}{7}\right).$$

As $E\alpha_1 = \alpha 1$ and $E\alpha_2 = 0 \implies [E]_{\{\alpha_1,\alpha_2\}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. Now we show that E is a projection. $E^2 = E$. As $[E^2]_\beta = [E]_\beta, E^2 = E$. Thus, E is a projection,

$$[E]_{\{e_1,e_2\}} = \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix}^{-1}.$$

Note that the column on the left of P is the range and on the right is the null.

Theorem 7.6.2.2. Sps $V = \oplus_{i=1}^{k} W_i$, then $\exists E_1, \ldots, E_k \in \mathcal{L}(V)$ s.t.

1. $E_j^2 = E_j, \forall j$
2. $\forall i \neq j, E_i E_j = 0$
3. $I = \sum_{i=1}^{k} E_i$
4. $\text{range}(E_i) = W_i, \forall i.$

Proof. Let $\beta_i = \{\alpha_{i,1}, \ldots, \alpha_{i,n}\}$ be a basis for W_i. Then, $\beta = \{\beta_1, \ldots, \beta_n\}$ is a basis for V. Using the basis for β, we can define $E_j \in \mathcal{L}(V)$ for $j = 1, \ldots, d$,

$$E_j(\alpha_{ik}) = \begin{cases} \alpha_{ik}, i = j \\ 0, i \neq j \end{cases}$$

Given this definition for E_i show that all conditions are satisfied.

$$E_j^2 = \begin{cases} E_j(\alpha_i k), i = j \\ E_j(0), i \neq j \end{cases} = \begin{cases} \alpha_{ik}, i = j \\ 0, i \neq j \end{cases}$$

Then,

$$i \neq j \implies E_i E_j(\alpha_{lk}) = \begin{cases} E_i(\alpha_{lk}), l = j \\ E_i(0), l \neq j \end{cases} = \begin{cases} 0, l = j \\ 0, l \neq j \end{cases} \implies (E_i \neq E_j \implies E_i E_j = 0).$$

Thus,

$$(E_i + E_j)(E_i + E_j) = E_i^2 + E_j^2 = E_i + E_j, \forall i \neq j.$$

□

7.7 Invariant Direct Sums

Definition 7.7.1 (Direct Sum of Linear Transformations)**.** Suppose $T \in \mathcal{L}(V)$ and $V = \oplus_{j=1}^{k} W_j$ s.t. W_j is *invariant for* T. Let T_j be the representation of $T \to W_j$. As V is a direct sum, every vector $\alpha \in V$ is represented uniquely by vectors in $W_1, \ldots, W_k$, i.e., $\forall \alpha = \alpha_1 + \cdots + \alpha_k \in V$. Applying T to α gives $T(\alpha) = T_1\alpha_1 + \cdots + T_k\alpha_k$. Then we say T is the *direct sum* of $T_j \in \mathcal{L}(W_j)$ and we write $T = \oplus_{j=1}^{k} T_j$.

Theorem 7.7.1.1 (Block diagonal form from an invariant direct sum)**.** Let $V = \bigoplus_{j=1}^{k} W_j$ with each W_j invariant under $T \in \mathcal{L}(V)$. For each j, let $T_j := T|_{W_j} \in \mathcal{L}(W_j)$. Then there exists a basis β of V such that

$$[T]_\beta = \begin{bmatrix} [T_1]_{\beta_1} & 0 & \cdots & 0 \\ 0 & [T_2]_{\beta_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [T_k]_{\beta_k} \end{bmatrix}$$

where β_j is a basis of W_j and $\beta = \beta_1 \cup \cdots \cup \beta_k$.

Proof. Pick a basis β_j for each W_j and set $\beta = \beta_1 \cup \cdots \cup \beta_k$. Because V is a direct sum, every $v \in V$ is uniquely $v = \sum_{j=1}^{k} v_j$ with $v_j \in W_j$. Since $T(W_j) \subseteq W_j$, applying T to a basis vector from W_j yields a linear combination of vectors *only* in W_j. Hence, in the basis β, the columns corresponding to β_j have nonzero entries only in the rows corresponding to β_j, and there are no cross terms between different W_i and W_j. Therefore $[T]_\beta$ is block diagonal with diagonal blocks $[T_j]_{\beta_j}$. □

Theorem 7.7.1.2 (Invariance $\iff$ commuting with projections)**.** Let $V = \bigoplus_{j=1}^{k} W_j$ and let $E_j \in \mathcal{L}(V)$ be the projection onto W_j (so $E_j^2 = E_j$, $\text{range}(E_j) = W_j$, and $E_iE_j = 0$ for $i \neq j$, $\sum_i E_i = I$). Then

$$W_j \text{ is invariant under } T \quad \iff \quad TE_j = E_jT.$$

Proof. ($\Rightarrow$) Assume $T(W_j) \subseteq W_j$. For any $v \in V$, $E_jv \in W_j$ and hence $T(E_jv) \in W_j$. As E_j acts as the identity on W_j, $E_jTE_jv = TE_jv$. Writing $v = E_jv + (I - E_j)v$ and using $E_j(I - E_j) = 0$ gives $E_jTv = TE_jv$, i.e. $E_jT = TE_j$.

($\Leftarrow$) Assume $TE_j = E_jT$. For $w \in W_j$ we have $E_jw = w$, hence $Tw = TE_jw = E_jTw \in \text{range}(E_j) = W_j$. Thus W_j is invariant. □

Theorem 7.7.1.3. Let $T \in \mathcal{L}(V), \dim(V) < \infty$. Sps T is diagonalizable and $c_1, \ldots, c_k$ are the distinct characteristic values of T. Then, $\exists E_1, \ldots, E_k \in \mathcal{L}(V)$ s.t.

1. $T = c_1 E_1 + \cdots + c_k E_k$
2. $I = E_1 + \cdots + E_k$
3. $E_i E_j = 0, \forall i \neq j$
4. $E_i^2 = E_i, \forall i$
5. range$(E_i) = \ker(T - ciI)$, which is the characteristic space of c_i for T.

Proof. Sps T is diagonalizable with distinct characteristic values $c_1, \ldots, c_k$. Let W_i be the space of characteristic vectors associated with the characteristic values c_i. As we have seen $V = W_1 + \cdots + W_k$. Let $E_1, \ldots, E_k$ be the projection onto W_i along $W_1 + \cdots + W_{i-1} + W_{i+1} + W_k$. We have shown that $\alpha = I\alpha = E_1\alpha + \ldots E_k\alpha$ and so $T\alpha = TE_1\alpha + \cdots + TE_k\alpha$, i.e., $T = c_1 E_1 + \cdots + c_k E_k$. □

Theorem 7.7.1.4 (Lagrange Polynomial to Compute E_i)**.** If T is diagonalizable with characteristic values $c_1, \ldots, c_k$, we can compute the projection E_i using *Lagrange Polynomials:* for $1 \leq j \leq k$,

$$p_j(x) := \prod_{i \neq j} \frac{x - c_i}{c_j - c_i} = \frac{x - c_1}{c_j - c_1} \cdots \frac{x - c_{j-1}}{c_j - c_{j-1}} \frac{x - c_{j+1}}{c_j - c_{j+1}} \cdots \frac{x - c_k}{c_j - c_k}.$$

Then, $p_j(T) = E_j$.

Exercise. Show that $p_j(T) = E_j$.

Ex. Let

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

Solve for $E_1, \ldots, E_k$ such that

1. $A = \sum_{i=1}^{k} c_i E_i$
2. $I = \sum_{i=1}^{k} E_i, \quad E_i E_j = 0$ for $i \neq j$
3. $E_1 E_2 = 0$
4. $E_1^2 = E_1 \quad$ and $\quad E_2^2 = E_2$
5. range$(E_i) =$ nullspace$(A - c_i I)$

Ex. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be represented in the standard basis by

$$A = \begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix}.$$

Determine if A is diagonalizable. If so, compute $E_1, \ldots, E_k$ such that:

1. $A = \sum_{i=1}^{k} c_i E_i$
2. $I = \sum_{i=1}^{k} E_i, \quad E_i E_j = 0$ for $i \neq j$
3. $E_1 E_2 = 0$
4. $E_1^2 = E_1 \quad$ and $\quad E_2^2 = E_2$
5. range$(E_i) =$ nullspace$(A - c_i I)$

Remarks 7.7.1.4.1 (Non-Diagonalizable and Non-Triangularizable Matrices)**.** We started the discussion about direct sum decompositions to create block diagonal matrices. We have shown that T is a direct sum of linear transformations if and only if T is block diagonalizable. In order to show that T is a direct sum of linear transformations we need to decompose V into a direct sum of invariant subspaces. Can we always do this?

It turns out the answer is yes! In the upcoming sections, we shall explore two ways to decompose V into the direct sum of invariant subspaces. The first technique, called the primary decomposition theorem, uses the prime factors of the characteristic polynomial. The second technique, called the cyclic decomposition theorem, uses the minimal polynomial.

While we will leave the proofs for both types of decompositions, the proofs are quite fascinating, and I encourage anyone interested in them to explore them further.

8 Decomposition and Jordan

8.1 The Primary Decomposition and Cyclic Decomposition

Definition 8.1.1 (T-annihilators of a vector)**.** Let $\alpha \in V$. Then, $M\left(\alpha; T\right) = \left\{f \in P\left(\mathbb{F}\right) \text{ s.t. } f(T)\alpha = 0\right\}$ is the *T-annihilator of* α.

Proposition 8.1.1.1. $M\left(\alpha; T\right)$ is an ideal.

Definition 8.1.2 (T-annihilator)**.** The monic generator of $M\left(\alpha; T\right)$ is called the *T-annihilator of* α.

Proposition 8.1.2.1 (A vector exists with T annihilator equal to the minimal polynomial)**.** Let V be a finite-dimensional vector space, $T \in \mathcal{L}(V)$ and $p(x)$ be the minimal polynomial of T. Let $\deg(p) = k$, $p(x) = p_1(x)^{r_1} \cdots p_k(x)^{r_k}$ be the prime factorization of p. Then for each i there exists a vector $\alpha \in V$ such that the T-annihilator of α is $p_i(T)^{r_i}$.

Theorem 8.1.2.2 (Primary Decomposition Theorem)**.** Let $T \in \mathcal{L}(V)$, $\dim(V) < \infty$. Let $p(x)$ be the minimal polynomial for T with prime factorization

$$p(x) = p_1(x)^{r_1} \cdots p_k(x)^{r_k},$$

(i.e., the p_j are distinct irreducible monic polynomials and $r_j \geq 1$).

Let $W_j = \text{nullspace}(p_j(T)^{r_j})$ for $j = 1, \ldots, k$. Then:

1. $V = W_1 \oplus \cdots \oplus W_k$
2. W_i is T invariant, $\forall i$
3. If T_{w_i} is the restriction of T to W_i, then the minimal polynomial of T_{w_i} is $p_i^{r_i}$.

Proof. Write $m_T = \prod_{i=1}^k p_i^{r_i}$ and, for each i, set

$$q_i(x) := \frac{m_T(x)}{p_i(x)^{r_i}} = \prod_{j \neq i} p_j(x)^{r_j}.$$

Note that $\gcd\bigl(q_i, p_i^{r_i}\bigr) = 1$ for every i.

Step 1: Construct commuting projections. Because the polynomials $q_1, \dots, q_k$ are pairwise coprime, there exist polynomials $c_1, \dots, c_k \in \mathbb{F}[x]$ such that

$$\sum_{i=1}^k c_i(x)\, q_i(x) = 1. \tag{1}$$

(Proof by induction using Bézout's identity.) Define

$$E_i := c_i(T)\, q_i(T) \in \mathcal{L}(V).$$

From (1) we get $\sum_{i=1}^k E_i = I_V$. Moreover, for $i \neq j$ we have $q_i(x)q_j(x)$ divisible by $m_T(x)$, hence

$$E_iE_j \;=\; c_i(T)c_j(T)\, q_i(T)q_j(T) \;=\; 0,$$

because $m_T(T) = 0$. Finally, since $m_T = p_i^{r_i}q_i$, there exist polynomials a_i, b_i with $a_i(x)q_i(x) + b_i(x)p_i(x)^{r_i} = 1$; evaluating at T shows

$$E_i = c_i(T)q_i(T) \equiv 1 \pmod{p_i(T)^{r_i}} \quad \text{and} \quad E_i \equiv 0 \pmod{p_j(T)^{r_j}}\ (j \neq i). \tag{2}$$

Step 2: Identify ranges with the primary kernels. We claim $\operatorname{Im}(E_i) = W_i$.

Inclusion $\subseteq$. Since $m_T = p_i^{r_i}q_i$, we have

$$p_i(T)^{r_i}E_i = p_i(T)^{r_i}c_i(T)q_i(T) = \bigl(c_i\, m_T\bigr)(T) = 0,$$

so $\operatorname{Im}(E_i) \subseteq \ker\bigl(p_i(T)^{r_i}\bigr) = W_i$.

Inclusion $\supseteq$. If $w \in W_i$, then $p_i(T)^{r_i}w = 0$. Applying (2) to w yields

$$w \;=\; E_iw + b_i(T)p_i(T)^{r_i}w \;=\; E_iw,$$

so $w \in \operatorname{Im}(E_i)$. Hence $\operatorname{Im}(E_i) = W_i$.

Step 3: Direct-sum decomposition. From $\sum_i E_i = I_V$ and $\operatorname{Im}(E_i) = W_i$ we obtain $V = \sum_{i=1}^k W_i$. If $\sum_i w_i = 0$ with $w_i \in W_i$, then applying E_j and using $E_j|_{W_j} = I$ and $E_j|_{W_i} = 0$ for $i \neq j$ (by (2)) gives $w_j = 0$ for each j. Thus the sum is direct, establishing (1).

Step 4: T-invariance. For any polynomial f we have $Tf(T) = f(T)T$. Hence

$$v \in W_i \implies p_i(T)^{r_i}v = 0 \implies p_i(T)^{r_i}(Tv) = T\, p_i(T)^{r_i}v = 0,$$

so $Tv \in W_i$. This proves (2).

Step 5: Minimal polynomials on the primary components. First, $p_i(T)^{r_i}$ annihilates W_i by definition, hence the minimal polynomial $m_{T_{W_i}}$ divides $p_i^{r_i}$; write $m_{T_{W_i}} = p_i^{s_i}$ with $0 < s_i \leq r_i$. Because $V = W_1 \oplus \cdots \oplus W_k$ and each W_i is T-invariant, the operator $\prod_{i=1}^k p_i(T)^{s_i}$ acts as 0 on each W_i, hence on V. Therefore m_T divides $\prod_i p_i^{s_i}$, i.e. $r_i \leq s_i$ for all i. Combining $s_i \leq r_i$ with $r_i \leq s_i$ forces $s_i = r_i$, so $m_{T_{W_i}} = p_i^{r_i}$, proving (3). □

Remarks 8.1.2.2.1. Let us now point out some important observations regarding the primary decomposition theorem:

1. The minimal polynomial of each of the T_i are a power of a single prime factor.
2. If the characteristic polynomial has the same prime factors as the minimal polynomial (i.e. $f(x) = p_1(x)^{d_1} \cdots p_k(x)^{d_k}$) then
$$\dim\left(W_i\right) = \deg\left(p_i\right) \cdot d_i.$$

Ex. Let V be a real vector space. Let $T \in \mathcal{L}(V)$ such that the characteristic polynomial is
$$f(x) = (x-3)^2(x^2+1)^2(x-1)$$
and the minimal polynomial is
$$p(x) = (x-3)(x^2+1)(x-1).$$
Using the primary decomposition theorem, decompose V and provide the minimal polynomial of each subspace W_i and $\dim(W_i)$.

Solution. $f_1(x) = (x-3)^2, f_2(x) = (x^2+1)^2, f_3(x) = (x-1); p_1(x) = (x-3), p_2(x) = (x^2+1), P_3(x) = (x-1)$. For W_1, the min polynomial
$$p_1(x) = (x-3), \dim\left(W_1\right) = \deg\left(x-3\right) \cdot 2 = 2.$$
For W_2,the min polynomial is x^2+1. Thus,
$$\deg\left(W_2\right) = \deg\left(x^2+1\right) \cdot 2 = 4.$$
For W_3, the min polynomial is $x-1$. In particular, $\dim\left(W_2\right) = \deg\left(x-1\right) \cdot 1 = 1$. Thus,
$$\exists \beta \text{ s.t. } [T]_\beta = \begin{bmatrix} A_1 & & \\ & A_2 & \\ & & A_3 \end{bmatrix}$$
where $A_1 \in M_{2\times2}, A_2 \in M_{4\times4}, A_3 \in M_{1\times1},$ s.t. $p_j\left(A_j\right) = 0, \forall j.$ ■

Theorem 8.1.2.3 (Cyclic-Subspaces)**.** Let $\alpha \in V$ and p be the T-annihilator of α. Let $\deg(p) = k$. Then

1. $\{\alpha, T\alpha, \cdots, T^{k-1}\alpha\}$ is linearly independent.
2. $\operatorname{span}(\{\alpha, T\alpha, \cdots, T^{k-1}\alpha\})$ is T-invariant.

The set
$$Z(\alpha; T) = \operatorname{span}(\{\alpha, T\alpha, \cdots, T^{k-1}\alpha\})$$
is called the cyclic subspace of α.

Theorem 8.1.2.4 (Cyclic subspace generated by α)**.** Let $T \in \mathcal{L}(V)$ and $\alpha \in V$. Let $p \in \mathbb{F}[x]$ be the T-annihilator of α, i.e. the monic polynomial of least degree such that $p(T)\alpha = 0$. Write $\deg p = k$. Then

1. $\{\alpha, T\alpha, \dots, T^{k-1}\alpha\}$ is linearly independent;
2. $W := \operatorname{span}\{\alpha, T\alpha, \dots, T^{k-1}\alpha\}$ is T-invariant.

Proof. **(1) Linear independence.** Suppose $\sum_{j=0}^{k-1} c_j T^j \alpha = 0$ with scalars c_j. Set $q(x) := \sum_{j=0}^{k-1} c_j x^j$. Then $q(T)\alpha = 0$ and $\deg q \le k-1 < k = \deg p$. By the minimality of p, the only polynomial of degree $< k$ that annihilates α is the zero polynomial; hence $c_j = 0$ for all j. Thus the listed k vectors are linearly independent.

(2) T-invariance. Let $W = \operatorname{span}\{\alpha, T\alpha, \dots, T^{k-1}\alpha\}$. Since p is the monic annihilator, write
$$p(x) = x^k + b_{k-1}x^{k-1} + \cdots + b_1 x + b_0,$$
so $p(T)\alpha = 0$ gives
$$T^k\alpha = -\sum_{j=0}^{k-1} b_j T^j \alpha \in W. \tag{3}$$

Take any $v = \sum_{j=0}^{k-1} a_j T^j \alpha \in W$. Then
$$Tv = \sum_{j=0}^{k-1} a_j T^{j+1}\alpha = a_{k-1}T^k\alpha + \sum_{j=0}^{k-2} a_j T^{j+1}\alpha \in W,$$

using (3). Hence $T(W) \subseteq W$, i.e. W is T-invariant. □

Theorem 8.1.2.5 (Matrix representation of a cyclic subspace)**.** Let V be a vector space, $\alpha \in V$, $T \in \mathcal{L}(V)$, $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be the T-annihilator of α, $\beta = \{\alpha, T\alpha, \ldots, T^{n-1}\alpha\}$ be a basis for $Z(\alpha; T)$, T_Z be the linear operator defined by restriction of T onto the subspace $Z(\alpha; T)$. Then

$$[T_Z]_\beta = C(p) = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}$$

Furthermore, the minimal polynomial and characteristic polynomial of T_Z are both equal to $p(x)$. We call the matrix $C(p)$ the *companion matrix* of the polynomial p.

Theorem 8.1.2.6 (Cyclic Decomposition Theorem)**.** Let $T \in \mathcal{L}(V)$, $f(x) = p_1(x)^{d_1} \cdots p_k(x)^{d_k}$ be the characteristic polynomial and $p(x) = p_1(x)^{r_1} \cdots p_k(x)^{r_k}$ be the minimal polynomial. Then there exists $\alpha_1, \ldots, \alpha_j$ such that

1. $V = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_j; T)$,
2. $p = p_1$,
3. p_i divides p_{i-1} for all $i \in \{2, \ldots, j\}$,

where p_j is the minimal polynomial of $Z(\alpha_j; T)$. Note that by construction

$$f = p_1 p_2 \cdots p_j.$$

Lemma 8.1.2.7. Let $T \in \mathcal{L}(V)$, p be the minimal polynomial of T, and $\alpha \in V$ such that p is the T-annihilator of α. Then there exists $W \subset V$ such that:

1. $V = Z(\alpha; T) \oplus W$
2. W is T-invariant.

Remarks 8.1.2.7.1 (Significance of the Cyclic Decomposition Theorem)**.** The Cyclic Decomposition Theorem tells us that every finite-dimensional vector space can be written as the direct sum of cyclic invariant subspaces. This has the following important consequences for matrix representations of linear transformations:

Let $T \in \mathcal{L}(V)$ with characteristic polynomial f and minimal polynomial p. Then there exists a basis β such that

$$[T]_\beta = \begin{bmatrix} C(p_1) & 0 & \cdots & 0 \\ 0 & C(p_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & C(p_j) \end{bmatrix},$$

where $p = p_1$, p_i divides p_{i-1} for all i, and $f = p_1 \cdots p_j$, and $C(p_i)$ are the companion matrices of the polynomial p_i.

Definition 8.1.3 (The Rational Canonical Form)**.** Let $T \in \mathcal{L}(V)$. Then $[T]_\beta$ is in Rational Canonical Form if

$$[T]_\beta = \begin{bmatrix} C(p_1) & 0 & \cdots & 0 \\ 0 & C(p_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & C(p_j) \end{bmatrix},$$

where $C(p_i)$ are the companion matrices of the polynomial p_i, $p = p_1$, p_i divides p_{i-1} for all i, and $f = p_1 \cdots p_j$.

Theorem 8.1.3.1. Two matrices are similar if and only if they have the same rational form.

Ex. Let V be a real vector space. Let $T \in \mathcal{L}(V)$ with characteristic polynomial

$$f(x) = (x-2)^2(x-3)^3(x^2+1)$$

and minimal polynomial

$$p(x) = (x-2)(x-3)(x^2+1).$$

Construct the rational canonical form of T.

Solution. $Z(\alpha_1; T)$ has min polynomial $p(x) = (x-2)\,(x-3)\,(x^2+1)$. Then, for W_1,

$$f_{w_1} \cdot p_1 = f \implies \frac{f}{p_1} = f_{w_1} \implies f_{w_1} = \frac{(x-2)^2\,(x-3)^3\,(x^2+1)}{(x-2)\,(x-3)\,(x^2+1)} \implies f_{w_1} = (x-2)\,(x-3)^2.$$

Thus, p_{w_1} containts *all rots of* f_{w_1} and $p_{w_1} | p_2$. Since $p_2 | p_1$. Thus, we have

$$p_{w_1} = (x-2)^{k_1}\,(x-3)^{k_2} \text{ s.t. } p_{w_1} |\, (x-2)\,(x-3)\,\left(x^2+1\right).$$

It follows that $p_{w_1} = (x-2)\,(x-3)$. $Z\,(\alpha_2; T)$ has min polynomial $p_2\,(x) = (x-2)\,(x-3)$. Thus, $f_{w_2} = \frac{f_{w_1}}{p_2} = \frac{(x-2)(x-3)^2}{(x-2)(x-3)} = (x-3)$. Note that $p_{w_2} | p_{w_1} \implies p_{w_2} |\, (x-2)\,(x-3)$. Thus,

$$p_{w_2} = (x-3)^{k_1} \text{ s.t. } p_{w_2} |\, (x-2)\,(x-3) \implies p_{w_2} = (x-3).$$

Lastly, $Z\,(\alpha_3; T)$ with minimal polynomial $(x-3)$. Thus,

$$V = Z\,(\alpha_1; T) \oplus Z\,(\alpha_2; T) \oplus Z\,(\alpha_3; T) \text{ s.t. } \begin{cases} p_1(x) = (x-2)\,(x-3)\,(x^2+1) = x^4 - 5x^3 + 7x^2 - 5x + 6 \\ p_2(x) = (x-2)\,(x-3) = x^2 - 5x + 6 \\ p_3(x) = (x-3). \end{cases}$$

Now, we **construct the rational form.**

$$[T]_\beta = \begin{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & -6 \\ 0 & 1 & 0 & 0 & 5 \\ 0 & 0 & 1 & 0 & -7 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix} & [0] & [0] \\ [0] & \begin{bmatrix} 0 & -6 \\ 1 & 5 \end{bmatrix} & [0] \\ [0] & [0] & 3 \end{bmatrix}$$

■

Ex. Decompose the rational cononical form of T further by applying the primary decomposition theorem.

Solution. We know that

$$W_1 = W_{11} \oplus W_{12} \oplus W_{13} \text{ s.t. } p_{11}(x) = x-2, p_{12}(x) = x-3, p_{22}(x) = x^2+1$$

$$W_2 = W_{21} \oplus W_{22} \text{ s.t. } p_{21}(x) = (x-2), p_{22}(x) = (x-3).$$

Thus,

$$[T]_\beta = \left[fillitup\right]$$

■

8.2 Jordan Canonical Form

Definition 8.2.1 (Jordan Blocks and Jordan Canonical Form)**.** Suppose $T \in \mathcal{L}(V)$. A basis of V is called a *Jordan basis* for T if with respect to this basis T has a block diagonal matrix

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_p \end{pmatrix},$$

in which each A_k is an upper-triangular matrix of the form

$$A_k = \begin{pmatrix} \lambda_k & 1 & 0 & \cdots & 0 \\ 0 & \lambda_k & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k & 1 \\ 0 & 0 & \cdots & 0 & \lambda_k \end{pmatrix}.$$

A matrix in this form is said to be in *Jordan Canonical Form.*

In such a case, these blocks are called *Jordan Blocks.*

A matrix representation of T in Jordan Canonical Form is said to be a Jordan Canonical form for T.

$$p(x) = (x-\lambda)^4.$$

Then, we have

$$\begin{bmatrix} \lambda & 0 & 0 & 0 \\ 1 & \lambda & 0 & 0 \\ 0 & 1 & \lambda & 0 \\ 0 & 0 & 1 & \lambda \end{bmatrix}$$

we have a jordan block if $p(x) = (x-\lambda)^k$.

Theorem 8.2.1.1 (Jordan Basis Exist for triangularizable linear transformations)**.** Let V be a finite-dimensional vector space over $\mathbb{F}$, $T \in \mathcal{L}(V)$.

If T is triangularizable, there exists a Jordan Basis for T.