MA2101

Linear Algebra II

Jia Xiaodong

April 26, 2020

1 Introduction

These are notes from Linear Algebra II, a more rigorous treatment of concepts introduced in LA I.

2 Matrices

2.1 Fields

We use Gaussian elimination to solve linear systems. Gaussian elimination works if one can add, subtract, multiply, and divide the coefficients similar to the real numbers. That is, the coefficients come from a *field*.

Definition 2.1. A *field* $(\mathbb{F}, +, \times)$ is a set $\mathbb{F}$ together with two binary operations $+$ and $\times$ called *addition* and *multiplication* respectively satisfying the *field axioms*[1]. □

In summary, the axioms state that the two operations are associative, commutative, have an identity, and have an inverse. Also multiplication is distributive over addition.

Axioms (A1) - (A4) says that $(\mathbb{F}, +)$ forms a *commutative group*. Axioms (M1) - (M4) says that $(\mathbb{F} - \{0\}, \times)$ forms another *commutative group*.

Example 2.1. $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$ are fields. $\mathbb{Z}$ is not a field, and neither is $\mathbb{R} \setminus \mathbb{Q}$. □

A *finite field* is a field which contains only finitely many elements. The number of elements in a finite field is of the form p^m, where p is prime and m is a positive integer. There is exactly one field (up to isomorphism) with q elements. We can safely call this $\mathbb{F}_q$.

Matrices are a familiar concept from Linear Algebra I. We can simply extend addition and product to any arbitrary field without additional work.

We denote $M_{mn}(\mathbb{F})$ as the set of all $m \times n$ matrices with entries taken from $\mathbb{F}$.

Definition 2.2. A square matrix A is *invertible* if there is a matrix B such that

$$AB = BA = I$$

B is called the *inverse* of A and we write $A^{-1} = B$. □

[1]These are stated in the Math 115 notes.

2.2 Determinants

To determine invertibility we previously we had a concept of *determinant*. We treat it more rigorously now:

Definition 2.3. A function

$$D : M_n(\mathbb{F}) \to \mathbb{F}$$

is called a *determinant function* if:

- (D1) it is multilinear:
  - For columns of the matrix, $D(\ldots, \alpha u + \beta v, \ldots) = \alpha D(\ldots, u, \ldots) + \beta D(\ldots, v, \ldots)$
  - In other words, it is linear for each individual column with the others held fixed.
- (D2) it is alternating:
  - If A' is formed from A by interchanging two columns, then $D(A') = -D(A)$.
  - Thus if A has two equal columns, $D(A) = 0$
- (D3) $D(1) = 1$

□

It can be shown that this definition of the determinant function will give us all the familiar properties of the determinant. However we will skip this, as well as enumerating through its properties since we have dealt with determinants plenty already in Linear Algebra I. What we are more interested in is if our familiar method of calculating determinants is the only way of doing so.

Example 2.2. The function $D : M_2(\mathbb{F}) \to \mathbb{F}$ given by:

$$D \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

is easily verified as a valid determinant function. □

Let $A \in M_n(\mathbb{F})$. For $1 \leq i, j \leq n$, let $\tilde{A}_{ij}$ be the $(n-1) \times (n-1)$ matrix obtained by deleting the i-th row and j-th column from A.

Theorem 2.1 (Cofactor expansion)**.**

$$E(X) = \sum_{j=1}^{n} (-1)^{i+j} x_{ij} D(\tilde{A}_{ij})$$

is a determinant function on $M_n(\mathbb{F})$.

Using this fact, we can inductively generate more determinant functions.

Corollary 2.1.1. *For each positive integer* n*, there is at least one determinant function on* $M_n(\mathbb{F})$*.*

Theorem 2.2. *The determinant function for $M_2(\mathbb{F})$ is unique and is the one given in Ex. 2.2.*

Proof.

$$\begin{aligned} D\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} &= D\left(\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}\right) \\ &= D\left(a_{11}\begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_{21}\begin{pmatrix} 0 \\ 1 \end{pmatrix}, a_{12}\begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_{22}\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \\ &= a_{11}a_{12}D(\mathbf{e}_1, \mathbf{e}_1) + a_{11}a_{22}D(\mathbf{e}_1, \mathbf{e}_2) + a_{21}a_{12}D(\mathbf{e}_2, \mathbf{e}_1) + a_{22}a_{22}D(\mathbf{e}_2, \mathbf{e}_2) \\ &= 0 + a_{11}a_{22}D(\mathbf{e}_1, \mathbf{e}_2) - a_{21}a_{12}D(\mathbf{e}_1, \mathbf{e}_2) + 0 \\ &= (a_{11}a_{22} - a_{21}a_{12})D(\mathbf{I}) \end{aligned}$$

By construction $D(\mathbf{I}) = 1$. Hence we get that any determinant function has to evaluate to the one given in Ex. 2.2. ■

We need the idea of permutations for higher order determinants.

Definition 2.4. A *permutation* of $\{1, 2, \ldots, n\}$ is an one-to-one function

$$\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$$

and we denote it as

$$\sigma = (\sigma 1, \sigma 2, \ldots, \sigma n)$$

□

The symmetric group of degree n is denoted as S_n, which is the set of all permutations of $\{1, 2, \ldots, n\}$. S_n forms a group under composition of functions.

If we want to extend the expansion above to larger n, we may note that actually we can instead write

$$D(A) = \sum_{\sigma \in S_n} a_{\sigma 1,1} \ldots a_{\sigma n,n} D(\mathbf{e}_{\sigma 1}, \ldots, \mathbf{e}_{\sigma n})$$

To see this fact, we can stare hard at the proof for Thm. 2.2.

From the original construction of the determinant function, swapping the rows

$$D(\ldots, \mathbf{e}_{\sigma 1}, \mathbf{e}_{\sigma 2}, \ldots) = -D(\ldots, \mathbf{e}_{\sigma 2}, \mathbf{e}_{\sigma 1}, \ldots)$$

Then any permutation of the rows causes

$$D(\mathbf{e}_{\sigma 1}, \ldots, \mathbf{e}_{\sigma n}) = \operatorname{sgn} \sigma D(\mathbf{I})$$

with the signum function sgn reacting to how many swaps we have performed:

$$\operatorname{sgn} \sigma = \begin{cases} 1, & \text{if even switches} \\ -1, & \text{if odd switches} \end{cases}$$

Theorem 2.3 (Uniqueness of determinants)**.** *For each n, there is only one determinant function on $M_n(\mathbb{F})$, given by*

$$\det(A) = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) a_{\sigma 1,1} \ldots a_{\sigma n,n}$$

Since there is only one determinant function, this is equivalent to our expression obtained from cofactor expansion in Thm. 2.1.

Theorem 2.4. *If a function $D : M_n(\mathbb{F}) \to \mathbb{F}$ is multilinear and alternating, then*

$$D(A) = (\det A)(D(\mathbf{I}))$$

Theorem 2.5. *If $A, B \in M_n(\mathbb{F})$, then*

$$\det(AB) = (\det A)(\det B)$$

Proof. Define a function $f : M_n(\mathbb{F}) \to \mathbb{F}$ as $f(\mathbf{X}) = \det(\mathbf{AX})$. It is easily checked that f is both alternating and multilinear. Then by Thm 2.4,

$$\begin{aligned} f(\mathbf{B}) &= (\det \mathbf{B}) f(\mathbf{I}) \\ &= (\det \mathbf{B})(\det \mathbf{A}) \end{aligned}$$

■

Theorem 2.6 (Determinant of transposes)**.**

$$\det A = \det A^T$$

Proof. First we note that for two permutations $\sigma, \tau \in S_n$, $\operatorname{sgn} \sigma\tau = \operatorname{sgn} \sigma \operatorname{sgn} \tau$, where $\sigma\tau$ represents their composition. This also implies that if σ^{-1} is the inverse function of σ, then $\operatorname{sgn} \sigma\sigma^{-1} = 1$, and $\operatorname{sgn} \sigma = \operatorname{sgn} \sigma^{-1}$. All these statements stem from sgn being an indicator of the parity of the number of switches we need to create a permutation from $(1, 2, \ldots, n)$.

Let $A = (a_{ij})$ and $A^T = (a^T_{ij}) = (a_{ji})$. The determinant of A^T is then

$$\det\left(A^T\right) = \sum_{\tau \in S_n} (\operatorname{sgn} \tau) a^T_{\tau 1,1} \ldots a^T_{\tau n,n} = \sum_{\tau \in S_n} (\operatorname{sgn} \tau) a_{1,\tau 1} \ldots a_{n,\tau n}$$

We can rearrange the terms $a_{1,\tau 1} \ldots a_{n,\tau n}$ such that they are in the form of $a_{\sigma 1,1} \ldots a_{\sigma n,n}$ without changing their value. Then it is clear that $\sigma = \tau^{-1}$, and $\operatorname{sgn} \sigma = \operatorname{sgn} \tau$. Thus,

$$\det A^T = \sum_{\tau \in S_n} (\operatorname{sgn} \tau) a_{1,\tau 1} \ldots a_{n,\tau n} = \sum_{\tau \in S_n} (\operatorname{sgn} \sigma) a_{\sigma 1,1} \ldots a_{\sigma n,n} = \det A$$

■

Naturally this leads to the following theorem.

Theorem 2.7 (Cofactor expansion along rows)**.**

$$\det X = \sum_{i=1}^{n} (-1)^{i+j} x_{ij} \det \tilde{A}_{ij}$$

Theorem 2.8 (Classical adjoint and inverses)**.** *Let $A = (a_{ij}) \in M_n(\mathbb{F})$ and $\operatorname{adj} A = (b_{ij}) \in M_n(\mathbb{F})$ be defined as*

$$b_{ij} = (-1)^{i+j} \det \tilde{A}_{ji}$$

$\operatorname{adj} A$ *is the* classical adjoint *of A, and*

$$(\det A)^{-1} (\operatorname{adj} A) A = \mathbf{I}$$

Proof. Let $C = (\text{adj } A)A = (c_{ij})$. Then

$$\begin{aligned} c_{ij} &= \sum_k (\text{adj } A)_{ik} a_{kj} \\ &= \sum_k (-1)^{i+k} (\det \tilde{A}_{ki}) a_{kj} \end{aligned}$$

We note that for $j \neq k$,

$$\sum_i (-1)^{i+j} a_{ik} \det \tilde{A}_{ij} = 0$$

since this is the expression of $\det B$ where B is formed by replacing the j-th row of A with the k-th row of A.

Thus,

$$c_{ij} = \begin{cases} 0 & j \neq i \\ \det A & \text{otherwise} \end{cases}$$

Hence $C = (\det A)\mathbf{I}$. ■

Thm. 2.8 also tells us that a matrix A is invertible iff $\det A \neq 0$.

Theorem 2.9 (Cramer's Rule)**.** *Let $A \in M_n(\mathbb{F})$ and $\mathbf{y} \in \mathbb{F}^n$. If A is invertible, then for the system of linear equations:*

$$A\mathbf{x} = \mathbf{y}$$

with $\mathbf{x} \in \mathbb{F}^n$, then for each $1 \leq j \leq n$,

$$\mathbf{x}_j = (\det A)^{-1} \det C(j)$$

where $C(j)$ is the $n \times n$ matrix obtained by replacing the j-th row of A by $\mathbf{y}$.

Proof. If A is invertible, then its inverse exists, and from Thm. 2.8, $A^{-1} = (\det A)^{-1} \text{adj } A$. Now consider $(\text{adj } A)\mathbf{y} = B = (b_i)$. Using the definition of the classical adjoint,

$$\begin{aligned} b_i &= \sum_{k=1}^{n} (\text{adj } A)_{ik} \mathbf{y}_k \\ &= \sum_{k=1}^{n} (-1)^{k+1} \det \tilde{A}_{ki} \mathbf{y}_k \end{aligned}$$

Now if we perform cofactor expansion over column k for $C(k)$, we find that

$$\det C(k) = \sum_{j=1}^{n} (-1)^{i+j} \mathbf{y}_j \det \tilde{A}_{jk}$$

Hence $b_i = \det C(k)$.

$$\begin{aligned} A\mathbf{x} &= \mathbf{y} \\ A^{-1}A\mathbf{x} &= A^{-1}\mathbf{y} \\ \mathbf{x} &= (\det A)^{-1} B \end{aligned}$$

Therefore $\mathbf{x}_i = (\det A)^{-1} b_i = (\det A)^{-1} \det C(i)$ ■

3 Vector spaces

3.1 Vector spaces

$\mathbb{R}^n$ consists of tuples of real numbers. In $\mathbb{R}^2$ and $\mathbb{R}^3$ we could represent them as vectors. We are quite familiar with them now; but we want to know if we can extend some of those concepts further:

- Linear combination
- Linear dependence/independence
- Subspaces, bases and dimensions
- Linear transformations

Actually, these concepts come not from R^n, but the operations on R^n. So with these two notions of vector addition and scalar multiplication, we can generalize beyond R^n.

We shall call any system with a notion of addition and scalar multiplication behaving in a certain way (like those in $\mathbb{R}^n$) *vector spaces* and their elements *vectors*.

Definition 3.1. A vector space consists of

i. A field $\mathbb{F}$ of scalars

ii. A set V of vectors

iii. A rule called *vector addition*,

$$\forall \mathbf{u}, \mathbf{v} \in V, (\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} + \mathbf{v}$$

iv. A rule called *scalar multiplication*,

$$\forall k \in F, \mathbf{v} \in V, (k, \mathbf{v}) \mapsto k\mathbf{v}$$

□

The operations must obey these rules:

(A1) V is closed under vector addition.

(A2) Vector addition is commutative.

(A3) Vector addition is associative.

(A4) A additive identity, the *zero vector* $\mathbf{0}$, exists.

(A5) Additive inverses exist for all elements.

(S1) V is closed under scalar multiplication.

(S2) $\forall k, l \in \mathbb{F}, \forall \mathbf{v} \in V, (kl)\mathbf{v} = k(l\mathbf{v})$

(S3) $\forall \mathbf{v} \in V, 1\mathbf{v} = \mathbf{v}$.

(S4,S5) Scalar multiplication is distributive over vector addition.

Example 3.1. For any field $\mathbb{F}$,

$$\mathbb{F}^n = \{(a_1, \ldots, a_n) \mid a_1, \ldots, a_n \in \mathbb{F}\}$$

with the operations

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n)$$
$$k(a_1, \ldots, a_n) = (ka_1, \ldots, ka_n)$$

forms a vector space over $\mathbb{F}$. □

Example 3.2. Take the set of all $m \times n$ matrices A with entries taken from field $\mathbb{F}$, $M_{mn}(\mathbb{F})$.

For $A = (a_{ij}), B = (b_{ij}) \in M_{mn}(\mathbb{F})$ and $k \in \mathbb{F}$,

$$A + B = (a_{ij} + b_{ij})$$
$$kA = (ka_{ij})$$

$M_{mn}(\mathbb{F})$ forms a vector space over $\mathbb{F}$. □

Example 3.3. Let S be a non-empty set, $\mathbb{F}$ be a field, and consider $\mathcal{F}(S, \mathbb{F})$ as the set of all functions $f : S \to \mathbb{F}$

Define for $f, g \in \mathcal{F}$ the function $f + g$ with some $s, k \in S$

$$(f + g)(s) = f(s) + g(s)$$
$$(kf)(s) = k(f(s))$$

$\mathcal{F}$ with the above operations form a vector space over $\mathbb{F}$. □

Example 3.4. Let $\mathcal{P}(\mathbb{F})$ be the set of all polynomials with coefficients in $\mathbb{F}$.

Define vector addition as resulting in a new polynomials whose coefficients are the sums of the respective coefficients in both operands.

Define scalar multiplication as returning a polynomial whose coefficients are the results of multiplying the respective coefficients in the vector with the scalar.

$\mathcal{P}(\mathbb{F})$ forms a vector space over $\mathbb{F}$. □

Similar to when we just started with fields, we should prove some of these facts that we usually take for granted.

Theorem 3.1. *Let V be a vector space over $\mathbb{F}$, $\mathbf{v} \in V$ and $k \in \mathbb{F}$. Then*

i. $0\mathbf{v} = \mathbf{0}$

ii. $k\mathbf{0} = \mathbf{0}$

iii. $(-1)\mathbf{v} = -\mathbf{v}$

iv. $k\mathbf{v} = \mathbf{0} \rightarrow (k = 0 \wedge = \mathbf{0})$.

Proof. i.

$$\begin{aligned} 0 &= 0 + 0 \\ 0\mathbf{v} &= (0+0)\mathbf{v} \\ 0\mathbf{v} &= 0\mathbf{v} + 0\mathbf{v} \\ 0\mathbf{v} + (-(0\mathbf{v})) &= 0\mathbf{v} + 0\mathbf{v} + (-(0\mathbf{v})) \\ \mathbf{0} &= 0\mathbf{v} + \mathbf{0} \\ \mathbf{0} &= 0\mathbf{v} \end{aligned}$$

■

3.2 Subspaces

Definition 3.2. Let V be a vector space over a field $\mathbb{F}$.

If a subset W of V forms a vector space of $\mathbb{F}$ as well with the same operations as those in V, then it is called a *subspace*. □

It might be noted that the zero vector being present is a necessary condition for a vector space to be a subspace.

Example 3.5. The xy plane in $\mathbb{R}^3$ forms a real vector space with the usual vector addition and scalar multiplication. Hence it is a subspace of $\mathbb{R}^3$. □

Often we want to know if a subset of some vector space is an subspace of it.

Theorem 3.2. *Let V be a vector space over $\mathbb{F}$ and W be a non-empty subset of V. Then W is a subspace of V iff $\forall \mathbf{u}, \mathbf{v} \in W, \forall \alpha, \beta \in \mathbb{F}, \alpha\mathbf{u} + \beta\mathbf{v} \in W$.*

Proof.
($\Longrightarrow$): If W is a subspace, then it is also a vector space. Hence it is closed under scalar multiplication and vector addition.

($\Longleftarrow$): We need to show that W satisfies all 10 conditions to be qualified as a vector space.

Firstly, associativity, commutativity, distributivity, multiplicative inverse(A2, A3, S2, S3, D1, D2) are true free of charge since W comes from V.

W is also obviously closed under addition due to our original supposition.

Next we quickly prove the remaining few conditions:

$$0\mathbf{u} + 0\mathbf{v} = 0 \in W$$

$$(-1)\mathbf{u} + 0\mathbf{v} = -\mathbf{u} \in W$$

$$\alpha\mathbf{u} + 0\mathbf{v} = \alpha\mathbf{u} \in W$$

■

Example 3.6. Take the solution space of a homogeneous linear system:

$$W = \left\{ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{F}^n \mid A\mathbf{x} = \mathbf{0} \right\}$$

W is a subspace of $\mathbb{F}^n$ □

Proof. If $\mathbf{x} = \mathbf{0}$, then $A\mathbf{x} = \mathbf{0}$, so $0 \in W$ and W is non-empty.

Take $u, v \in W$ and $\alpha, \beta \in \mathbb{F}$. Then $A\mathbf{u} = A\mathbf{v} = 0$.

$$\begin{aligned} A(\alpha\mathbf{u} + \beta\mathbf{v}) &= \alpha A\mathbf{u} + \beta A\mathbf{v} \\ &= \alpha 0 + \beta 0 \\ &= 0 \in W \end{aligned}$$

Hence by Thm. 3.2, W is a subspace of F^n. ■

Example 3.7. Let $\mathbb{F}\{\mathbb{R}, \mathbb{R}\}$ be the set of all functions $f : \mathbb{R} \to \mathbb{R}$. Let $\mathcal{C}\{\mathbb{R}\}$ be the set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$. The zero vector is represented by the zero function $\mathcal{C}\{\mathbb{R}\} \ni f_0(x) = 0$.

We also know that if we have two continuous functions f, g, then $f + g$ is also continuous. Also, $\forall \alpha$, αf is continuous. Therefore $\alpha f + \beta g$ is also continuous. Hence $\mathcal{C}\{\mathbb{R}\}$ is a subspace of $\mathbb{F}\{\mathbb{R}, \mathbb{R}\}$. □

Example 3.8. Let V be the set of all sequences of real numbers. For $(a_b), (b_n) \in V$, $\alpha \in \mathbb{R}$,

$$(a_n)_n + (b_n)_n = (a_n + b_n)_n \qquad \alpha(a_n)_n = (\alpha a_n)_n$$

V with these operations form a vector space. The set of all convergent sequences hence also form a subspace. □

If we want to combine subspaces, what comes to mind first when is that we may take their union. However it turns out that most of the time this will not work. For example, the union of the x and y axes in $\mathbb{R}^3$ clearly does not create a subspace.

Theorem 3.3. *Let U and W be subspaces of a vector space V.*

i. $U \cap W$ is a subspace of V.

ii. If $U \cup W$ is a subspace of V, then either $U \subseteq W$ or $W \subseteq U$.

Proof.

i. Take $\mathbf{u}, \mathbf{w} \in U \cap W$. Then $\mathbf{u}, \mathbf{w} \in U \wedge \mathbf{u}, \mathbf{w} \in W$. Hence $\alpha\mathbf{u} + \beta\mathbf{w} \in U$ and $\alpha\mathbf{u} + \beta\mathbf{w} \in W$. Thus $\alpha\mathbf{u} + \beta\mathbf{w} \in U \cap W$.

ii. Take $\mathbf{u} \in U$, and $\mathbf{w} \in W$. Then $\mathbf{u}, \mathbf{w} \in U \cup W$. Since it is a subspace, $\mathbf{u} + \mathbf{w} \in U \cup W$. Hence $(\mathbf{u}+\mathbf{w} \in U) \vee (\mathbf{u}+\mathbf{w} \in W)$. But then since they are both subspaces, $(\mathbf{u}+\mathbf{w}-\mathbf{u} \in U) \vee (\mathbf{u}+\mathbf{w}-\mathbf{w} \in W)$. Thus either $U \subseteq W$ or $W \subseteq U$.

■

The proper way to combine subspaces is to actually take their sum.

Definition 3.3. Define $V + W = \{\mathbf{v} + \mathbf{w} \mid \mathbf{v} \in V, \mathbf{w} \in W\}$. This can be verified to be a subspace. □

3.3 Linear spans

Definition 3.4. A vector $\mathbf{u}$ of the form $\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$ where $a_1, \ldots, a_n \in \mathbb{F}$, is called a *linear combination* of $\mathbf{v}_1, \ldots, \mathbf{v}_n$. □

Definition 3.5. The *span* of S is the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_n \in S$. It is defined that $\operatorname{span}(\emptyset) = \{\mathbf{0}\}$. If $W = \operatorname{span}(S)$, we say W is spanned by S, or S is the spanning set for W. □

Theorem 3.4. *Let V be a vector space over $\mathbb{F}$ and S be a finite subset of vectors in V. Then:*

i. $\operatorname{span}(S)$ *is a subspace of V.*

ii. If W is a subspace of V and $S \subseteq W$, then $\operatorname{span}(S) \subseteq W$.

In other words, the span of S is the smallest subspace containing S.

Proof.

i. Use Thm. 3.2.

ii. Suppose the finite set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq W$. Take $\mathbf{v} \in \operatorname{span}(S)$. Then $\mathbf{v}$ is of the form $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$. Since $S \subseteq W$, $\mathbf{v}_1, \ldots, \mathbf{v}_n \in W$. W is closed under scalar multiplication and vector addition. Hence $\mathbf{v} \in W$. Thus $\operatorname{span}(S) \subseteq W$.

■

In fact, we may say that

$$\operatorname{span}(S) = \bigcap_{\substack{S \subseteq W \\ W \text{ is a subspace}}} W$$

Example 3.9. Let $\mathbb{F}$ be a field. Consider the n-tuples with the ith entry as 1:

$$\mathbf{e}_i = (\underbrace{0, \ldots, 1}_{i \text{ entries}}, 0, \ldots, 0)$$

$\mathbb{F}^n = \operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_n)$. □

Example 3.10. Let E_{ij} be the $m \times n$ matrix with its ijth entry as 1 and 0 elsewhere.

If $A = (a_{ij}) \in M_{mn}(\mathbb{F})$, then

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} E_{ij}$$

Therefore $M_{mn}(\mathbb{F}) = \operatorname{span}(E_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq n)$. □

Example 3.11. Consider

$$W = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{22}(\mathbb{R}) \mid a + b + c + d = 0 \right\}$$

It can be shown that W is a subspace of $M_{22}(\mathbb{R})$.

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in W$. Then using the fact that $d = -a - b - c$,

$$A = a \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + b \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} + c \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}$$

Then it is clear what the spanning set of W is. □

Example 3.12. Let $\mathcal{P}_n(\mathbb{F})$ be the space of polynomials in $\mathbb{F}$ of degree at most n.

$\mathcal{P}_n(\mathbb{F}) = \operatorname{span}(\{1, x, x^2, \ldots, x^n), \text{ with } 1, x \in \mathbb{F}$. □

However we can also have more complicated spanning sets.

Example 3.13. Consider $p_1(x) = x^2 + 3x - 2$, $p_2(x) = 2x^2 + 5x - 3$, $p_3(x) = -x^2 - 4x + 4$. They span $\mathcal{P}_2(\mathbb{R})$.

For any $q(x) = ax^2 + bx + c \in \mathcal{P}_2(\mathbb{R})$, we want to show that $c_1p_1(x) + c_2p_2(x) + c_3p_3(x) = q(x)$ has a solution for all $a, b, c \in \mathbb{R}$. After simplification, it is easy to show with Gaussian elimination that there are indeed always solutions. □

Example 3.14. We can also use the rows and columns of matrices as spanning sets. We call the space spanned by all the rows of a matrix its *row space*, and the space spanned by all the columns of a matrix its *column space*. □

It might be noted that linear combinations are only defined for a finite set of vectors. We will get into trouble if we try to allow infinite sums. Consider $\mathbb{Q}$ as a vector space over $\mathbb{Q}$.

$$\sum_{n=0}^{\infty} \frac{1}{n!} = e \notin \mathbb{Q}$$

However we can still have spans of infinite sets. We take a finite subset of S, and form all linear combinations over it. In other words,

$$\operatorname{span}(S) = \bigcup_{\substack{s \subset S \\ s \text{ finite}}} \operatorname{span}(s)$$

The span of S can also be thus stated as the set of all linear combinations of elements in S. Thm. 3.4 still holds for this definition.

It can be seen that every subspace spans itself. However, realistically we want more useful, or even better, the smallest spanning set possible.

Example 3.15. $\mathcal{P}(\mathbb{F}) = \operatorname{span}(\{1, x, x^2, x^3, \ldots\})$. □

To achieve this we need the idea of linear dependence.

3.4 Linear dependence

Let $\mathbf{v}_1, \mathbf{v}_n \in V$, where V is a vector space. Consider the equation

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$$

A trivial solution is to let all $a_1, \ldots, a_n = 0$

Definition 3.6. The vectors $\mathbf{v}_1, \mathbf{v}_n \in V$ are *linearly dependent* if there exists a non-trivial solution to the following equation:

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$$

If there are not linearly dependent, we say they are *linearly independent*, i.e only the trivial solution can satisfy the equation above □

Any finite set containing the zero vector is linearly dependent since we can cheat by doing

$$0\mathbf{v}_1 + \cdots + 0\mathbf{v}_n + 1 \cdot \mathbf{0} = \mathbf{0}$$

Also, if we only have one vector, it is linearly independent, by the uniqueness of the 0 element.

Definition 3.7. Let S be a subset (potentially infinite) of a vector space. If S has a finite subset that is linearly dependent, we also say that S is linearly dependent. If every finite subset of S is linearly independent, then we also say S is linearly independent. Since the empty set is in all sets, it is linearly independent. □

Theorem 3.5. *If $A \in M_n(\mathbb{F})$ is invertible, then its columns form a linearly independent set in $\mathbb{F}^n$.*

Proof. Let the i-th column of A be represented as $\mathbf{a}_i$. Consider the linear system

$$c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n = \mathbf{0}$$

We can rewrite this as

$$A\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \mathbf{0}$$

Since A is invertible, multiplying A^{-1} on both sides gets us that every c_i is 0. ■

Theorem 3.6. *If a subspace W of a vector set is spanned by S and S is linearly dependent, then there exists a vector $\mathbf{v}_i \in S$ such that $W = \operatorname{span}(S \setminus \{\mathbf{v}_i\})$.*

Proof. Since S is linearly dependent, $\exists 1, \cdots, c_n \in \mathbb{F}$, which are not all 0, such that $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = 0$.

Assume $c_i \neq 0$. Then after some arrangement we can obtain $\mathbf{v}_i = -\frac{c_1}{c_i}\mathbf{v}_1 - \cdots - \frac{c_n}{c_i}\mathbf{v}_n$. This means that $\mathbf{v}_i$ is a linear combination of the other vectors in S.

Now we have to show that removing $\mathbf{v}_i$ does not change the set spanned by S.

Take some $\mathbf{w} \in W = \text{span}(S \setminus \{\mathbf{v}_i\})$. Then $\mathbf{w} = k_1\mathbf{v}_1 + \cdots + k_n\mathbf{v}_n$ for some $k_1, \ldots, k_n \in \mathbb{F}$. But we could just replace $\mathbf{v}_i$ with the expression found above and obtain $\mathbf{w} = (k_1 - \frac{k_i c_1}{c_i})\mathbf{v}_1 + \cdots + (k_{i-1} - \frac{k_i c_{i-1}}{c_i})\mathbf{v}_{i-1} + \ldots + (k_{i+1} - \frac{k_i c_{i+1}}{c_i})\mathbf{v}_{i+1} + (k_n - \frac{k_i c_n}{c_i})\mathbf{v}_n$. This shows that $\mathbf{w}$ is a linear combination of vectors in $S \setminus \{\mathbf{v}_i\}$. ■

In other words, linearly dependent spanning sets contain redundant vectors that can be removed. Furthermore, the smallest spanning set should therefore be linearly independent.

3.5 Bases and dimensions

Definition 3.8. A subset B of a vector space V is called a *basis* for V if

i. B spans V.

ii. B is linearly independent.

If V has a finite basis, then we say V is *finite dimensional.* Otherwise it is *infinite dimensional.* □

Theorem 3.7. *Let $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for the vector space V. Then every vector $\in V$ can be expressed in the form of*

$$\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$$

uniquely.

Proof. Since B spans V, there should be at least one way to express $\mathbf{v}$ as a linear combination of the vectors in B.

Now suppose there are two ways of writing such a linear combination.

$$\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$$
$$\mathbf{v} = b_1\mathbf{v}_1 + \ldots + b_n\mathbf{v}_n$$

Consider their difference:

$$0 = (a_1 - b_1)\mathbf{v}_1 + \ldots + (a_n - b_n)\mathbf{v}_n$$

Since B is linearly independent, all the coefficients $a_1 - b_1, \ldots, a_n - b_n$ should all be 0. Hence $a_1 = b_1, \ldots, a_n = b_n$. ■

Example 3.16. We have encountered the spanning set for $\mathbb{F}^n$ before,

$$\mathbf{e}_i = (\underbrace{0, \ldots, 1}_{i \text{ entries}}, 0, \ldots, 0)$$

They are all linearly independent, and we call them the standard basis for $\mathbb{F}^n$ □

Example 3.17. Let E_{ij} be the $m \times n$ matrix with its ijth entry as 1 and 0 elsewhere. Then $\{E_{ij} \mid 1 \le i \le m, 1 \le j \le n\}$ is the standard basis for $M_{mn}(\mathbb{F})$. □

Example 3.18. The set $\{1, x, \ldots, x^n\}$ is the standard basis for $P_n(\mathbb{F})$. The set $\{1, x, x^2, \ldots\}$ is the standard basis for $P(\mathbb{F})$. □

Example 3.19. Since the empty set spans the zero vector space, it is also its basis. □

Theorem 3.8. *Consider the homogeneous linear system*

$$A\mathbf{x} = 0$$

where $A \in M_{mn}(\mathbb{F})$, If $m < n$, that is, there are more variables than equations, then the system will have non-trivial solutions.

Proof. If we imagine performing Gaussian elimination on such a matrix, then we will see why this is so. ■

Theorem 3.9. *Suppose that the vector space V is spanned by a finite set S. Then any subset L of V such that $|L| > |S|$ is linearly dependent.*

Proof. Say we have some $L = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, with $m > n$. Since S spans V, we can write

$$\begin{aligned} \mathbf{u}_1 &= a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n \\ \mathbf{u}_2 &= b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n \\ &\vdots \end{aligned}$$

We want to find $\alpha_1, \alpha_2, \ldots$ not all zero that satisfies

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \ldots = \mathbf{0}$$

Substituting,

$$\alpha_1(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) + \alpha_2(b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n) + \cdots = \mathbf{0}$$

We can consider the linear system

$$\begin{cases} a_1\alpha_1 + b_1\alpha_2 + \cdots = 0 \\ a_2\alpha_1 + b_2\alpha_2 + \cdots = 0 \\ \qquad \vdots \end{cases}$$

There are more variables than equations. By Thm. 3.8, there exists a non-trivial solution, and therefore L is linearly dependent. ■

Corollary 3.9.1. *If a vector space V is spanned by a finite set S and L is a linearly independent subset of V, then $|L| \leq |S|$.*

Theorem 3.10 (Dimension theorem)**.** *If V is a finite dimensional vector space, then every basis of V is finite and as the same number of elements.*

Proof. Since V is finite dimensional vector space, it has a finite basis B.

Let B' be another basis of V. By Cor. 3.9.1, B is a spanning set, and B' is linearly independent. So $|B| \geq |B'|$. But we can also reverse their roles and say $|B'| \geq |B|$. Hence $|B| = |B'|$. ■

Definition 3.9. Let V be a finite dimensional vector space. The *dimension* of V, denoted by

$$\dim V$$

is the number of elements of any basis of V. □

Example 3.20. $\dim \mathbb{F}^n = n$ □

Theorem 3.11. *Suppose* $\dim V = n$*.*

i. Any subset S of V which contains more than n elements is linearly dependent.

ii. No subset of V with less than n elements span V.

Proof. Let B be a basis of V. Then $|B| = n$.

i. By Thm. 3.9.

ii. By Cor. 3.9.1.

■

Theorem 3.12. *Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a linearly independent subset of a vector space V. If $\mathbf{w} \in V$ such that $\mathbf{w} \notin \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, then the set $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}\}$ is linearly independent.*

Proof. Suppose not. Suppose that $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}\}$ is instead linearly dependent. Then

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n + a_{n+1}\mathbf{w} = 0$$

has non-trivial solutions. $a_{n+1} \neq 0$ since $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ are linearly independent, and $a_{n+1} = 0$ would suggest otherwise. Then we have

$$\mathbf{w} = -\frac{a_1}{a_{n+1}}\mathbf{v}_1 - \cdots - \frac{a_n}{a_{n+1}}\mathbf{v}_n$$

which means $\mathbf{w} \in \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, contradiction. ■

Corollary 3.12.1. *Suppose* $\dim V = n$ *and $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a set of vectors in V.*

i. If S is linearly independent, then it is a basis for V.

ii. If S spans V, then S is a basis for V.

Proof.

i. Assume S is linearly independent but it does not span V. By Thm. 3.12, we can add some vector $\mathbf{w} \in V$ into S. However now $|S \cup \{\mathbf{w}\}| = n + 1 > n$, which contradicts Thm. 3.11.

ii. Assume S spans V but S is linearly dependent. By Thm. 3.6, we can remove a vector from S, but then then V will be spanned by $n - 1 < n$ vectors, contradicting Thm. 3.11.

■

Corollary 3.12.2. *Let V be a finite n-dimensional vector space and $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ is a linearly independent subset of V where $m < n$. There exists $n - m$ vectors $\{\mathbf{w}_1, \dots, \mathbf{w}_m\} \in V$ such that $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}_1, \dots, \mathbf{w}_m\}$ is a basis for V.*

Proof. We can iteratively perform Thm. 3.12 $(m - n)$ times, and from Cor. 3.12.1 this new set is a basis. ■

Theorem 3.13. *Let V be a finite dimensional vector space and W_1, W_2 be two subspaces of V. Then*

$$\dim W_1 + \dim W_2 = \dim(W_1 + W_2) + \dim(W_1 \cap W_2).$$

Proof. The proof is simple, but long. We provide a sketch here.

Let $\{\mathbf{a}_1, \dots, \mathbf{a}_r\}$ be a basis for $W_1 \cap W_2$. Then we can add more vectors using Thm. 3.12 to form a basis $\mathcal{B}_1$ for W_1 and a basis $\mathcal{B}_2$ for W_2. It can then be shown that $\mathcal{B}_1 \cup \mathcal{B}_2$ forms a basis for $W_1 + W_2$. ■

Theorem 3.14. *Let V be a vector space and let $\mathcal{B}_1 = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ and $\mathcal{B}_2 = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be two bases for V. There exists $\mathbf{v}_j \in \mathcal{B}_2$ such that $\{\mathbf{v}_j, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is again a basis of V.*

Proof. Let $W = \operatorname{span}\{\mathbf{u}_2, \dots, \mathbf{u}_n\}$. $\mathcal{B}_2 \nsubseteq W$, because otherwise $V = \operatorname{span} \mathcal{B}_2 \subseteq W$. Then by Thm. 3.12, there exists some $\mathbf{v}_j \in \mathcal{B}_2$ such that $\{\mathbf{v}_j, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is a linearly independent set with $n = \dim V$ vectors, hence it is also a basis by Cor. 3.12.1. ■

3.6 Direct sums of subspaces

If we have subspaces W_1, W_2 in V, $W_1 + W_2$ forms another subspace in V. We ask: how many ways are there to write $\mathbf{v} \in W_1 + W_2$ in the form of $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2$, with $\mathbf{w}_1 \in W_1$ and $\mathbf{w}_2 \in W_2$.

Example 3.21. Take W_1 as the xy-plane and W_2 as the yz-plane in $\mathbb{R}^3$. There is more than one way to write $(1, 2, 3)$ as a sum $\mathbf{w}_1 + \mathbf{w}_2$ with $\mathbf{w}_1 \in W_1$ and $\mathbf{w}_2 \in W_2$. □

Definition 3.10. $W_1 + W_2$ is the *direct sum* of W_1 and W_2, as every vector $\mathbf{v} \in W_1 + W_2$ can be expressed uniquely as the sum $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2$, with $\mathbf{w}_1 \in W_1$ and $\mathbf{w}_2 \in W_2$. We denote it as $W_1 \oplus W_2$. □

Example 3.22. The previous example with the xy and yz planes in $\mathbb{R}^3$ is not a direct sum. However, the xy plane and the z axis can form a direct sum that equals $\mathbb{R}^3$ itself. □

Theorem 3.15. *The subspace $W_1 + W_2$ is a direct sum of W_1 and W_2 iff $W_1 \cap W_2 = \{\mathbf{0}\}$.*

Proof.

($\implies$) Take $\mathbf{w}_1 \in W_1$, $\mathbf{w}_2 \in W_2$, $\mathbf{w} \in W_1 \cap W_2$. We can write

$$\mathbf{w}_1 + \mathbf{w}_2 = (\mathbf{w}_1 + \mathbf{w}) + (\mathbf{w}_2 - \mathbf{w})$$

Now $(\mathbf{w}_1 + \mathbf{w}) \in W_1$ and $(\mathbf{w}_2 - \mathbf{w}) \in W_2$. If $W_1 + W_2$ is a direct sum, then $\mathbf{w}_1 = (\mathbf{w}_1 + \mathbf{w})$ and $\mathbf{w}_2 = (\mathbf{w}_2 + \mathbf{w})$. Then $\mathbf{w} = \mathbf{0}$.

($\impliedby$) If $W_1 \cap W_2 = \{\mathbf{0}\}$, and there are two ways of writing the same vector $\mathbf{v} \in W_1 + W_2$:

$$\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2 = \mathbf{w}_1' + \mathbf{w}_2'$$

with $\mathbf{w}_1, \mathbf{w}_1' \in W_1$ and $\mathbf{w}_2, \mathbf{w}_2' \in W_2$.

Then we can also write

$$\mathbf{w}_1 - \mathbf{w}_1' = \mathbf{w}_2' - \mathbf{w}_2$$

Then $(w_1 - \mathbf{w}_1') \in W_1$ and $(\mathbf{w}_2' - \mathbf{w}_2) \in W_2$ so they are both in $W_1 \cap W_2$, which means that they are equal to $\mathbf{0}$. Therefore $\mathbf{w}_1 = \mathbf{w}_1'$ and $\mathbf{w}_2 = \mathbf{w}_2'$ and the sum of W_1 and W_2 is direct. ■

Definition 3.11. Let $W_1, \dots, W_k$ be subspaces in V. The sum of $W_1, \dots, W_k$ is the subspace

$$W_1 + \cdots + W_k = \{\mathbf{w}_1 + \cdots + \mathbf{w}_k \mid \forall 1 \le i \le k, \mathbf{w}_i \in w_i\}$$

□

Definition 3.12. We say the subspace $W_1 + \cdots + W_k$ is the direct sum of $W_1, \dots, W_k$ if every vector $\mathbf{v}$ in $W_1 + \cdots + W_k$ can be expressed uniquely as

$$\mathbf{v} = \mathbf{w}_1 + \cdots + \mathbf{w}_k$$

with $\mathbf{w}_i \in W_i$, $1 \le i \le k$. We write the direct sum as

$$W_1 \oplus \cdots \oplus W_k$$

□

We see that the left hand side here is actually equal to the span of the unions of all the sets:

$$W_1 + \cdots + W_k = \operatorname{span}(W_1 \cup \cdots \cup W_k)$$

Hence when a sum is a direct sum, there is a notion of linear independence between the subspaces.

Example 3.23. $\mathbb{R}^3$ is the direct sum of the x, y, and z axes. □

Theorem 3.16. *Let $W_1, \dots, W_k$ be subspaces of the vector space V and $W = W_1 + \cdots + W_k$. Then the following are equivalent:*

i. $W = W_1 \oplus \cdots \oplus W_k$

ii. *For* $2 \le j \le k$, $W_k \cap (W_1 + \cdots + W_{j-1}) = \{\mathbf{0}\}$

iii. If $\mathcal{B}_i$ is a basis for W_i for $1 \leq i \leq k$ and $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$ for all $i \neq j$, then $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$ is a basis for W.

Proof.

($(i) \implies (ii)$) Assume $W = W_1 \oplus \cdots \oplus W_k$. Let $2 \leq j \leq k$ and $\mathbf{v} \in W_j \cup (W_1 + \cdots + W_{j-1})$. Then $\mathbf{v} \in W_j$ and $\mathbf{v} \in W_1 + \cdots + W_{j-1}$.

We can write $\mathbf{v} = \mathbf{w}_1 + \cdots + \mathbf{w}_{j-1}$ for some $\mathbf{w}_1 \in W_1, \mathbf{w}_2 \in W_2, \ldots, \mathbf{w}_{j-1} \in W_{j-1}$. We claim that $\mathbf{v}$ has 2 expressions.

$$\begin{aligned} \mathbf{v} &= \underbrace{\mathbf{0} + \cdots + \mathbf{0} + \mathbf{v}}_{j\text{-th}} + \mathbf{0} + \cdots + \mathbf{0} \\ &= \mathbf{w}_1 + \cdots + \mathbf{w}_{j-1} + \mathbf{0} + \cdots + \mathbf{0} \end{aligned}$$

But since the expression for $\mathbf{v}$ is unique, we get that $\mathbf{w}_1 = \mathbf{0}, \ldots, \mathbf{w}_{j-1} = \mathbf{0}, \mathbf{v} = \mathbf{0}$. Therefore $W_j \cap (W_1 + \cdots + W_{j-1}) = \{\mathbf{0}\}$.

($(ii) \implies (i)$) Assume $W_j \cap (W_1 + \cdots + W_{j-1}) = \{\mathbf{0}\}$ for $2 \leq j \leq k$. Let $\mathbf{w} \in (W_1 + \cdots + W_k)$. Suppose $\mathbf{w} = \mathbf{w}_1 + \cdots + \mathbf{w}_k = \mathbf{w}'_1 + \cdots + \mathbf{w}'_k$, with $\mathbf{w}_1, \mathbf{w}'_1 \in W_1, \ldots, \mathbf{w}_k, \mathbf{w}'_k \in W_k$. Then

$$\overbrace{\underbrace{(\mathbf{w}_1 - \mathbf{w}'_1)}_{\in W_1} + \cdots + \underbrace{(\mathbf{w}_{k-1} - \mathbf{w}'_{k-1})}_{\in W_{k-1}}}^{\in W_1 + \cdots + W_{k-1}} = \underbrace{(\mathbf{w}'_k - \mathbf{w}_k)}_{\in W_k} \in W_k \cap (W_1 + \cdots + W_{k-1}) = \{\mathbf{0}\}$$

We get $\mathbf{w}'_k = \mathbf{w}_k$. We can apply this repeatedly to show that all $\mathbf{w}_i = \mathbf{w}'_i$, and so $W_1 + \cdots + W_k$ is a direct sum.

($(i) \implies (iii)$) $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$ spans W. We only need to show that it is linearly independent. Suppose not. Let $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$. Then

$$\alpha_1 \mathbf{w}_1 + \cdots + \alpha_n \mathbf{w}_n = \mathbf{0}$$

has non-trivial solutions. But this means that there are multiple ways of writing the zero vector in W, which contradicts the fact that W is created by direct sums. Hence $\mathcal{B}$ is a basis of V.

($(iii) \implies (ii)$) Let $\mathcal{B}_k = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$. $W_1 + \cdots + W_{j-1} = \text{span}\{\mathcal{B}_1 \cup \cdots \cup \mathcal{B}_{j-1}\}$, so let $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_{j-1} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.

Take $\mathbf{w} \in W_k \cap (W_1 + \cdots + W_{j-1})$ for any $2 \leq j \leq k$. Then we can write

$$\alpha_1 \mathbf{u}_1 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{w} = \beta_1 \mathbf{v}_1 + \cdots + \beta_n \mathbf{v}_n$$

for some $\alpha, \ldots, \alpha_n, \beta_1, \ldots, \beta_n \in \mathbb{F}$.

Rearranging, we get

$$\alpha_1 \mathbf{u}_1 + \cdots + \alpha_m \mathbf{u}_m - \beta_1 \mathbf{v}_1 - \cdots - \beta_n \mathbf{v}_n = \mathbf{0}$$

The vectors $\{u_1, \ldots, u_m, v_1, \ldots, v_n\} = W_k \cup \mathcal{B}$ are all unique and linearly independent, since they are mutually disjoint and form a basis together. Hence only the trivial solution exists, and $\mathbf{w} = \mathbf{0}$. Therefore $W_k \cap (W_1 + \ldots + W_{j-1}) = \{\mathbf{0}\}$. ■

In a way, direct sums are a way to break up and organize larger vector spaces into "components". We can break them up into smaller pieces that have minimal interaction with one another.

4 Linear transformations

4.1 Linear transformations

Definition 4.1. Let V, W be vector spaces over $\mathbb{F}$. A function $T : V \to W$ is called a *linear transformation* if:

i. $\forall \mathbf{v}_1, \mathbf{v}_2 \in V, T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$

ii. $\forall c \in \mathbb{F}, \mathbf{v} \in V, T(c\mathbf{v}) = cT(\mathbf{v})$.

More concisely,

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2)$$

□

Roughly speaking linear transformations respects the vector space operations. Note that above V and W are defined over the same field. That is the only constraint imposed on the choice of V and W.

Definition 4.2. A linear transformation $T : V \to V$ is called a *linear operator* on V. A linear transformation $T : V \to \mathbb{F}$ is called a *linear functional* on V. □

Example 4.1. Define $T : \mathbb{F}^n \to \mathbb{F}^m$ by $T(\mathbf{x}) = A\mathbf{x}$, for some $m \times n$ matrix A.

$$T\left[\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right] = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}$$

Using the rules for matrix multiplication it is easy to see that T is a linear transformation. In fact, every linear transformation $T : \mathbb{F}^n \to \mathbb{F}^m$ must be of this form. □

Theorem 4.1. *Let $\mathbb{F}$ be a field and let $T : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation. Then there exists an unique $A \in M_m(\mathbb{F})$ such that*

$$T(\mathbf{u}) = A\mathbf{u}$$

for every $\mathbf{u} \in \mathbb{F}^n$.

Proof. First we show existence. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis for $\mathbb{F}^n$. Then for any $\mathbf{u} \in \mathbb{F}^n$, we can express it in terms of the standard basis:

$$\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = u_1\mathbf{e}_1 + \cdots + u_n\mathbf{e}_n$$

And therefore

$$T(\mathbf{u}) = u_1T(\mathbf{e}_1) + \cdots + u_nT(\mathbf{e}_n)$$

Consider the following matrix with its columns as such

$$A = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

It can be easily verified that this matrix is what we are looking for.

Suppose there are matrices A, B such that $T(\mathbf{u}) = A\mathbf{u} = B\mathbf{u}$. Then more specifically,

$$T(\mathbf{e}_1) = A\mathbf{e}_1 = B\mathbf{e}_1, \ldots, T(\mathbf{e}_n) = A\mathbf{e}_n = B\mathbf{e}_n$$

Then A and B have the same columns. Therefore A is unique. ■

Example 4.2. Let the zero transformation be $T_0 : V \to W$ be $T_0(\mathbf{v}) = \mathbf{0}$, $\forall \mathbf{v} \in V$. □

Example 4.3. Let the identity operator be $I_V : V \to V$ be $I_V(\mathbf{v}) = \mathbf{v}$, $\forall \mathbf{v} \in V$. □

Example 4.4. Let $C(\mathbb{R})$ denote the set of all continuous real functions, and $C^1(\mathbb{R})$ be the set of all continuously differentiable functions. They are both real vector spaces, as we have shown before.

Define $D : C^1(\mathbb{R}) \to C(R)$ as $D(f) = \frac{\mathrm{d}f}{\mathrm{d}x}$, $\forall f \in C^1(\mathbb{R})$. Using the rules for derivatives, we can see that D is a linear transformation.

We can also have another transformation $T : C(\mathbb{R}) \to C^1(\mathbb{R})$, given by $T(f) = \int_0^x f(t)\,\mathrm{d}t$. Using the rules for integrals, we can see that T is also a linear transformation. □

Below are a few observations. They are fairly straightforward so the proofs have been left out.

Theorem 4.2. *A linear transformation can be completely determined by its image of basis basis vectors.*

Theorem 4.3. *A linear transformation sends the zero vector to the zero vector.*

Theorem 4.4. *If*

i. $\{\mathbf{v}_i, \ldots, \mathbf{v}_n\}$ *is a basis for* V*, and*

ii. $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ *is* any *set of vectors in* W*,*

then there is exactly one *linear transformation* $T : V \to W$*, with the property that*

$$T(\mathbf{v}_i) = \mathbf{w}_i \qquad i = 1, 2, \ldots, n$$

Proof. For $\mathbf{v} \in V$, we can write it as a linear combination of the basis $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. Then

$$\begin{aligned} T(\mathbf{v}) &= c_1 T(\mathbf{v}_1) + \cdots + c_n T(\mathbf{v}_n) \\ &= c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n \end{aligned}$$

■

4.2 Range and Kernel

Definition 4.3. Let $T : V \to W$ be a linear transformation.

i. The *kernel* of T is the subset $\{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}\}$. They are all the vectors being sent to the zero vector. We denote it as $\ker(T)$.

ii. The *range* of T is the subset $\{T(\mathbf{v}) \mid \mathbf{v} \in V\}$ of W. It is the set of all images of T. We denote it as $\mathcal{R}(T)$. □

$\mathbf{0} \in \ker(T)$ always. Also, T is surjective if $\mathcal{R}(T) = W$. T is injective if $\forall \mathbf{v}_1, \mathbf{v}_2 \in V, T(\mathbf{v}_1) = T(\mathbf{v}_2) \implies \mathbf{v}_1 = \mathbf{v}_2$. In other places, injective linear transformations may be called *nonsingular* linear transformations.

Lemma 4.5. *Let $T : V \to W$ be a linear transformation. Then T is injective iff $\ker(T) = \{\mathbf{0}\}$.*

Proof.

($\implies$) Assume T is injective. Take $\mathbf{v} \in \ker(T)$. Then $T(\mathbf{v}) = \mathbf{0} = T(\mathbf{0})$. Since T is injective, $\mathbf{v} = \mathbf{0}$.

($\impliedby$) Assume $\ker(T) = \{\mathbf{0}\}$. Take any $\mathbf{v}_1, \mathbf{v}_2 \in V$. $T(\mathbf{v}_1) = T(\mathbf{v}_2)$ implies that $T(\mathbf{v}_1) - T(\mathbf{v}_2) = \mathbf{0}$ since T is linear. This implies that $\mathbf{v}_1 - \mathbf{v}_2 \in \ker(T) = \{\mathbf{0}\}$. Therefore $\mathbf{v}_1 = \mathbf{v}_2$. ■

Theorem 4.6. *Let $T : V \to W$ be an injective linear transformation. If $\{\mathbf{v}, \dots, \mathbf{v}_n\}$ is a basis for V, then $B = \{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$ is a basis for $\mathcal{R}(T)$.*

Proof. First we show that B spans $\mathcal{R}(T)$. It is quite clear that $\operatorname{span} B \subseteq \mathcal{R}(T)$. On the other hand, for any $\mathbf{u} \in \mathcal{R}(T)$, there exists some $\mathbf{w} \in V$ such that $T(\mathbf{w}) = \mathbf{v}$. It follows that since $\mathbf{w}$ can be expressed as $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$, therefore $T(\mathbf{w}) = a_1T(\mathbf{v}_1) + \cdots + a_nT(\mathbf{v}_n) \in \operatorname{span} B$. This shows that $\mathcal{R}(T) \subseteq \operatorname{span} B$, and therefore $\mathcal{R}(T) = \operatorname{span} B$.

Now we show that B is linearly independent. Consider the equation

$$\begin{aligned} a_1T(\mathbf{v}_1) + \cdots + a_nT(\mathbf{v}_n) &= \mathbf{0} \\ T(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) &= \mathbf{0} = T(\mathbf{0}) \end{aligned}$$

Since T is injective this means that $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$. But as $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent, we get that $a_1 = \cdots = a_n = 0$ and hence B is linearly independent. ■

Theorem 4.7. *Let $T : V \to W$ be a linear transformation. Then*

i. $\ker(T)$ *is a subspace of V, and*

ii. $\mathcal{R}(T)$ *is a subspace of W.*

Proof.

i. $\ker(T)$ is non-empty since $\mathbf{0} \in \ker(T)$ at least. Also by definition $\ker(T) \subseteq V$. Now take any $\mathbf{u}, \mathbf{v} \in \ker(T)$. Then for any scalars α, β, $T(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v}) = \mathbf{0}$ so $\alpha\mathbf{u} + \beta\mathbf{v} \in \ker(T)$, and it is a subspace.

ii. Now $\mathbf{0} \in \mathcal{R}(T)$, and $\mathcal{R}(T) \subseteq W$ by definition. Taking any $\mathbf{u}, \mathbf{v} \in \mathcal{R}(T)$, there must exist some $\mathbf{u}', \mathbf{v}' \in V$ such that $T(\mathbf{u}') = \mathbf{u}$ and $T(\mathbf{v}') = \mathbf{v}$. Hence $\alpha\mathbf{u} + \beta\mathbf{v} = T(\alpha\mathbf{u}' + \beta\mathbf{v}')$, and since $\alpha\mathbf{u}' + \beta\mathbf{v}' \in V$, $\alpha\mathbf{u} + \beta\mathbf{v} \in \mathcal{R}(T)$ and so it is also a subspace.

■

Theorem 4.8. *Let A be an $m \times n$ matrix over $\mathbb{F}$, and let $T : \mathbb{F}^n \to \mathbb{F}^m$. Consider $T(\mathbf{u}) = A\mathbf{u}, \forall \mathbf{u} \in \mathbb{F}^n$. Some observations:*

i. $\ker(T)$ *is the solution space of the system* $A\mathbf{x} = \mathbf{0}$.

ii. $\mathcal{R}(T)$ *is the column space of* A.

iii. Every subspace of $\mathbb{R}^n$ *(or* $\mathbb{F}^n$*) is the solution space of a linear system* $A\mathbf{x} = \mathbf{0}$.

Proof.

i. $\mathbf{u} \in \ker(T) \iff A\mathbf{u} = T(\mathbf{u}) = 0 \iff \mathbf{u}$ solves $A\mathbf{x} = \mathbf{0}$.

ii. $\mathbf{v} \in \mathcal{R}(T) \iff \exists \mathbf{u} \in \mathbb{F}^n, A\mathbf{u} = T(\mathbf{u}) = \mathbf{v}$.

iii. This is claiming that every subspace of $\mathbb{F}^n$ is the kernel of some linear transformation $T : \mathbb{F}^n \to \mathbb{F}^n$.

   Call the subspace V. Let a basis for the subspace be $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$. We can extend it to be a basis for $\mathbb{F}^n$, $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{v}_{m+1}, \dots, \mathbf{v}_n\}$. Consider the linear transformation given by

$$T(\mathbf{v}_i) = \begin{cases} 0, & \text{if } 1 \leq i \leq m \\ v_i, & \text{otherwise} \end{cases}$$

   Then it is easy to see that in this case $T(\mathbf{u}) = \mathbf{0} \iff \mathbf{u} \in V$, i.e. $\ker(T) = V$.

■

Example 4.5. The map $T : P_2(\mathbb{R}) \to M_2(\mathbb{R})$ from the vector space of real polynomials of degree two $f(x) = ax^2 + bx + c$ to the vector space of real 2×2 matrices defined by

$$T(f(x)) = \begin{pmatrix} f(1) - f(2) & 0 \\ 0 & f(0) \end{pmatrix} = \begin{pmatrix} -b - 3c & 0 \\ 0 & a \end{pmatrix}$$

is a linear transformation.

Assume $f(x) \in \ker(T)$. Then $T(f(x)) = \mathbf{0} \in M_2(\mathbb{R})$. Then we have constraints

$$\begin{cases} -b - 3c = 0 \\ a = 0 \end{cases}$$

Then every $f(x) = c(x^2 - 3x)$. Then $\{x^2 - 3x\}$ forms a basis for $\ker(T)$. By Lemma 4.5, T is not injective.

$\mathcal{R}(T) = \{T(f(x)) \mid f(x) \in P_2(\mathbb{R})\} \ni T(f(x))$. We can go through the same process to find a basis for the range:

$$\mathcal{R}(T) = \text{span}\left\{\begin{pmatrix}1 & 0\\0 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\0 & 1\end{pmatrix}\right\}$$

□

Definition 4.4. Let $T : V \to W$ be a linear transformation. Assume V is finite dimensional. Then

i. The *rank* of T is the dimension of the range of T.

ii. The *nullity* of T is the dimension of the kernel of T.

□

$\text{rank}(T) = \dim W$ iff T is surjective. $\text{nullity}(T) = 0$ iff T is injective.

Theorem 4.9 (Rank Nullity Theorem)**.** *Let $T : V \to W$ be a linear transformation. Suppose V is finite dimensional. Then*

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

Proof. Let $n = \dim V$, $r = \text{nullity}(T)$ Let the basis for $\ker(T)$ be $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$. By Thm. 3.12, we can extend this to a basis for V. Let the new basis be $\{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots \mathbf{v}_n\}$. The vectors $\{\mathbf{v}_1, \ldots \mathbf{v}_r\}$ are mapped to $\mathbf{0} \in W$. We claim that the other vectors $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ are mapped to a basis for the range.

Take $\mathbf{w} \in \mathcal{R}(T)$. Then $\exists \mathbf{v} \in V, T(\mathbf{v}) = \mathbf{w}$. We can write $\mathbf{v} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$. Therefore $T(\mathbf{v}) = \alpha_1 T(\mathbf{v}_1) + \cdots + \alpha_r T(\mathbf{v}_r) + \ldots + \alpha_n T(\mathbf{v}_n)$. But the first r terms are all sent to zero. Therefore $\{T(\mathbf{v}_{r+1}), \ldots, T(\mathbf{v}_n)\}$ spans the range.

Now consider the homogeneous system

$$\beta_{r+1}T(\mathbf{v}_{r+1}) + \cdots + \beta_n T(\mathbf{v}_n) = \mathbf{0}$$

Since T is linear, we can write

$$T(\underbrace{\beta_{r+1}\mathbf{v}_{r+1} + \cdots + \beta_n\mathbf{v}_n}_{\in\ker(T)=\text{span}\{\mathbf{v}_1,\ldots,\mathbf{v}_r\}}) = \mathbf{0}$$

Then we can write $\beta_{r+1}\mathbf{v}_{r+1} + \cdots + \beta_n\mathbf{v}_n = \gamma_1\mathbf{v}_1 + \cdots + \gamma_r\mathbf{v}_r$, for some $\gamma_1, \ldots, \gamma_r \in \mathbb{F}$. We can shift them around, $-\gamma_1\mathbf{v}_1 - \cdots - \gamma_r\mathbf{v}_r + \beta_{r+1}\mathbf{v}_{r+1} + \cdots + \beta_n\mathbf{v}_n = \mathbf{0}$. But since all the $\mathbf{v}$'s form a basis, all the coefficients must be zero. Hence $\beta_{r+1} = \cdots = \beta_n = 0$, and $\{T(\mathbf{v}_{r+1}, \ldots, T(\mathbf{v}_n)\}$ is linearly independent. Hence it spans the range.

We get that the range is of dimension $n - r$ and the kernel is of dimension r. ■

Example 4.6. Referring back to Ex. 4.5, we can see that $\text{rank}(T) = 2$ and $\text{nullity}(T) = 1$, and $\dim P_2(\mathbb{R}) = 3$. □

Corollary 4.9.1. *If V and W are finite dimensional vector spaces, with*

$$\dim V = \dim W$$

and we have a linear transformation $T : V \to W$, then the following are equivalent:

i. T is injective

ii. T is surjective

iii. T is bijective

Proof.

($(i) \implies (ii)$) Let $n = \dim V = \dim W$. From Thm. 4.9, $\text{rank}(T) + \text{nullity}(T) = n$. Since T is injective, by Lemma 4.5, $\ker(T) = \{\mathbf{0}\}$ and therefore $\text{nullity}(T) = 0$. Hence $\text{rank}(T) = n = \dim W$. It follows that $W = \mathcal{R}(T)$, and so T is surjective.

($(ii) \impliedby (i)$) In this case, $\dim W = \text{rank}(T)$. Then with Thm. 4.9, we end up with $\text{nullity}(T) = 0$. Therefore T is injective.

The other equivalences then come from definition. ■

Very intuitively, we also have the following relation.

Corollary 4.9.2. *Let V and W be finite dimensional vector spaces and $T : V \to W$ a linear transformation. Then*

i. If $\dim V < \dim W$, then T is not surjective.

ii. If $\dim V > \dim W$, then T is not injective.

Proof. We use Thm. 4.9 throughout the proof for both parts.

i. $\dim V = \text{rank}(T) + \text{nullity}(T) \implies \text{rank}(T) \leq \dim V$. Then $\text{rank}(T) < \dim W$ so T cannot be surjective.

ii. Suppose not. Suppose T is injective, then $\text{nullity}(T) = 0$. Then

$$\dim V = \underbrace{\text{rank}(T)}_{\leq \dim W} + \underbrace{\text{nullity}(T)}_{0} \leq \dim W.$$

Contradiction.

■

Theorem 4.10 (Equality of Row Rank and Column Rank)**.** *Let $A \in M_{mn}(\mathbb{F})$. The rank of the row space of A (row rank) is equal to the rank of the column space of A (column rank).*

Proof. Let us denote the rows of the matrix A with $\mathbf{r}_1, \dots, \mathbf{r}_n$, and the columns with $\mathbf{c}_1, \dots, \mathbf{c}_m$. Let there be a $T : \mathbb{F}^n \to \mathbb{F}^m$ with $T(x) = A\mathbf{x}$. Then $\ker(T)$ is the solution space of $A\mathbf{x} = \mathbf{0}$. Let B be the reduced row echelon form of A. Then $\ker(T)$ is also the solution space of $B\mathbf{x} = \mathbf{0}$. Also, the row space (and hence the row rank) of A is the same as that of B. The row rank of B is determined by the number of non-zero rows of B. Call this value r. Then the dimension of the solution space is $n - r$.

By Thm.4.9, with $\text{rank}(T)$ being the column rank of A, and $\text{nullity}(T)$ being the dimension of the solution space, we see that the column rank is also r. ■

Due to this fact, we can simply call the row rank and column rank the *rank* of a matrix.

Suppose we have two linear transformations and we wish to create new ones from them. There are two obvious ways to do this.

Definition 4.5. Let $T_1, T_2 : V \to W$ be linear transformations. Define $T_1 + T_2 : V \to W$ by

$$(T_1 + T_2)(\mathbf{v}) = T_1(\mathbf{v}) + T_2(\mathbf{v})$$

□

Definition 4.6. Let $T : V \to W$ be linear transformation, $\alpha \in \mathbb{F}$. Define $\alpha T : V \to W$ by

$$(\alpha T)(\mathbf{v}) = \alpha T(\mathbf{v})$$

□

At first glance, it may seem that there is a hint of vector spaces in these two definitions. Indeed!

Theorem 4.11. *Let $\mathcal{L}(V, W)$ be the set of all linear transformations $T : V \to W$. Then $\mathcal{L}(V, W)$ forms a vector space over $\mathbb{F}$ with vector addition and scalar multiplication as defined in Def. 4.5 and Def. 4.6, and with the zero transformation as the zero vector.*

In addition, if V and W are finite dimensional, then

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

Proof. It is easy (perform the routine) to check that $\mathcal{L}(V, W)$ is indeed a vector space. This part is left out.

Let $\dim V = n$ and $\dim W = m$. Assume that $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis for V and $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ is a basis for W. For $1 \le i \le m$, $1 \le j \le n$, define $T_{ij} : V \to W$ as the linear transformation that maps $\mathbf{v}_j$ to $\mathbf{w}_i$ and everything else to $\mathbf{0}$.

Take the set $A = \{T_{ij} \mid 1 \le j \le m, 1 \le j \le n\}$. It spans $\mathcal{L}(V, W)$ since all linear transformations in $\mathcal{L}(V, W)$ can be defined by its images on the basis vectors of V, and the images can also be defined in terms of the basis vectors for W. A is also obviously linearly independent. Thus A forms a basis for $\mathcal{L}(V, W)$. Then $\dim \mathcal{L}(V, W) = \|A\| = mn = (\dim V)(\dim W)$. ■

Example 4.7. Consider $\mathcal{L}(V, \mathbb{F})$, the set of all linear functionals $f : V \to \mathbb{F}$. This is called the *dual space* of V and is denoted by V^*. By Thm. 4.11, $\dim V^* = (\dim V)(\dim \mathbb{F}) = \dim V$. The basis constructed in the way detailed in the proof of Thm. 4.11 is called the *dual basis* for the basis of V. □

Definition 4.7. Let $T : V \to W$ and $S : W \to U$ be linear transformations. Define the composite function

$$S \circ T : V \to U \qquad (S \circ T)(\mathbf{v}) = S[T(\mathbf{v})]$$

We may just write $S \circ T$ as ST. ☐

It can be shown that $ST : V \to U$ is also a linear transformation.

In a special case $S, T \in \mathcal{L}(V, V) \implies ST \in \mathcal{L}(V, V)$. Hence $\mathcal{L}(V, V)$ has more structure than $\mathcal{L}(V, W)$ since it has an additional operation of multiplication. We say that $\mathcal{L}(V, V)$ is an *algebra* over $\mathbb{F}$.

Definition 4.8. If $T \in \mathcal{L}(V, V)$, we will denote TT as T^2. We may define recursively $T^0 = \mathbf{I}$ and $T^n = TT^{n-1}$. ☐

Example 4.8. If $f(x) = a_0 + a_1x + \ldots + a_nx^n$ is a polynomial, we can consider $f(T)$, with T being a linear operator. Then $f(T) : V \to V$ is also a linear operator, since it is made up of the sum of linear operators. ☐

Example 4.9. If $T : P(\mathbb{R}) \to P(\mathbb{R})$ is given by the derivative

$$T(f) = \frac{\mathrm{d}f}{\mathrm{d}x}$$

Then in general

$$T^n(f) = \frac{\mathrm{d}^n f}{\mathrm{d}x^n}$$

☐

Example 4.10 (Application to Differential Equations)**.** Take the set $\mathcal{F}(\mathbb{R}, \mathbb{C})$, the set of all functions $f : \mathbb{R} \to \mathbb{C}$. Such functions are in the form of $f(t) = u(t) + iv(t)$. $C^\infty(\mathbb{R}, \mathbb{C}) \subset \mathcal{F}(\mathbb{R}, \mathbb{C})$. Let $D : C^\infty \to C^\infty$ be the operator $D(f) = \frac{\mathrm{d}f}{\mathrm{d}x} = u' + iv'$.

Consider the differential equation

$$y^{(n)} + a_{n-1}y^{n-1} + \cdots + a_1y' + a_0y = 0$$

with $a_0, a_1, \ldots a_{n-1} \in \mathbb{C}$. We can recast it in this form

$$(D^n + a_{n-1}D^{n-1} + \cdots + a_1D + a_0\mathbf{I})(y) = f(D)(y)$$

with $f(x)$ is a polynomial with complex coefficients of degree n, called the auxiliary polynomial of the differential equation. $f(D) : C^\infty \to C^\infty$ is a linear operator.

$y \in C^\infty$ is a solution to this equation iff $f(D)(y) = 0$ iff $y \in \ker f(D)$. There are a few more properties we will state but will not prove since they are out of scope.

i. It is not immediately clear that the kernel is finite dimensional, since it is a subspace of a infinite dimensional vector space. But it can be shown that $\dim \ker f(D) = n$.

ii. If $f(x)$ has n distinct zeroes, $c_1, \ldots, c_n$, then $\{e^{c_1t}, e^{c_2t}, \ldots, e^{c_nt}\}$ is a basis for $\ker f(D)$.

□

Theorem 4.12. *Let U and V be finite dimensional vector spaces over a field $\mathbb{F}$ and let W be a subspace of V. Suppose that $T : U \to V$ is a linear transformation and $S = \{\mathbf{u} \in U \mid T(\mathbf{u}) \in W\}$. Then*

$$\dim S \geq \dim U - \dim V + \dim W$$

Proof. S is a subspace. We skip the proof for this part.

We claim that $\ker(T) \subseteq S$. Let $\mathbf{u} \in \ker(T)$. Then $T(\mathbf{u}) = \mathbf{0} \in W$, and thus $\mathbf{u} \in S$.

Now pick a basis for $\ker(T)$, say $\{\mathbf{k}_1, \ldots, \mathbf{k}_n\}$, and extend it to a basis for S, $\{\mathbf{k}_1, \ldots, \mathbf{k}_n, \mathbf{u}_1, \ldots, \mathbf{u}_p\}$. Then we can extend it further to become a basis for U, $\{\mathbf{k}_1, \ldots, \mathbf{k}_n, \mathbf{u}_1, \ldots, \mathbf{u}_p, \mathbf{u}'_1, \ldots, \mathbf{u}'_q\}$. Now consider the action of T on this basis of U:

$$\{\mathbf{0}, \ldots, \mathbf{0}, \underbrace{T(\mathbf{u}_1), \ldots, T(\mathbf{u}_p)}_{\in W}, \underbrace{T(\mathbf{u}'_1), \ldots, T(\mathbf{u}'_q)}_{\notin W}\}$$

We also claim that $\{T(\mathbf{u}_1), \ldots, T(\mathbf{u}_p)\}$ is linearly independent. Consider the equation

$$\alpha_1 T(\mathbf{u}_1) + \cdots + \alpha_p T(\mathbf{u}_p) = T(\alpha_1 \mathbf{u}_1 + \cdots + \alpha_p \mathbf{u}_p) = \mathbf{0}.$$

Therefore $\alpha_1 \mathbf{u}_1 + \cdots + \alpha_p \mathbf{u}_p \in \ker(T) = \operatorname{span}\{\mathbf{k}_1, \ldots, \mathbf{k}_n\}$. Therefore

$$\alpha_1 \mathbf{u}_1 + \cdots \alpha_p \mathbf{u}_p = \beta_1 \mathbf{k}_1 + \cdots + \beta_n \mathbf{k}_n = \mathbf{0}.$$

We can rearrange this to show that indeed there is only the trivial solution.

Now we extend this to a basis for W, $\{T(\mathbf{u}_1), \ldots, T(\mathbf{u}_p), \mathbf{w}_1, \ldots, \mathbf{w}_l\}$. We have three groups of vectors that we can now put together, $\{T(\mathbf{u}_1), \ldots, T(\mathbf{u}_p), \mathbf{w}_1, \ldots, \mathbf{w}_l, T(\mathbf{u}'_1), \ldots, T(\mathbf{u}'_q)\}$. We claim that this is a linearly independent subset of V. Again, consider the equation

$$\underbrace{\alpha_1 T(\mathbf{u}_1) + \cdots + \alpha_p T(\mathbf{u}_p) + \beta_1 \mathbf{w}_1 + \cdots \beta_l \mathbf{w}_l}_{\mathbf{w} \in W} + \underbrace{\gamma_1 T(\mathbf{u}'_1) + \cdots + \gamma_q T(\mathbf{u}'_q)}_{T(\gamma_1 \mathbf{u}'_1 + \cdots + \gamma_q \mathbf{u}'_q)} = \mathbf{0}$$

Therefore $T(\gamma_1 \mathbf{u}'_1 + \cdots + \gamma_q \mathbf{u}'_q) = -\mathbf{w} \in W$. Then $\gamma_1 \mathbf{u}'_1 + \cdots + \gamma_q \mathbf{u}'_q \in S = \operatorname{span}\{\mathbf{k}_1, \ldots, \mathbf{k}_n, \mathbf{u}_1, \ldots, \mathbf{u}_p\}$. We can write

$$\gamma_1 \mathbf{u}'_1 + \cdots + \gamma_q \mathbf{u}'_q = \delta_1 \mathbf{k}_1 + \cdots + \delta_n \mathbf{k}_n + \varepsilon_1 \mathbf{u}_1 + \cdots + \varepsilon_p \mathbf{u}_p$$

Performing more arrangements, we will find that only the trivial solution exists.

The number of vectors in this big set, $p + l + q \leq \dim V$. But $\dim W = p + l$, $\dim S = n + p$, and $\dim U = n + p + q$. Substituting, we get that $\dim W + \dim U - \dim S \leq \dim V$. ■

Theorem 4.13. *Let V be a finite dimensional vector space and $T : V \to V$ and $S : V \to V$ be linear operators.*

i. $\ker(T) \subseteq \ker(ST)$.

ii. $\mathcal{R}(ST) \subseteq \mathcal{R}(S)$.

iii. Let $n_T = \text{nullity}(T)$, $n_S = \text{nullity}(S)$ and $n_{ST} = \text{nullity}(ST)$. Then

$$\max(n_S, n_T) \leq n_{ST} \leq n_S + n_T$$

Proof.

Take any $\mathbf{v} \in \ker(T)$. Then $S(T(\mathbf{v})) = S(\mathbf{0}) = \mathbf{0}$. Thus $\mathbf{v} \in \ker(ST)$.

ii.

$$\begin{aligned}\mathcal{R}(S) &= \{S(\mathbf{v}) \mid \mathbf{v} \in V\} \\ \mathcal{R}(ST) &= \{ST(\mathbf{v}) \mid \mathbf{v} \in V\} \\ &= \{S(\mathbf{v}) \mid \mathbf{v} \in \mathcal{R}(T)\}\end{aligned}$$

Since $\mathcal{R}(T) \subseteq V$, it follows that $\mathcal{R}(ST) \subseteq \mathcal{R}(S)$.

iii. From the above two points, $n_T \leq n_{ST}$, and $\text{rank}(ST) \leq \text{rank}(S)$. Using Thm. 4.9,

$$n_S = \dim V - \text{rank}(S) \leq \dim V - \text{rank}(ST) = n_{ST}$$

The other part of the inequality is more tedious. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ be a basis for $\ker(T)$. Note that this means $r = n_T$. We can extend this to be a basis of $\ker(ST)$, $\{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{v}_1, \ldots, \mathbf{v}_s\}$. Now $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_s) \in \ker(S)$. We can verify that $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_s)\}$ is linearly independent, and this part is skipped. This tells us that $s \leq n_S$. Hence,

$$n_{ST} = r + s \leq n_T + n_S$$

■

4.3 Isomorphisms

Definition 4.9. We say $T : V \to W$ is *invertible* if there exists a function $S : W \to V$

$$S \circ T = \mathbf{I}_V : V \to V \qquad T \circ S = \mathbf{I}_W : W \to W$$

in other words,

$$\forall \mathbf{v} \in V, \mathbf{w} \in W, \qquad S[T(\mathbf{v})] = \mathbf{v} \qquad T[S(\mathbf{w})] = \mathbf{w}$$

We write $S = T^{-1}$. □

Also note that T is invertible iff it is bijective.

Lemma 4.14. *If $T : V \to W$ is an invertible linear transformation, then T^{-1} is also a linear transformation.*

Proof. T is surjective. Suppose $T(\mathbf{v}_1) = \mathbf{w}_1$, and $T(\mathbf{v}_2) = \mathbf{w}_2$. Then $T^{-1}(\mathbf{w}_1) = \mathbf{v}_1$ and $T^{-1}(\mathbf{w}_2) = \mathbf{v}_2$.

Since T is linear, we have

$$T(\alpha\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1 T(\mathbf{v}_1) + \alpha_2 T(\mathbf{v}_2) = \alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2$$

By the definition of the inverse,

$$T^{-1}(\alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2) = \alpha_1 T^{-1}(\mathbf{w}_1) + \alpha_2 T^{-1}(\mathbf{w}_2)$$

■

Definition 4.10. An invertible linear transformation $T : V \to W$ is called an *isomorphism* of V onto W. □

So for a linear transformation T, T is bijective iff T is invertible iff T is an isomorphism. Furthermore, by Lemma 4.14, T is an isomorphism iff T^{-1} is an isomorphism.

Definition 4.11. We say that the vector spaces V and W are *isomorphic* if there is an isomorphism $T : V \to W$. We write

$$V \cong W$$

□

Isomorphism means that two vector spaces may be different on the surface, but have identical structures.

Theorem 4.15. *If V and W are finite dimensional vector spaces over the same field, then*

$$V \cong W \iff \dim V = \dim W$$

Proof.

($\Longrightarrow$) Cor. 4.9.2.

($\Longleftarrow$). Assume $\dim V = \dim W = n$. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for V and $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be a basis for W. Let $T : V \to W$ be the linear transformation such that

$$T(\mathbf{v}_i) = \mathbf{w}_i.$$

It can be checked that $W = \text{span}\{\mathbf{w}_1, \ldots, \mathbf{w}_n\} = \mathcal{R}(T)$. So T is surjective, and by Cor. 4.9.1, T is bijective and hence is an isomorphism. ■

Example 4.11. Here are some isomorphisms:

- If $\dim V = n$ then $V \cong \mathbb{F}^n$
- $M_2(\mathbb{R}) \cong \mathbb{R}^4$
- $P_2(\mathbb{C}) \cong \mathbb{C}^3$
- $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \cong M_{mn}(\mathbb{R})$.

□

4.4 Coordinates

From the previous section we have learned that a vector space V with dimension n has the same structure as $\mathbb{F}^n$. This section will aim to describe an explicit way to connect these two spaces together.

Definition 4.12. An *ordered basis* of a vector space V is a basis of V with a specific ordering. □

We will write it the same way as a normal basis, but it will be made clear when a basis is ordered. The purpose of this ordering can be seen in the next part.

For some ordered basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V, then for any $\mathbf{v} \in V$,

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n$$

with unique $\alpha_1, \ldots, \alpha_n$. Let

$$[\mathbf{v}]_{\mathcal{B}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

If the basis was not ordered then there would be ambiguity in the writing of this matrix.

Definition 4.13. The column matrix $[\mathbf{v}]_{\mathcal{B}}$ is called the *coordinate matrix of* $\mathbf{v}$ *relative to the ordered basis* $\mathcal{B}$. □

Example 4.12. Take $\mathbb{R}^3$ and its ordered basis $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, the standard basis. Then for any $\mathbf{v} = a\mathbf{e}_1 + b\mathbf{e}_2 + c\mathbf{e}_3 \in \mathbb{R}^3$:

$$[\mathbf{v}]_{\mathcal{B}} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{v}$$

It seems like we have done nothing. In fact, if $\mathcal{B}$ is the standard basis for $\mathbb{F}^n$, then $[\mathbf{v}]_{\mathcal{B}} = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$. □

Actually, the association of a vector to its coordinate matrix is an isomorphism.

Theorem 4.16. *Every n-dimensional vector space V over the field $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$.*

Proof. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an ordered basis for V. Define a transformation $T : V \to \mathbb{F}^n$ such that

$$\forall \mathbf{v} \in V, T(\mathbf{v}) = [\mathbf{v}]_{\mathcal{B}}$$

Firstly, T is linear. The following can be easily checked:

$$T(\alpha\mathbf{u} + \beta\mathbf{v}) = [\alpha\mathbf{u} + \beta\mathbf{v}]_{\mathcal{B}} = \alpha[\mathbf{u}]_{\mathcal{B}} + \beta[\mathbf{v}]_{\mathcal{B}} = \alpha T(\mathbf{u}) + \beta T(\mathbf{v})$$

Next, T is injective. Let $\mathbf{u} \in \ker(T)$. Then

$$[\mathbf{u}]_{\mathcal{B}} = T(\mathbf{u}) = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{F}^n$$

So $\ker(T) = \{\mathbf{0}\}$ and therefore T is injective. Since $\dim V = \dim \mathbb{F}^n$, T is bijective by Cor. 4.9.1. ■

4.5 Representation with matrices

Thm. 4.16 tells us that vectors in a finite dimensional vector space behaves like column matrices. We like a space to look like column matrices since they have a simple structure and are easier to manipulate.

Let $T : V \to W$ be a linear transformation, and $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an ordered basis for V, and $\mathcal{B}' = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be an ordered basis for W. For any $\mathbf{v}_j \in V$ and $1 \le j \le n$,

$$T(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i$$

Let $\mathbf{v} \in V$. $\mathbf{v} = \sum_{j=1}^{n} \alpha_j \mathbf{v}_j$. We can them obtain

$$\begin{aligned} T(\mathbf{v}) &= \sum_{j=1}^{n} \alpha_j T(\mathbf{v}_j) \\ &= \sum_{j=1}^{n} \alpha_j \sum_{i=1}^{m} a_{ij}\mathbf{w}_i \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}\alpha_j \mathbf{w}_i \end{aligned}$$

Then

$$\begin{aligned} [T(\mathbf{v})]_{\mathcal{B}'} &= \begin{pmatrix} \sum_{j=1}^{n} a_{1j}\alpha_j \\ \sum_{j=1}^{n} a_{2j}\alpha_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}\alpha_j \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \end{aligned}$$

Then we see that $[T(\mathbf{v})]_{\mathcal{B}'} = A[\mathbf{v}]_{\mathcal{B}}$, with A being the first matrix on the right hand side.

Definition 4.14. The matrix we called A in the exposition above is called the *matrix of T relative to the ordered bases $\mathcal{B}$ and $\mathcal{B}'$*. We write

$$A = [T]_{\mathcal{B}',\mathcal{B}}$$

Then $[T(\mathbf{v})]_{\mathcal{B}'} = [T]_{\mathcal{B}',\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}$.

If $V = W$ and $\mathcal{B} = \mathcal{B}'$ then we just call $[T]_{\mathcal{B}',\mathcal{B}} = [T]_{\mathcal{B}}$. □

We also have to keep in mind that in the definition above, $\mathcal{B}'$ is for the co-domain and $\mathcal{B}$ is for the domain. A helpful mnemonic is to keep the same symbols together, for example above we keep the $\mathcal{B}$ facing each other.

The way to understand the long calculation above is as follows. For two finite dimensional vector spaces V, W, they both have isomorphisms to some $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively. Now if we have a $T : V \to W$, for any $\mathbf{v} \in V$, $[\mathbf{v}]_{\mathcal{B}} \in F^n$ and $[T(\mathbf{v})]_{\mathcal{B}'} \in \mathbb{F}^m$. Then what we have done is find some linear map A that sends $[\mathbf{v}]_{\mathcal{B}}$ to $[T(\mathbf{v})]_{\mathcal{B}'}$. It transfers the computations in the general vector spaces to $\mathbb{F}^n$ and $\mathbb{F}^m$.

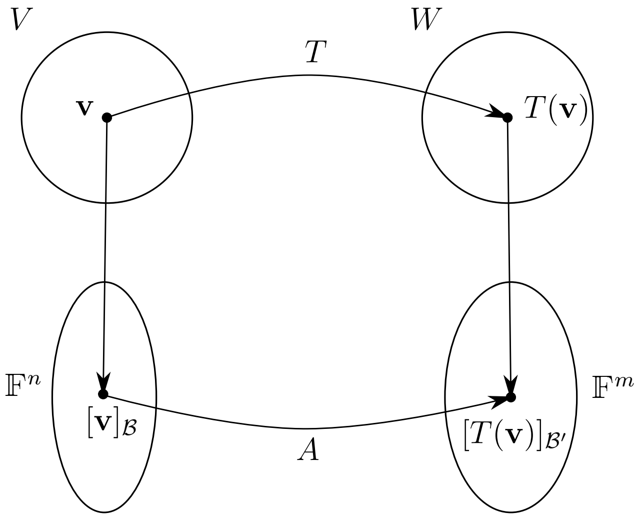

Figure 1: An illustration of the effect of A.

Now of course we want to know how to construct A. For $1 \leq j \leq n$, we have $T(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i$, so

$$[T(\mathbf{v}_j)]_{\mathcal{B}'} = \begin{pmatrix} a_{ij} \\ \vdots \\ a_{mj} \end{pmatrix}$$

Then looking at the definition of A above, we have

$$A = [T]_{\mathcal{B}',\mathcal{B}} = \left(\ [T(\mathbf{v}_1)]_{\mathcal{B}'} \ \middle| \ [T(\mathbf{v}_2)]_{\mathcal{B}'} \ \middle| \ \cdots \ \middle| \ [T(\mathbf{v}_n)]_{\mathcal{B}'} \ \right)$$

Example 4.13. Consider the identity $\mathbf{I}_V : V \to V$ and let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an ordered basis for V. Then $[\mathbf{I}_V]_{\mathcal{B}}$ of $\mathbf{I}_V$ with respect to $\mathcal{B}$ is the identity matrix in $M_n(\mathbb{F})$, since

$$[\mathbf{I}_V]_{\mathcal{B}} = \left(\ [\mathbf{v}_1]_{\mathcal{B}} \ \middle| \ [\mathbf{v}_2]_{\mathcal{B}} \ \middle| \ \cdots \ \middle| \ [\mathbf{v}_n]_{\mathcal{B}} \ \right) = \mathbf{I}$$

□

Let us consider another question: how much information is needed to specify a vector $\mathbf{v}$ in a vector space V? Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an ordered basis for V. Then $\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$. The information is recorded in the coordinate matrix of $\mathbf{v}$. This correspondence is an isomorphism.

Theorem 4.17. *Let V, W be vector spaces over $\mathbb{F}$ with $\dim V = n$ and $\dim W = m$. Let $\mathcal{B}$ be an ordered basis for V and $\mathcal{B}'$ an ordered basis for W. Then the map $\varphi : \mathcal{L}(V, W) \to M_{mn}(\mathbb{F})$ given by*

$$\varphi(T) = [T]_{\mathcal{B}',\mathcal{B}}$$

is an isomorphism. So $\mathcal{L}(V, W) \cong M_{mn}(\mathbb{F})$.

Proof. φ is linear. It follows from the linearity of matrix addition:

$$\varphi(c_1T_1 + c_2T_2) = [c_1T_1 + c_2T_2]_{\mathcal{B}',\mathcal{B}} = c_1[T_1]_{\mathcal{B}',\mathcal{B}} + c_2[T_2]_{\mathcal{B}',\mathcal{B}} = c_1\varphi(T_1) + c_2\varphi(T_2)$$

φ is bijective. $\dim \mathcal{L}(V, W) = (\dim V)(\dim W) = nm = \dim M_{mn}(\mathbb{F})$ by Thm. 4.11. Then we only need to show that φ is injective due to Cor. 4.9.1. So let $T \in \ker(\varphi)$. Then $\varphi(T) = \mathbf{0}_{mn} = [T]_{\mathcal{B}',\mathcal{B}}$. From the way we constructed $[T]_{\mathcal{B}',\mathcal{B}}$ above, we can see that $[T(\mathbf{v}_j)]_{\mathcal{B}'} = \mathbf{0}$ for all $\mathbf{v}_j \in \mathcal{B}$. So T must be the zero transformation. Hence φ is injective and is thus bijective. ∎

The isomorphism φ depends on ordered bases $\mathcal{B}$ and $\mathcal{B}'$. A different choice would give us a different φ.

Theorem 4.18. *Let $A \in M_{mn}(\mathbb{F})$ and let $T : \mathbb{F}^n \to \mathbb{F}^m$ be given by*

$$\forall \mathbf{v} \in \mathbb{F}^n, T(\mathbf{v}) = A\mathbf{v}$$

Let $\mathcal{B}$ and $\mathcal{B}'$ be the standard ordered bases for $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively. Then

$$[T]_{\mathcal{B}',\mathcal{B}} = A$$

Proof.

$$[T]_{\mathcal{B}',\mathcal{B}} = \left(\begin{array}{c|c|c} [T(\mathbf{e}_1)]_{\mathcal{B}'} & \cdots & [T(\mathbf{e}_n)]_{\mathcal{B}'} \end{array} \right)$$

We can check that for all $1 \leq i \leq n$,

$$T(\mathbf{e}_i) = A\mathbf{e}_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix}$$

Then the result follows. ∎

Example 4.14. Let $D : P_3(\mathbb{R}) \to P_3(\mathbb{R})$ be

$$D(a_0 + a_1x + a_2x^2 + a_3x^3) = \frac{\mathrm{d}}{\mathrm{d}x}\left(a_0 + a_1x + a_2x^2 + a_3x^3\right)$$

Let $\mathcal{B} = \{f_1 = 1, f_2 = x, f_3 = x^2, f_4 = x^3\}$ be the standard basis for $P_3(\mathbb{R})$.

$$\begin{aligned} [D]_{\mathcal{B}} &= \left(\begin{array}{c|c|c} [D(\mathbf{f}_1)]_{\mathcal{B}} & \cdots & [D(\mathbf{f}_4)]_{\mathcal{B}} \end{array} \right) \\ &= \left(\begin{array}{c|c|c|c} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 2 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ 3 \\ 0 \end{pmatrix} \end{array} \right) \end{aligned}$$

□

We now wish to generalize the results to more numbers of vector spaces. Let $T : U \to V$ and $S : V \to W$ be linear transformations. Also let $\mathcal{B}$, $\mathcal{B}'$, and $\mathcal{B}''$ be the ordered bases for U, V, W respectively. Consider the composition $ST : U \to W$. What is the relationship between $[T]_{\mathcal{B}',\mathcal{B}}$, $[S]_{\mathcal{B}'',\mathcal{B}'}$ and $[ST]_{\mathcal{B}'',\mathcal{B}}$?

To find out, take $\mathbf{u} \in U$ and

$$[ST(\mathbf{u})]_{\mathcal{B}''} = [ST]_{\mathcal{B}'',\mathcal{B}}[\mathbf{u}]_{\mathcal{B}}$$

However,

$$\begin{aligned}[ST(\mathbf{u})]_{\mathcal{B}''} &= [S(T(\mathbf{u}))]_{\mathcal{B}''} \\ &= [S]_{\mathcal{B}'',\mathcal{B}'}[T(\mathbf{u})]_{\mathcal{B}'} \\ &= [S]_{\mathcal{B}'',\mathcal{B}'}[T]_{\mathcal{B}',\mathcal{B}}[\mathbf{u}]_{\mathcal{B}}\end{aligned}$$

Since this is true for any $\mathbf{u}$, we can pick $\mathbf{u}$ as the basis vectors and then $[ST]_{\mathcal{B}'',\mathcal{B}} = [S]_{\mathcal{B}'',\mathcal{B}'}[T]_{\mathcal{B}',\mathcal{B}}$. Of course we can generalize this procedure for any number of transformations. Here, the mnemonic mentioned earlier again comes in handy.

Theorem 4.19. *Let $\mathbb{F}$ be a field. For $1 \leq i \leq r+1$, let V_i be a finite dimensional vector space over $\mathbb{F}$ and let $\mathcal{B}_i$ be an ordered basis for V_i. For $1 \leq j \leq r$, let $T_j : V_j \to V_{j+1}$ be a linear transformation. Consider the composition*

$$T = \bigcirc_{k=r}^{1} T_k$$

Then

$$[T]_{\mathcal{B}_{r+1},\mathcal{B}_1} = [T_r]_{\mathcal{B}_{r+1},\mathcal{B}_r}[T_{r-1}]_{\mathcal{B}_r,\mathcal{B}_{r-1}} \cdots [T_1]_{\mathcal{B}_2,\mathcal{B}_1}$$

Corollary 4.19.1. *Let $T : V \to W$ be an invertible linear transformation. Then $[T]^{-1}_{\mathcal{B}',\mathcal{B}} = [T^{-1}]_{\mathcal{B},\mathcal{B}'}$.*

Proof. $T^{-1} : W \to V$ is also an invertible linear transformation. $TT^{-1} = \mathbf{I}_W$ and $T^{-1}T = \mathbf{I}_V$ are the identity operators. Then by Thm. 4.19, we get the result. ∎

4.6 Change of basis

A special case may be interesting. If we take the identity operator $\mathbf{I}_V$ on V, and let $\mathcal{B}$ and $\mathcal{B}'$ be two ordered bases for V, then

$$[\mathbf{I}_V]_{\mathcal{B},\mathcal{B}'} = [\mathbf{I}_V]^{-1}_{\mathcal{B}',\mathcal{B}}$$

Perhaps now we can ask, if we have two bases $\mathcal{B}$ and $\mathcal{B}'$ for a vector space V, what are the relationships between $[\mathbf{v}]_{\mathcal{B}}$ and $[\mathbf{v}]_{\mathcal{B}'}$, and between $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{B}'}$?

Theorem 4.20 (Change of basis)**.** *Let V be a finite dimensional vector space and let $\mathcal{B}$ and $\mathcal{B}'$ be two ordered bases of V. Then for any $\mathbf{v} \in V$,*

$$[\mathbf{v}]_{\mathcal{B}} = [I_V]_{\mathcal{B},\mathcal{B}'}[\mathbf{v}]_{\mathcal{B}'}$$

Proof.

$$[\mathbf{v}]_{\mathcal{B}} = [\mathbf{I}_V(\mathbf{v})]_{\mathcal{B}} = [\mathbf{I}_V]_{\mathcal{B},\mathcal{B}'}[\mathbf{v}]_{\mathcal{B}'}$$

■

Definition 4.15. The matrix $P = [\mathbf{I}_V]_{\mathcal{B},\mathcal{B}'}$ is called the *transition matrix* from $\mathcal{B}'$ to $\mathcal{B}$. □

Theorem 4.21 (Change of matrix)**.** *Let V be a finite dimensional vector space and let $T : V \to V$ be a linear operator. If $\mathcal{B}$ and $\mathcal{B}'$ are two ordered bases of V and $P = [\mathbf{I}_V]_{\mathcal{B},\mathcal{B}'}$ is the transition matrix from $\mathcal{B}'$ to $\mathcal{B}$, then*

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

Proof. $T = \mathbf{I}_V \circ T \circ \mathbf{I}_V$. Then with Thm. 4.19,

$$\begin{aligned}[T]_{\mathcal{B}'} &= [\mathbf{I}_V \circ T \circ \mathbf{I}_V]_{\mathcal{B}'} \\ &= [\mathbf{I}_V]_{\mathcal{B}',\mathcal{B}}[T]_{\mathcal{B}}[\mathbf{I}_V]_{\mathcal{B},\mathcal{B}'} \\ &= P^{-1}[T]_{\mathcal{B}}P\end{aligned}$$

■

Example 4.15. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear operator given by

$$T\left[\begin{pmatrix} x \\ y \end{pmatrix}\right] = \begin{pmatrix} x+y \\ -2x+4y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

and $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$ and $\mathcal{B}' = \left\{\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right\}$. Then $[T]_{\mathcal{B}} = \begin{pmatrix} 1 & 1 \\ 2 & 4 \end{pmatrix}$.

Let P be the transition matrix from $\mathcal{B}'$ to $\mathcal{B}$. Then

$$P = [\mathbf{I}]_{\mathcal{B},\mathcal{B}'} = \left(\ [\mathbf{u}_1]_{\mathcal{B}} \mid [\mathbf{u}_2]_{\mathcal{B}}\ \right) = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Using Thm. 4.21,

$$[T]_{\mathcal{B}'} = P^{-1}[T]_{\mathcal{B}}P = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

This illustrates a reason why we may be interested in a change in basis. The new $[T]_{\mathcal{B}'}$ is diagonal and much simpler than the original. This would be something we will take on in a future section. □

Definition 4.16. Let A and B be two $n \times n$ matrices over the field $\mathbb{F}$. We say that B is *similar* to A over $\mathbb{F}$ if there is an invertible $n \times n$ matrix P such that

$$B = P^{-1}AP.$$

This is an equivalence relation on $M_n(\mathbb{F})$. □

The meaning of this definition is apparent from the following theorem.

Theorem 4.22. *Let V be a vector space and $T : V \to V$ be a linear operator. Then matrices A, B, are similar iff A, B represent the same operator T.*

Proof.

($\implies$) Let $\mathcal{B}$ be an ordered basis of V such that $[T]_{\mathcal{B}} = A$, and some invertible P such that $B = P^{-1}AP$. We want to find another ordered basis $\mathcal{B}'$ of V such that $[T]_{\mathcal{B}'} = B$. Now since P is invertible, by Thm. 3.5 its columns are linearly independent. Hence if we write $P = \begin{pmatrix} [\mathbf{v}_1]_{\mathcal{B}} & \cdots & [\mathbf{v}_n]_{\mathcal{B}} \end{pmatrix}$, it can be shown that $\mathcal{B}' = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is also linearly independent and hence forms a basis for V, and thus $P = [\mathbf{I}]_{\mathcal{B}\mathcal{B}'}$.

($\impliedby$) Thm. 4.21. ■

This allows us to define the determinant of a linear transformation. From Thm. 4.22, we can see that it is well defined.

Definition 4.17 (Determinant of a linear transformation)**.** Let $T : V \to V$ be a linear operator. Let $\mathcal{B}$ be an ordered basis of V. We define

$$\det T = \det[T]_{\mathcal{B}}$$

□

5 Diagonalization and Jordan canonical form

5.1 Diagonalization

In this chapter we shall focus only on linear operators $T : V \to V$ where V is a finite dimensional vector space.

Using ordered bases, T can be represented by matrices. Each basis will give us a different matrix. Naturally, we will want to know how to choose the basis to find the "nicest" matrices to represent T.

Example 5.1. Let $T : V \to V$ be a linear operator. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an ordered basis of V, and let $[T]_{\mathcal{B}} = D$ such that D is diagonal.

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_n \end{pmatrix} = [T]_{\mathcal{B}} = \left(\, [T(\mathbf{v}_1)]_{\mathcal{B}} \mid [T(\mathbf{v}_2)]_{\mathcal{B}} \mid \ldots \mid [T(\mathbf{v}_n)]_{\mathcal{B}} \, \right)$$

Then

$$\begin{aligned} T(\mathbf{v}_1) &= \lambda_1\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_n = \lambda_1\mathbf{v}_1 \\ T(\mathbf{v}_2) &= 0\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + 0\mathbf{v}_n = \lambda_2\mathbf{v}_2 \\ &\vdots \\ T(\mathbf{v}_n) &= \lambda_n\mathbf{v}_n \end{aligned}$$

Assume that $\lambda_1, \dots, \lambda_r \neq 0$, and $\lambda_{r+1} = \lambda_n = 0$. For any $\mathbf{v} \in V$, we can write $\mathbf{v} = \sum_i^n c_i \mathbf{v}_i$. Then

$$\begin{aligned} T(\mathbf{v}) &= \sum_{i=1}^{n} c_i T(\mathbf{v}_i) \\ &= \sum_{i-1}^{r} c_i \lambda_i \mathbf{v}_i \end{aligned}$$

Therefore, $\mathcal{R}(T) = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ and $\ker(T) = \text{span}\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$. Diagonal matrices make it convenient if we want to know the range and kernel for an operator. Another convenience brought about by diagonal matrices is

$$[T^k]_{\mathcal{B}} = D^k = \begin{pmatrix} \lambda_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^k \end{pmatrix}$$

□

This motivates us to the following definition:

Definition 5.1. A linear operator $T : V \to V$ is called *diagonalizable* if V has an ordered basis $\mathcal{B}$ such that the matrix $[T]_{\mathcal{B}}$ is diagonal. □

Moving forward, we will address these questions.

1. Is every linear operator diagonalizable?
2. If not, which linear operators are diagonalizable?
3. If operator T is diagonalizable, how do we find a basis $\mathcal{B}$ for which $[T]_{\mathcal{B}}$ is diagonal?
4. If operator T is not diagonalizable, what is the simplest form of their matrix representation?

The way we answer these questions is to consider diagonalizing matrices instead of linear transformations, and move the computation to something more familiar.

Definition 5.2. Let V be a vector space over a field $\mathbb{F}$ and $T : V \to V$ be a linear operator. A scalar $\lambda \in \mathbb{F}$ is an *eigenvalue* of T is there exists a **non-zero** vector $\mathbf{v} \in V$ such that

$$T(\mathbf{v}) = \lambda \mathbf{v}$$

and $\mathbf{v}$ is called the *eigenvector* of T corresponding to the eigenvalue λ. □

Then we can make the observation that $T : V \to V$ is diagonalizable iff V has a basis $\mathcal{B}$ consisting entirely of eigenvectors of T. Then the diagonal entries of $[T]_{\mathcal{B}}$ are the eigenvalues of T. Therefore in order to determine if T is diagonalizable, we have to find all the eigenvalues and eigenvectors and check if there are enough to make up a basis. This begs the question, how do we find the eigenvectors?

Let $\mathbf{I}_V : V \to V$ denote the identity operator on V. Then

$$\begin{aligned} T(\mathbf{v}) = \lambda \mathbf{v} &\iff T(\mathbf{v}) = \lambda \mathbf{I}_V(\mathbf{v}) \\ &\iff (T - \lambda \mathbf{I}_V)(\mathbf{v}) = \mathbf{0} \\ &\iff \mathbf{v} \in \ker(T - \lambda \mathbf{I}_V) \end{aligned}$$

Therefore,

$$\begin{aligned} \lambda \text{ is an eigenvalue of } T &\iff \exists \mathbf{v} \in \ker(T - \lambda \mathbf{I}_V), \mathbf{v} \neq 0 \\ &\iff \ker(T - \lambda \mathbf{I}_V) \neq \{\mathbf{0}\} \\ &\iff T - \lambda \mathbf{I}_V \text{ not invertible} \qquad (\because \text{ not injective}) \\ &\iff \det(T - \lambda \mathbf{I}_V) = 0 \end{aligned}$$

Definition 5.3. If λ is an eigenvalue of the linear operator $T : V \to V$ then the subspace of V

$$E_\lambda = \ker(T - \lambda \mathbf{I}_V) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \lambda \mathbf{v}\}$$

called the *eigenspace* corresponding to the eigenvalue λ. It contains all eigenvectors associated with λ and the zero vector. □

If $\mathcal{B}$ is an ordered basis of V and $A = [T]_\mathcal{B}$, then

$$[T - \lambda \mathbf{I}_V]_\mathcal{B} = A - \lambda \mathbf{I}$$

From Def. 4.17, we know that

$$\det(T - \lambda \mathbf{I}) = \det(A - \lambda \mathbf{I})$$

and

$$\begin{aligned} \lambda \text{ is an eigenvalue of } T &\iff \det(A - \lambda \mathbf{I}) = 0 \\ &\iff x = \lambda \text{ is a solution of } \det(x\mathbf{I} - A) = 0 \end{aligned}$$

Let

$$c_A(x) = \det(x\mathbf{I} - A) = \det \begin{pmatrix} x - a_{11} & -a_{12} & -a_{13} & \dots & -a_{1n} \\ -a_{21} & x - a_{22} & -a_{23} & \dots & -a_{2n} \\ -a_{31} & -a_{32} & x - a_{23} & \dots & -a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x - a_{n1} & -a_{n2} & -a_{n3} & \dots & -a_{nn} \end{pmatrix}$$

$c_A(x)$ is a *monic* polynomial of degree n. A monic polynomial is one whose leading coefficient is 1, that is $c_A(x) = x^n + b_{n-1}x^{n-1} + \dots + b_1 x + b_0$. The eigenvalues of T are the roots of $c_A(x)$.

Definition 5.4. Let $A \in M_n(\mathbb{F})$. Then

i. The characteristic polynomial of A is $c_A(x) = \det(x\mathbf{I} - A)$.

ii. The characteristic equation of A is $c_A(x) = 0$.

iii. A scalar $\lambda \in \mathbb{F}$ is an eigenvalue of A is there is a **non-zero** vector $\mathbf{v} \in \mathbb{F}^n$ such that $A\mathbf{v} = \lambda\mathbf{v}$. Then $\mathbf{v}$ is an *eigenvector* of A corresponding to λ.

iv. If λ is an eigenvalue of A, then the eigenspace of A corresponding to λ is the subspace $E_\lambda = \{\mathbf{v} \in \mathbb{F}^n \mid A\mathbf{v} = \lambda\mathbf{v}\}$ of $\mathbb{F}^n$. E_λ is also the solution space of the linear system $(A - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$.

□

The eigenvalue-eigenvector problem for matrices. Let $A \in M_n(\mathbb{F})$.

1. To find the eigenvalue of A, we solve the characteristic equation of the matrix.
2. To find the eigenvectors corresponding to the eigenvalue λ, we solve the homogeneous linear system

$$(A - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$$

We have seen that the eigenvalues of $T : V \to V$ are the eigenvalues of any matrix representation of T. We fix an ordered basis $\mathcal{B}$ and solve the eigenvalue-eigenvector problem for the matrix $A = [T]_\mathcal{B}$.

Also, $\mathbf{v}$ is an eigenvector of T corresponding to the eigenvalue λ iff $[\mathbf{v}]_\mathcal{B}$ is an eigenvector of $A = [T]_\mathcal{B}$ corresponding to the eigenvalue λ, since $[T(\mathbf{v})]_\mathcal{B} = A[v]_\mathcal{B}$.

Definition 5.5. Let $T : V \to V$ be a linear operator. Then the *characteristic polynomial* $c_T(x)$ of T is the characteristic polynomial of any matrix which represents T.

If $\mathcal{B}$ is a basis for V and $A = [T]_\mathcal{B}$, then $c_T(x) = c_A(x)$. The equation $c_T(x) = 0$ is the *characteristic equation* of T.

□

The characteristic polynomials are the same regardless of the matrix that we choose to represent T, since they are all similar, and have the same determinant.

Definition 5.6. Let $A \in M_n(\mathbb{F})$. A is *diagonalizable* over $\mathbb{F}$ if there exists an invertible matrix P over $\mathbb{F}$ such that the matrix

$$P^{-1}AP$$

is diagonal. □

Theorem 5.1. *A linear operator $T : V \to V$ is diagonalizable iff any of the matrices which represent T is diagonalizable.*

Proof.

($\Longrightarrow$) Pick any basis $\mathcal{B}$ and represent T relative to this basis. Let $A = [T]_\mathcal{B}$. Suppose we also find some basis $\mathcal{B}'$ consisting entirely of eigenvectors such that $D = [T]_{\mathcal{B}'}$ is diagonal. From Thm. 4.21, we can find P such that $[T]_{\mathcal{B}'} = P^{-1}[T]_\mathcal{B}P$. ■

Does every matrix have an eigenvalue? Since not all polynomial equations have roots, the answer is no. It also depends on the field we are in. A simple example is the equation $x^2+1=0$. Nevertheless, from the fundamental theorem of algebra we know that every polynomial of degree n has n solutions in $\mathbb{C}$. Then we can write our monic complex polynomial as

$$c_0 + c_1 x + \cdots + c_{n-1}x^{n-1} + x^n = (x-b_1)^{m_1} \cdots (x-b_r)^{m_r}$$

where $b_1, \ldots, b_r$ are distinct and $m_i, \ldots, m_r$ are positive integers called the *multiplicity* of the corresponding root b_i. Then in fact all matrices in $M_n(\mathbb{C})$ have eigenvalues in $\mathbb{C}$.

From Def. 5.4, for some $A \in M_n(\mathbb{R})$, the eigenvalues will differ depending on if we treat it as a real matrix or a complex matrix, since the eigenvalues and hence the eigenvectors must come from the same field.

Example 5.2. Let $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Then $c_A(x) = x^2 + 1$.

A has no eigenvalues if we treat A as a real matrix.

If A is a complex matrix, then $x = \pm i$ are its eigenvalues and $\begin{pmatrix} i \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -i \\ 1 \end{pmatrix}$ are its eigenvectors.

□

5.2 Diagonalizability

Let linear operator $T : V \to V$ be diagonalizable, and $\mathcal{B}$ be a basis for V consisting entirely of eigenvectors. Suppose $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of T. Rearrange the basis elements so that the first d_1 vectors all have eigenvalue λ_1, the next d_2 vectors have eigenvalue λ_2, and so on. Then

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 \mathbf{I}_{d_1} & & & & 0 \\ & \lambda_2 \mathbf{I}_{d_2} & & & \\ & & \lambda_3 \mathbf{I}_{d_3} & & \\ & & & \ddots & \\ & 0 & & & \lambda_k \mathbf{I}_{d_k} \end{pmatrix}$$

where $\mathbf{I}_{d_i}$ is an identity matrix of size d_i and the zeros are all appropriately sized.

Then $c_T(x) = (x-\lambda_1)^{d_1} \cdots (x-\lambda_k)^{d_k}$. This can be seen if we write the entire matrix out:

$$A = \begin{pmatrix} \lambda_1 & & & & & & & & & & & 0 \\ & \ddots & & & & & & & & & & \\ & & \lambda_1 & & & & & & & & & \\ & & & \lambda_2 & & & & & & & & \\ & & & & \ddots & & & & & & & \\ & & & & & \lambda_2 & & & & & & \\ & & & & & & \ddots & & & & & \\ & & & & & & & \ddots & & & & \\ & & & & & & & & \ddots & & & \\ & & & & & & & & & \lambda_k & & \\ & 0 & & & & & & & & & \ddots & \\ & & & & & & & & & & & \lambda_k \end{pmatrix}$$

With Def. 5.4 we can also see why for the eigenspaces E_{λ_i}, $\dim E_{\lambda_i} = d_i$. Also of course $d_1 + \cdots + d_k = n$. These form a set of necessary conditions for diagonalizability.

Theorem 5.2. *Let $T : V \to V$ be a linear operator. If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a set of eigenvectors, each corresponding to a distinct eigenvalue, then it is linearly independent.*

Proof. We perform induction on the number of eigenvectors. $\{\mathbf{v}_1\}$ is linearly independent. Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent for $1 \leq n \leq m-1$.

Let us consider the case where $n = m$, so we check the equation $a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0}$. We can write $-a_m\mathbf{v}_m = a_1\mathbf{v}_1 + \cdots + a_{m-1}\mathbf{v}_{m-1}$. Applying T on both sides and rearranging, we get

$$\begin{aligned} a_1\lambda_1\mathbf{v}_1 + \cdots + a_{m-1}\lambda_{m-1}\mathbf{v}_{m-1} &= -a_m\lambda_m\mathbf{v}_m \\ &= a_1\lambda_m\mathbf{v}_1 + \cdots + a_{m-1}\lambda_m\mathbf{v}_{m-1} \\ a_1(\lambda_1 - \lambda_m)\mathbf{v}_1 + \cdots + a_{m-1}(\lambda_{m-1} - \lambda_m)\mathbf{v}_{m-1} &= \mathbf{0} \end{aligned}$$

Since the eigenvalues are distinct, $(\lambda_i - \lambda_m) \neq 0$ for all $1 \leq i < m$. But since $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ are linearly independent by our original supposition, $a_1, \ldots, a_{m-1} = 0$. Since $\mathbf{v}_m \neq \mathbf{0}$, then $a_m = 0$ as well. Hence they are linearly independent. ∎

Corollary 5.2.1. *Let $T : V \to V$ be a linear operator. If $\dim V = n$ and T has n distinct eigenvalues, then it is diagonalizable.*

Proof. Let $\lambda_1, \ldots, \lambda_n$ be n distinct eigenvalues of T, and let the corresponding eigenvectors be $\mathbf{v}_1, \ldots, \mathbf{v}_n$. By Thm. 5.2, $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent, and by Cor. 3.12.1, it is a basis for V. Then $[T]_B$ is diagonal. ∎

Note that this does not mean that operators with less than n distinct eigenvalues are not diagonalizable.

Theorem 5.3. *Let $T : V \to V$ be a linear operator and $\lambda_1, \ldots, \lambda_r$ are distinct eigenvalues of T. The following are equivalent:*

i. *T is diagonalizable.*

ii. *The characteristic polynomial for T is*

$$c_T(x) = (x - \lambda_2)^{d_1} \cdots (x - \lambda_r)^{d_r}$$

where $\dim E_{\lambda_i} = d_i$ for $i = 1, \ldots, r$.

iii. $\dim E_{\lambda_1} + \cdots + E_{\lambda_r} = \dim V$

iv. $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}$.

Proof.

($(i) \implies (ii)$) The necessary conditions expounded above show this.

($(ii) \implies (iii)$) We can see that $d_1 + \cdots + d_r$ gives the degree of $c_T(x)$. The degree is equal to $n = \dim V$ since it is derived from the determinant of a $n \times n$ matrix.

(($(iii) \implies (iv)$) For each $1 \le i \le r$, let $\mathcal{B}_i = \{\mathbf{v}_{i1}, \dots, \mathbf{v}_{id_i}\}$ be a basis for E_{λ_i}. Let $\mathcal{B} = \bigcup_i^r \mathcal{B}_i = \{\mathbf{v}_{11}, \dots, \mathbf{v}_{1d_1}, \dots, \mathbf{v}_{r1}, \dots, \mathbf{v}_{rd_r}\}$. Now all the $\mathcal{B}_i$ are pairwise disjoint since a vector cannot be an eigenvector for two eigenvalues. There are exactly $n = \dim V$ vectors in $\mathcal{B}$. We claim that they are linearly independent. Consider the equation

$$\underbrace{\alpha_{11}\mathbf{v}_{11} + \cdots + \alpha_{1d_1}\mathbf{v}_{1d_1}}_{\mathbf{w}_1} + \cdots + \underbrace{\alpha_{r1}\mathbf{v}_{r1}, \dots, \alpha_{rd_r}\mathbf{v}_{rd_r}}_{\mathbf{w}_r} = \mathbf{0}$$

All $\mathbf{w}_i$ are also eigenvectors corresponding to eigenvalue λ_i. However, from Thm. 5.2, all $\mathbf{w}_i$'s are linearly independent. Therefore all $w_i = \mathbf{0}$. Then this means

$$\alpha_{i1}\mathbf{v}_{i1} + \cdots + \alpha_{id_i}\mathbf{v}_{id_i} = \mathbf{w}_i = \mathbf{0}$$

so we get that all α's are 0.

Then, $\mathcal{B}$ forms a basis for V. By Thm. 3.16, condition (iv) follows.

(($(iv) \implies (i)$) Taking the bases of the eigenspaces, Thm. 3.16 tells us that the union of all these bases form a basis for V. We now have a basis formed of only eigenvectors. By Cor. 5.2.1, T is diagonalizable. ■

Hence, to know if $A \in M_n(\mathbb{F})$ is diagonalizable, we find all the eigenvalues of A, and a basis for each of the eigenspaces. Then A is diagonalizable only if there are a total of n vectors in all the bases.

Theorem 5.4 (Diagonalizing a matrix)**.** *Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis of $\mathbb{F}^n$ consisting of eigenvectors of A. Form the matrix $P = \left(\mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_n \right)$. Then $P^{-1}AP$ is diagonal. More specifically, if each $\mathbf{v}_i$ is the eigenvector associated with eigenvalue λ_i, then*

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_n \end{pmatrix}$$

Proof. This is in fact just a change of basis. Let $L_A : \mathbb{F}^n \to \mathbb{F}^n$ defined by $L_A(\mathbf{v}) = A\mathbf{v}$. Then $[L_A]_S = A$ where S is the standard ordered basis of $\mathbb{F}^n$. Then

$$[L_A]_{\mathcal{B}} = \left([L_A(\mathbf{v}_1)]_{\mathcal{B}} \mid \cdots \mid [L_A(\mathbf{v}_n)]_{\mathcal{B}} \right) = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_n \end{pmatrix}$$

We also know that $[L_A]_{\mathcal{B}} = [\mathbf{I}]_{\mathcal{B},S}[L_A]_S[\mathbf{I}]_{S,\mathcal{B}}$. We can see that $P = [\mathbf{I}]_{S,\mathcal{B}}$. ■

Example 5.3. We have seen before that if we have a complex matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, then $x = \pm i$ are its eigenvalues and $\begin{pmatrix} i \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -i \\ 1 \end{pmatrix}$ are its eigenvectors. These two eigenvectors form a basis for $\mathbb{C}^2$. Furthermore, if $P = \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}$, then $P^{-1}AP = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$. □

5.3 Invariant subspaces

Definition 5.7. Let $T : V \to V$ be a linear operator. A subspace W of V is called a *T-invariant subspace* of V if $T(W) \subseteq W$. $\square$

Some observations:

- Clearly $\{\mathbf{0}\}$ and V itself are T-invariant.
- Since $\mathbf{0} \in \ker(T)$, $\ker(T)$ is also T-invariant .
- $\mathcal{R}(T)$ is also T-invariant since applying T to vectors in it must still result in something in the range.
- Eigenspaces are T-invariant. Applying T to any eigenvector only scales it, hence the result is still in the eigenspace.

Example 5.4. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+y+z \\ y+z \\ z \end{pmatrix}$$

We see that for vectors on the xy-plane,

$$T\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} x+y \\ y \\ 0 \end{pmatrix}$$

is also in the xy-plane. So the xy-plane is T-invariant. $\square$

Example 5.5. Let $D : P_3(\mathbb{F}) \to P_3(\mathbb{F})$ be the linear operator given by

$$D(a_0 + a_1x + a_2x^2 + a_3x^4) = \frac{\mathrm{d}}{\mathrm{d}x}\left(a_0 + a_1x + a_2x^2 + a_3x^4\right)$$

Consider $P_2(\mathbb{F})$:

$$D(a_0 + a_1x + a_xx^2) = a_1 + 2a_2x \in P_2(\mathbb{F})$$

So $P_2(\mathbb{F})$ is D-invariant. $\square$

Definition 5.8. Let $T : V \to V$ be a linear operator and let $W \subseteq V$ be a T-invariant subspace of V. Then define $T_W : W \to W$ as

$$\forall \mathbf{w} \in W, T_W(\mathbf{w}) = T(\mathbf{w})$$

T_W is a linear operator on W called the *restriction of T to W*. $\square$

Theorem 5.5. *If $T : V \to V$ is a linear operator and $W \subseteq V$ is a T-invariant subspace of V, then the characteristic polynomial of T_W divides the characteristic polynomial of T.*

Proof. Let $\dim V = n$ and $\dim W = m$. Take a basis for W, $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$. We can extend this to a basis for V, $\mathcal{B}' = \{\mathbf{w}_1, \ldots, \mathbf{w}_m, \mathbf{w}_{m+1}, \ldots \mathbf{w}_n\}$. Let $A = [T]_{\mathcal{B}'} = (a_{ij})$, and $B = [T_W]_{\mathcal{B}} = (b_{ij})$.

Then

$$
\begin{aligned}
T(\mathbf{w}_j) &= \sum_{i=1}^{n} a_{ij}\mathbf{w}_i, & 1 \le j \le n \\
T_W(\mathbf{w}_j) &= \sum_{i=1}^{m} b_{ij}\mathbf{w}_i, & 1 \le j \le m.
\end{aligned}
$$

Since W is T-invariant, $T(\mathbf{w}_1), \ldots, T(\mathbf{w}_m) \in W = \operatorname{span}\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$. So $T(\mathbf{w}_1), \ldots, T(\mathbf{w}_m)$ are linear combinations of $\mathbf{w}_1, \ldots, \mathbf{w}_m$ only. Thus for $1 \le j \le m$, $T(\mathbf{w}_j) = \sum_{i=1}^{m} b_{ij}\mathbf{w}_i + \sum_{i=m+1}^{n} \mathbf{0}$.

Therefore we know that (the other parts in asterisk we have no information about):

$$
A = \begin{pmatrix}
b_{11} & \ldots & b_{1m} & *** \\
\vdots & & \vdots & \\
b_{m1} & \ldots & b_{mm} & *** \\
0 & \ldots & 0 & *** \\
\vdots & & \vdots & \\
0 & \ldots & 0 & ***
\end{pmatrix}
= \left(\begin{array}{c|c} B & * \\ \hline \mathbf{0}_m & ** \end{array}\right)
$$

The characteristic equation of T

$$
\begin{aligned}
c_T(x) &= \det(x\mathbf{I}_n - A) \\
&= \det \begin{pmatrix} x\mathbf{I}_m - B & -* \\ \mathbf{0} & x\mathbf{I}_m - ** \end{pmatrix} \\
&= c_{T_W}(x) q(x)
\end{aligned}
$$

where $q(x)$ is some polynomial, since for triangular matrices the determinant is just the product of diagonal entries. ■

Below is one of the simplest ways to create an invariant subspace.

Definition 5.9. Let $T : V \to V$ be a linear operator and let $\mathbf{u} \in V$ be a non-zero vector. The subspace

$$W = \operatorname{span}\{\mathbf{u}, T(\mathbf{u}), T^2(\mathbf{u}), \ldots\}$$

is called the *T-cyclic subspace of V generated by* $\mathbf{u}$. □

Lemma 5.6. *Let $T : V \to V$. If W is a T-cyclic subspace of a vector space V generated by some non-zero vector $\mathbf{u} \in V$, then W is T-invariant.*

Proof. Let $\mathbf{w} \in W$. Then $\mathbf{w} = c_0\mathbf{u} + c_1T(\mathbf{u}) + c_2T^2(\mathbf{u}) + \cdots + c_nT^n(\mathbf{u})$. Note that we cannot go on forever, since by definition even for spans of infinite sets we can only take finite linear combinations. Then $T(\mathbf{w}) = c_0T(\mathbf{u}) + c_1T^2(\mathbf{u}) + \cdots + c_nT^{n+1}(\mathbf{u}) \in \operatorname{span}\{\mathbf{u}, T(\mathbf{u}), T^2(\mathbf{u}), \ldots\}$. ■

Are the spanning sets always infinite?

Example 5.6. Let us revisit T from Ex. 5.4 again. Take $\mathbf{u} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. But $\mathbf{u} = T(\mathbf{u}) = T^2(\mathbf{u}) = \cdots$. So the T-cyclic subspace generated by $\mathbf{u}$ is just spanned by $\mathbf{u}$ itself, i.e. the x-axis.

However if we take $\mathbf{u} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, we get that

$$T(\mathbf{u}) = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad T^2(\mathbf{u}) = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \qquad T^3(\mathbf{u}) = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} \qquad \cdots \qquad T^n(\mathbf{u}) = \begin{pmatrix} n \\ 1 \\ 0 \end{pmatrix}$$

and the T-cyclic subspace generated by this new $\mathbf{u}$ is actually the xy-plane. □

Theorem 5.7. *Let $T : V \to V$ be a linear operator. Let W be the T-cyclic subspace of V generated by a non-zero $\mathbf{u} \in V$ and $\dim W = n$.*

i. The set $\{\mathbf{u}, T(\mathbf{u}), T^2(\mathbf{u}), \ldots, T^{n-1}(\mathbf{u})\}$ is a basis for W.

ii. There exist scalars $a_0, \ldots, a_{n-1}$ such that

$$a_0\mathbf{u} + a_1 T(\mathbf{u}) + \cdots + a_{n-1}T^{n-1}(\mathbf{u}) + T^n(\mathbf{u}) = \mathbf{0}$$

and the characteristic polynomial of T_W is given by

$$c_{T_W}(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n$$

Proof.

i. Since $\mathbf{u} \neq \mathbf{0}$, $\{\mathbf{u}\}$ is linearly independent. Take $\{\mathbf{u}, T(\mathbf{u})\}$. If this is still linearly independent, we take $\{\mathbf{u}, T(\mathbf{u}), T^2(\mathbf{u})\}$. If this is still linearly independent, we add $T^3(\mathbf{u})$, and so on. This will terminate at some point since W is finite dimensional.

   Let k be the largest integer such that $B = \{\mathbf{u}, T(\mathbf{u}), \ldots, T^{k-1}(\mathbf{u})\}$ is linearly independent. There are k vectors in B. Let $W' = \operatorname{span}(B)$. Since $B \subseteq W$, $W' \subseteq W$.

   By the definition of k, $B \cup \{T^k(\mathbf{u})\}$ is linearly dependent. Then there exists scalars $b_0, \ldots, b_k$, not all zero, such that

$$b_0\mathbf{u} + b_1 T(\mathbf{u}) + \cdots + b_{k-1}T^{k-1}(\mathbf{u}) + b_k T^k(\mathbf{u}) = \mathbf{0}$$

   $b_k \neq 0$ since that would imply B is linearly dependent. Then we can rearrange and express $T^k(\mathbf{u})$ as a linear combination of the vectors in B.

   Then W' is T-invariant. For any $\mathbf{w} \in W'$, we can write $\mathbf{w} = c_0\mathbf{u} + \cdots + c_{k-1}T^{k-1}(u)$, and $T(\mathbf{w}) = c_0 T(\mathbf{u}) + \cdots + c_{k-1}T^k(u) \in W'$ as we have shown above.

   Hence we can also inductively show that $T^n(\mathbf{u}) \in W'$. Then all the spanning vectors for W are in W'. Therefore $W \subseteq W'$, and we conclude that $W = W'$, with B as a basis for W.

ii. By definition,

$$\begin{aligned}
[T_W]_B &= \left([T_w(\mathbf{u})]_B \mid [T_w(T(\mathbf{u}))]_B \mid \cdots \mid [T_w(T^{n-1}(\mathbf{u}))]_B \right) \\
&= \left([T(\mathbf{u})]_B \mid [T^2(\mathbf{u})]_B \mid \cdots \mid [T^n(\mathbf{u})]_B \right) \\
&= \left(\begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \middle| \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \middle| \cdots \middle| \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \middle| \begin{pmatrix} -a_0 \\ -a_1 \\ -a_2 \\ \vdots \\ -a_{n-1} \end{pmatrix} \right)
\end{aligned}$$

where $a_0, \ldots, a_{n-1}$ are scalars such that

$$a_0\mathbf{u} + \cdots a_{n-1}T^{n-1}(\mathbf{u}) + T^n(\mathbf{u}) = \mathbf{0}$$

as we have shown previously.

Then the characteristic polynomial can be shown to be (with lots of induction):

$$\begin{aligned}
c_{T_W}(x) &= \det \begin{pmatrix} x & 0 & 0 & \ldots & 0 & a_0 \\ -1 & x & 0 & \ldots & 0 & a_1 \\ 0 & -1 & x & \ldots & 0 & a_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & x + a_{n-1} \end{pmatrix} \\
&= \underbrace{x \det \begin{pmatrix} x & 0 & \ldots & 0 & a_1 \\ -1 & x & \ldots & 0 & a_2 \\ 0 & -1 & \ldots & 0 & a_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & -1 & x + a_{n-1} \end{pmatrix}}_{=a_1 + a_2x + \cdots + a_{n-1}x^{n-2} + x^{n-1}} + (-1)^{n+1} a_0 \underbrace{\det \begin{pmatrix} -1 & x & 0 \\ \vdots & \ddots & \ddots \\ 0 & \ldots & -1 \end{pmatrix}}_{=(-1)^{n-1}} \\
&= a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + x^n
\end{aligned}$$

■

5.4 Cayley Hamilton theorem

Definition 5.10. If $T : V \to V$ is a linear operator on a vector space V over $\mathbb{F}$ and $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ is a polynomial over $\mathbb{F}$, then $f(T)$ is the linear operator

$$f(T) = a_0\mathbf{I}_V + a_1T + a_2T^2 + \cdots + a_nT^n$$

Similarly for matrix $A \in M_n(\mathbb{F})$, $f(A)$ is the $n \times n$ matrix given by

$$f(A) = a_0\mathbf{I}_n + a_1A + a_2A^2 + \cdots + a_nA^n$$

If $\mathcal{B}$ is an ordered basis for V and $[T]_\mathcal{B} = A$, then

$$[f(T)]_\mathcal{B} = f(A) \tag{1}$$

□

Exercise. *The characteristic polynomial for* $A = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix}$ *can be shown to be* $c_A(x) = x^3 - 5x^2 + 8x - 4$.

Surprisingly, a simple calculation shows that

$$c_A(A) = \mathbf{0}_3$$

Is this a coincidence? It is not, and the Cayley Hamilton theorem will tell us so. Perhaps some motivation or intuition at first:

Assume a linear operator $T : V \to V$ is diagonalizable. Then V is the direct sum of the eigenspaces of T, by Thm. 5.3. Not only is an eigenspace the space containing some eigenvectors, it is also $\ker(T - \lambda_i \mathbf{I})$ (see Def. 5.3).

If $c_T(x) = (x - \lambda_1)^{d_1} \cdots (x - \lambda_n)^{d_n}$, then $c_T(T) = (T - \lambda_1 \mathbf{I}_V)^{d_1} \cdots (T - \lambda_n \mathbf{I}_V)^{d_n}$. Since V is the direct sum of these eigenspaces, we can write every $\mathbf{v} \in V$ as a linear combination of eigenvectors. Now we can see that $c_T(T)$ will kill off eigenvectors in each space one by one, since the eigenvectors are in the kernel of one of the terms in $c_T(T)$. So it sends $\mathbf{v}$ to $\mathbf{0}$.

However this is a special case that the transformation is diagonalizable. In general, can we still say the same?

Theorem 5.8 (Cayley Hamilton theorem)**.** *Let T be a linear operator on a finite-dimensional vector space V, and let $c_T(x)$ be the characteristic polynomial of T. Then*

$$c_T(T) = T_0$$

where T_0 is the zero transformation.

Proof. Similar to above, we shall show that $\forall \mathbf{v} \in V, c_T(T)(\mathbf{v}) = \mathbf{0}$. It is trivial if $\mathbf{v} = \mathbf{0}$. Otherwise, let $W = \operatorname{span}\{\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v}), \ldots\}$, the T-cyclic subspace generated by $\mathbf{v}$. Let $m = \dim W$.

By Thm. 5.7, there are scalars $a_0, \ldots, a_{m-1}$ such that

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

and

$$c_{T_W}(x) = a_0 x + a_1 x + \cdots + a_{m-1}x^{m-1} + x^m = 0$$

where T_W is the restriction of T to W. Then,

$$c_{T_W}(T) = a_0 \mathbf{I}_V + a_1 T + \cdots + a_{m-1}T^{m-1} + T^m = \mathbf{0}$$

Looking again, we see that we have shown that $(c_{T_W}(T))(\mathbf{v}) = \mathbf{0}$. However, Thm. 5.5 tells us that $c_{T_W}(x)$ divides $c_T(x)$. Therefore for some polynomial $q(X)$,

$$\begin{aligned} c_T(x) &= q(x)c_{T_W}(x) \\ c_T(T) &= q(T)c_{T_W}(T) \\ c_T(T)(\mathbf{v}) &= q(T)[c_{T_W}(T)(\mathbf{v})] \\ &= \mathbf{0} \end{aligned}$$

■

Theorem 5.9 (Cayley Hamilton theorem for matrices)**.** *Let $A \in M_n(\mathbb{F})$ and let $c_A(x)$ be its characteristic polynomial. Then*

$$c_A(A) = \mathbf{0}_n$$

is the $n \times n$ zero matrix.

Proof. Define $T : \mathbb{F}^n \to \mathbb{F}^n$ with $T(\mathbf{v}) = A\mathbf{v}$. Let $\mathcal{B}$ be the standard ordered basis of $\mathbb{F}^n$. Then $[T]_{\mathcal{B}} = A$, and $\mathbf{0} = [c_T(T)]_{\mathcal{B}} = c_T(A) = c_A(A)$. ■

We are now interested to know, are any other polynomials $p(x)$ such that $p(A)$ is the zero matrix? Yes, and an easy example will be any multiple of $c_A(x)$. However this is not very exciting. Could we perhaps find a polynomial with a lower degree than $c_A(x)$?

Exercise. *Reusing the same matrix* $A = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3, \end{pmatrix}$ *a quick calculation shows that for* $p(x) = (x-1)(x-2)$,

$$p(A) = \mathbf{0}_3$$

but $p(x)$ has a lower degree that $c_A(x)$.

Definition 5.11. Let A be an $n \times n$ matrix over $\mathbb{F}$. The *minimal polynomial* of A is the polynomial $m_A(x)$ with coefficients in $\mathbb{F}$ such that

i. $m_A(x)$ is monic,

ii. $m_A(A) = \mathbf{0}_n$,

iii. $m_A(x)$ has the smallest degree possible.

□

Theorem 5.10. *If $p(x)$ is a polynomial such that $p(A) = \mathbf{0}_n$, then the minimal polynomial $m_A(x)$ divides $p(x)$.*

Proof. $p(x) = m_A(x)q(x) + r(x)$, from division of polynomials. Now either $r(x)$ is zero or it is a polynomial with degree less than degree of $m_A(x)$.

Suppose $r(x)$ is not zero. Then

$$\begin{aligned} r(A) &= p(A) - m_A(A)q(A) \\ &= \mathbf{0}_n - \mathbf{0}_n q(A) \end{aligned}$$

Now $m_A(x)$ has higher degree than $r(x)$, but also $r(A) = \mathbf{0}_n$ which means $m_A(x)$ cannot be the minimal polynomial. Hence $r(x)$ must be zero. ■

Theorem 5.11. *Similar matrices have the same minimal polynomial.*

Proof. Suppose we have two similar matrices A, B, and some invertible matrix P such that $B = P^{-1}AP$.

Take a non-zero monic polynomial $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0$. Notice:

$$\begin{aligned} f(B) &= P^{-1}A^nP + a_{n-1}P^{-1}A^{n-1}P + \ldots + a_0 \\ &= P^{-1}f(A)P \end{aligned}$$

Therefore $m_A(B) = P^{-1}m_A(A)P = 0$, and $0 = m_B(B) = P^{-1}m_B(A)P$, so with the properties of minimal polynomials, we gather that m_A and m_B divides each other. Therefore $m_A = m_B$. ■

Definition 5.12. The minimal polynomial $m_T(x)$ of a linear operator $T : V \to V$ is defined as the minimal polynomial of any matrix which represents T. □

Theorem 5.12. $m_T(\lambda) = 0 \iff c_T(\lambda) = 0$. *In other words, the minimal polynomial and characteristic polynomial have the same roots.*

Proof.

($\implies$) $m_T(x)$ divides $c_T(x)$. So if $m_T(\lambda) = 0$, then $c_T(\lambda) = q(\lambda)m_T(\lambda) = 0$.

($\impliedby$) Suppose $c_T(\lambda) = 0$. Then there exists an eigenvector $\mathbf{v}$ such that $T(\mathbf{v}) = \lambda\mathbf{v}$. Also if λ is an eigenvalue for A, then λ^k is an eigenvalue for A^k. Thus $[m_T(\lambda)]\mathbf{v} = [m_T(T)](\mathbf{v})$. However $[m_T(T)](\mathbf{v}) = 0$. So since $\mathbf{v} \neq \mathbf{0}$, $m_T(\lambda) = 0$. ■

Theorem 5.13. *If $A \in M_n(\mathbb{F})$ is invertible, then there is a polynomial $g(x)$ with coefficients in $\mathbb{F}$ such that $g(A) = A^{-1}$.*

Proof. Take the characteristic polynomial of A, $c_A(x) = a_0 + a_1x + \ldots + a_{n-1}x^{n-1} + x^n$. $a_0 \neq 0$ otherwise $c_A(x) = x(a_1 + \ldots + a_{n-1}x^{n-1} + x^{n-1})$ and 0 is an eigenvalue for A. But A is invertible, so 0 cannot be its eigenvalue. So therefore we have

$$\begin{aligned} c_A(x) - x(a_1 + \ldots + a_{n-1}x^{n-1} + x^{n-1}) &= a_0 \\ \frac{c_A(A) - A(a_1 + \ldots + A^{n-1})}{a_0} &= \mathbf{I} \\ Ag(A) &= \mathbf{I} \end{aligned}$$

thus $g(A) = A^{-1}$ where $g(x) = -\frac{a_1 + \ldots + a_{n-1}x^{n-2} + x^{n-1}}{a_0}$. ■

5.5 Jordan canonical form

In this section, we will assume all our matrices are complex matrices.

Not all matrices can be diagonalized. What is the next best thing we can get?

Definition 5.13. Let $\lambda \in \mathbb{C}$. The $k \times k$ matrix

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ & & \ddots & \ddots & \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}$$

is called a *Jordan block* of order k corresponding to λ. □

The characteristic polynomial for a Jordan block is $c_{J_k(\lambda)}(x) = (x-\lambda)^k$. Its minimal polynomial is the same as its characteristic polynomial. To see this, we can look at the action of $(J_k(\lambda)-\lambda\mathbf{I})$ on itself – it just shifts all the rows up once. Also we can also see that it only has one eigenvalue λ and one eigenspace, span$\{\mathbf{e}_1\}$.

Theorem 5.14 (Jordan canonical form)**.** *Let A be an $n \times n$ complex matrix. Then there exists an invertible matrix P such that (in block form):*

$$P^{-1}AP = J = \begin{pmatrix} J_{k_1}(\lambda_1) & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & J_{k_2}(\lambda_2) & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & J_{k_{r-1}}(\lambda_{r-1}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & J_{k_r}(\lambda_r) \end{pmatrix}$$

where $J_{k_i}(\lambda_i)$ are Jordan blocks. They are unique up to reordering. The matrix J is called a Jordan canonical form *of A.*

For linear operators, let $T : V \to V$ be a linear operator. Take any ordered basis $\mathcal{B}$ of V and consider $A = [T]_{\mathcal{B}}$. Then A is similar to a matrix J in Jordan canonical form. There is a basis $\mathcal{B}'$ for V such that $[T]_{\mathcal{B}'} = J$. We call J a *Jordan canonical form of T*.

This theorem is remarkable since it applies to any square complex matrix, unlike diagonalization. Every square complex matrix is similar to a matrix in Jordan canonical form. In some sense this is the nicest matrix that is similar to A, as it gives us information regarding A.

The characteristic polynomial of A, $c_A(x) = c_J(x) = (x-\lambda_1)^{k_1}(x-\lambda_2)^{k_2}\cdots(x-\lambda_r)^{k_r}$ and $\lambda_1, \ldots, \lambda_r$, not necessarily distinct, are the eigenvalues of A.

If we rearrange the blocks such that the largest blocks associated to the s distinct eigenvalues are placed at the front, the minimal polynomial of A, $m_A(x) = m_J(x) = (x-\lambda_1)^{k_1}(x-\lambda_2)^{k_2}\cdots(x-\lambda_s)^{k_s}$.

If all the Jordan blocks in a matrix's Jordan canonical form are of order 1, then it is diagonalizable. This gives us the following result.

Corollary 5.14.1. *A matrix A is diagonalizable iff the roots of its minimal polynomial are distinct, i.e. $m_A(x) = (x-\lambda_1)\cdots(x-\lambda_r)$.*

It is generally difficult to determine the Jordan canonical form of a matrix. However using information from the characteristic and minimal polynomials we can limit the domain of our search.

Example 5.7. Let $A \in M_n(\mathbb{C})$ such that

$$c_A(x) = (x-5)^4(x-2)^2 \qquad m_A(x) = (x-5)^2(x-2)$$

Since $c_A(x)$ is of degree 6, $n = 6$. The eigenvalues of A are 5 and 2. From $c_A(x)$ we can tell the number of times the eigenvalues appear on its diagonal. The minimal polynomial then tells us that the largest Jordan blocks are $J_2(5)$ and $J_1(2)$.

Hence in total, we have $J_2(5)$, two copies of $J_1(2)$, and a choice between a single $J_2(5)$ or two copies of $J_1(5)$. $\square$

Example 5.8. Let $T : V \to V$ be a linear operator on a complex vector space V. Let there be an ordered basis $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_8\}$ for V and let

$$[T]_{\mathcal{B}} = \begin{pmatrix} J_3(2) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & J_1(2) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & J_2(3) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & J_1(1) \end{pmatrix}$$

It has eigenspaces

$$E_2 = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_4\} \qquad E_3 = \operatorname{span}\{\mathbf{v}_5\} \qquad E_1 = \operatorname{span}\{\mathbf{v}_7\}$$

by inspecting the position of the blocks. So this operator is clearly not diagonalizable.

We can express $T(\mathbf{v}_i)$ in linear combinations of elements in $\mathcal{B}$.

$$\begin{aligned} T(\mathbf{v}_1) &= 2\mathbf{v}_1 && \implies (T - 2\mathbf{I})(\mathbf{v}_1) = \mathbf{0} \\ T(\mathbf{v}_2) &= \mathbf{v}_1 + 2\mathbf{v}_2 && \implies (T - 2\mathbf{I})^2(\mathbf{v}_2) = \mathbf{0} \\ T(\mathbf{v}_3) &= \mathbf{v}_2 + 2\mathbf{v}_3 && \implies (T - 2\mathbf{I})^3(\mathbf{v}_3) = \mathbf{0} \\ T(\mathbf{v}_4) &= \mathbf{v}_4 && \implies (T - 2\mathbf{I})(\mathbf{v}_4) = \mathbf{0} \end{aligned}$$

and so on. We see something interesting here. The four vectors are eliminated using different powers of $(T - 2\mathbf{I})$. However, all of them are eliminated by $(T - 2\mathbf{I})^3$. Hence if we let $K_2 = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$, we see that $K_2 = \ker(T - 2\mathbf{I})^3$. Recall that eigenspaces can be expressed as $\ker(T - \lambda\mathbf{I})$. We call subspaces like K_2 that are kernels of higher powers of $(T - \lambda\mathbf{I})$ a *generalized eigenspace*.

In fact, $V = \ker(T - 2\mathbf{I})^2 \oplus \ker(T - 3\mathbf{I})^2 \oplus \ker(T)^2$. All of them are T-invariant subspaces. This means that for more complicated vector spaces, we can find the restriction of T on each of these subspaces and T will behave in a very simple way on them. □

6 Inner product spaces

6.1 Dot product

The length of a vector $\mathbf{u} \in \mathbb{R}^3$, denoted $\|\mathbf{u}\|$, is given by Pythagoras' theorem. The familiar dot product is also defined as $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$ where θ is the angle between the two vectors. Using the cosine rule on the triangle with two sides formed by the two vectors, we get

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$$

and rearranging,

$$\mathbf{u} \cdot \mathbf{v} = \frac{1}{2}\left(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2\right)$$

If we have the coordinates of the two vectors, say $\mathbf{u} = (a_u, b_u, c_u)$ and $\mathbf{v} = (a_v, b_v, c_v)$, plugging them in and simplifying gives us

$$\mathbf{u} \cdot \mathbf{v} = a_u a_v + b_u b_v + c_u c_v$$

This is a familiar relationship but it highlights a few properties of the dot product. It relates length, angle and distance, and is surprisingly easy to compute. This serves as our motivation to study this further.

So far we have been in $\mathbb{R}^3$ where the idea of length and distance is easily conceptualized. However for other vector spaces, can we define such notions? Furthermore, can we transfer our geometrical ideas to such spaces?

Above we know what is length, distance and angle, and we used that to define the dot product. However, in a general vector space we have no physical interpretation of lengths and angles, hence we will first define a "dot product" and use that to investigate length, distance, and angle.

6.2 Inner product

We assume in this section that our vector spaces are either real or complex since we are trying to imitate $\mathbb{R}^3$ for now.

Definition 6.1 (Inner product)**.** Let V be a vector space over $\mathbb{F}$ where $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. An *inner product* on V is a function

$$(\mathbf{u}, \mathbf{v}) \to \langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{F}$$

such that

- (IP1) $\forall \mathbf{u}, \mathbf{v} \in V, \langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$ where the bar denotes complex conjugation.
- (IP2) $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V, \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
- (IP3) $\forall \mathbf{u}, \mathbf{v} \in V, \forall \alpha \in \mathbb{F}, \langle \alpha\mathbf{u}, \mathbf{w} \rangle = \alpha \langle \mathbf{u}, \mathbf{w} \rangle$.
- (IP4) $\forall \mathbf{v} \in V, \mathbf{v} \neq \mathbf{0} \implies \langle \mathbf{v}, \mathbf{v} \rangle > 0$. Also $\langle \mathbf{0}, \mathbf{0} \rangle = 0$. □

The necessity for complex conjugate in (IP1) is demanded by (IP4). These set of rules will also help us define lengths. We will see them later below. We follow with a few observations about the inner product.

If V is a real vector space, then $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$. The inner product is symmetric.

We can also combine (IP2) and (IP3) into

$$\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{u}, \mathbf{w} \rangle + \beta \langle \mathbf{u}, \mathbf{w} \rangle$$

for some $\alpha, \beta \in \mathbb{F}$. We say that the inner product is *linear in the first variable.* Only when the field is $\mathbb{R}$ does the inner product become linear in the second variable as well:

$$\begin{aligned} \langle \mathbf{w}, \alpha\mathbf{u} + \beta\mathbf{v} \rangle &= \overline{\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle} \\ &= \overline{\alpha \langle \mathbf{u}, \mathbf{w} \rangle + \beta \langle \mathbf{v}, \mathbf{w} \rangle} \\ &= \overline{\alpha} \langle \mathbf{w}, \mathbf{u} \rangle + \overline{\beta} \langle \mathbf{w}, \mathbf{v} \rangle \end{aligned}$$

We say that the inner product is *conjugate linear in the second variable.*

For any $\mathbf{v} \in V$, $\langle \mathbf{0}, \mathbf{v} \rangle = \mathbf{0}$. This is quite simple to show:

$$\langle \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{0} + \mathbf{0}, \mathbf{v} \rangle = 2 \langle \mathbf{0}, \mathbf{v} \rangle$$

Definition 6.2. An *inner product space* is a real or complex vector space V together with an inner product $\langle \cdot, \cdot \rangle$. $\square$

Example 6.1. $\mathbb{R}^n$ with $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots + u_n v_n$, called the dot product or Euclidean inner product, is an inner product space. $\square$

Example 6.2. In $\mathbb{C}^n$, $\langle, \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = u_1 \overline{v}_1 + \cdots + u_n \cdot \overline{v}_n$. Now we see why (IP1) requires the complex conjugate – it guarantees us $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ since $z\overline{z} = |z|^2 \geq 0$. $\square$

Example 6.3. Let $V = M_n(\mathbb{F})$ where $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. For $A, B \in V$, we can define

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

where B^* is the conjugate transpose of B. This is an inner product on V. If we let $A = (a_{ij})$ and $B = (b_{ij})$, then

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

$\square$

Example 6.4. Let $V = C[0, 1]$, the space of all continuous real-valued functions on the interval $[0, 1]$. For $f, g \in V$, define

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

This is an inner product on V. $\square$

6.3 Norm and distance

Recall that $\mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2$. Similarly we have the following definitions:

Definition 6.3. For an inner product space V,

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$$

is called the *norm* of $\mathbf{u}$.

The *distance* between $\mathbf{u}, \mathbf{v} \in V$ is defined by

$$\operatorname{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

$\square$

In $\mathbb{R}^3$, since $0 \leq \cos\theta \leq 1$, we know that $\mathbf{u} \cdot \mathbf{v} \leq \|\mathbf{u}\| \|\mathbf{v}\|$. In fact, we can generalize this fact to any inner product space.

Theorem 6.1 (Cauchy-Schwarz inequality)**.** *Let there be an inner product space V, and let $\mathbf{u}, \mathbf{v} \in V$. Then*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

Equality holds iff $\mathbf{u} = c\mathbf{v}$ for some scalar c.

Proof. If $\mathbf{u} = \mathbf{0}$, then $|\langle \mathbf{u}, \mathbf{v} \rangle| = 0 = \|\mathbf{u}\|\|\mathbf{v}\|$.

Now assume $\mathbf{u} \neq \mathbf{0}$. Let $\mathbf{w} = \mathbf{u} - c\mathbf{w}$ where $c = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{v}\|^2}$. Then

$$\begin{aligned} \|w\|^2 = \langle \mathbf{w}, \mathbf{w} \rangle &= \langle \mathbf{u} - c\mathbf{v}, \mathbf{u} - c\mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle - \langle \mathbf{u}, c\mathbf{v} \rangle - \langle c\mathbf{v}, \mathbf{u} \rangle + \langle c\mathbf{v}, c\mathbf{v} \rangle \\ &= \|\mathbf{u}\|^2 - \overline{c}\langle \mathbf{u}, \mathbf{v} \rangle - c\overline{\langle \mathbf{u}, \mathbf{v} \rangle} + c\overline{c}\|\mathbf{v}\|^2 \\ &= \|\mathbf{u}\|^2 - \frac{\overline{\langle \mathbf{u}, \mathbf{v} \rangle}}{\|\mathbf{v}\|^2}\langle \mathbf{u}, \mathbf{v} \rangle - \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{v}\|^2}\overline{\langle \mathbf{u}, \mathbf{v} \rangle} - \frac{\overline{\langle \mathbf{u}, \mathbf{v} \rangle}}{\|\mathbf{v}\|^2}\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{v}\|^2}\|\mathbf{v}\|^2 \\ &= \|\mathbf{u}\|^2 - \frac{|\langle \mathbf{u}, \mathbf{v} \rangle|^2}{\|\mathbf{v}\|^2} \geq 0. \end{aligned}$$

After some rearrangement,

$$|\langle \mathbf{u}, \mathbf{v} \rangle|^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2.$$

Retracing our steps, equality happens iff $\|\mathbf{w}\| = \mathbf{0}$ which means $\mathbf{u} = c\mathbf{v}$. ■

Theorem 6.2 (Triangle inequality)**.** *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are two vectors in an inner product space* V*,*

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

Proof.

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

■

6.4 Orthogonal sets

For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, we have a notion of orthogonality: $\mathbf{u} \cdot \mathbf{v} = 0 \iff \mathbf{u} \perp \mathbf{v}$. We can generalize this notion.

Definition 6.4. Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space.

- We say two vectors $\mathbf{u}, \mathbf{v} \in V$ are *orthogonal* and write $\mathbf{u} \perp \mathbf{v}$ if $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{0}$.
- A set S of vectors in V is called an *orthogonal set* if all pairs of distinct vectors in S are orthogonal.
- If every vector in an orthogonal set S of vectors in V has norm 1, then S is called an *orthonormal set.* □

Theorem 6.3 (Pythagoras' theorem)**.** *If* $\mathbf{u}, \mathbf{v}$ *are orthogonal vectors in an inner product space* V*, then*

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

Proof.

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

■

Theorem 6.4. *An orthogonal set of non-zero vectors is linearly independent.*

Proof. Let S be an orthogonal set in an inner product space V. Note that S is not assumed to be finite. Pick a finite subset of distinct vectors from S, $\mathbf{v}_1, \ldots, \mathbf{v}_k$ and consider the equation

$$a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = \mathbf{0}$$

for scalars $a_1, \ldots, a_k$.

For $1 \leq j \leq k$, we can take the inner product on both sides,

$$\begin{aligned} \langle a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k \rangle &= 0 \\ a_1 \cancelto{0}{\langle \mathbf{v}_1, \mathbf{v}_j \rangle} + \cdots + a_j \langle \mathbf{v}_j, \mathbf{v}_j \rangle + \cdots + a_k \cancelto{0}{\langle \mathbf{v}_k, \mathbf{v}_j \rangle} &= 0 \end{aligned}$$

Since the norm is non-zero, we conclude that all $a_j = 0$. ■

Corollary 6.4.1. *If* V *is a finite dimensional inner product space and* $n = \dim V$*, then any orthogonal set of non-zero vectors in* V *is finite and contains at most* n *vectors.*

6.5 Orthonormal bases

Definition 6.5. Let V be an inner product space. A basis B of V is called an *orthonormal basis* if B is also an orthonormal set in V. □

From our experience in $\mathbb{R}^3$, it is very easy to compute the coordinates of a vector relative to an orthonormal basis. Similarly, the calculation of a matrix of a linear operator is also much easier.

Lemma 6.5. *Let* $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *be an orthonormal basis of* V*. Then for any* $\mathbf{v} \in V$*,*

$$\mathbf{v} = \langle \mathbf{v}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \cdots + \langle \mathbf{v}, \mathbf{v}_n \rangle \mathbf{v}_n$$

Proof. Suppose $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$. Then

$$\begin{aligned} \langle \mathbf{v}, \mathbf{v}_j \rangle &= \langle a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n, \mathbf{v}_j \rangle \\ &= a_1 \cancelto{0}{\langle \mathbf{v}_1, \mathbf{v}_j \rangle} + \cdots a_j \langle \mathbf{v}_j, \mathbf{v}_j \rangle + \cdots a_n \cancelto{0}{\langle \mathbf{v}_n, \mathbf{v}_j \rangle} \\ &= a_j \end{aligned}$$

■

Corollary 6.5.1. *Let $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an orthonormal basis of V, and let $T : V \to V$ be a linear operator and $[T]_B = (a_{ij})$. Then*

$$a_{ij} = \langle T(\mathbf{v}_j), \mathbf{v}_i \rangle$$

Proof. From the definition of $[T]_B$, $T(\mathbf{v}_j) = a_{1j}\mathbf{v}_1 + \cdots + a_{nj}\mathbf{v}_n$. By Lemma 6.5, each $a_{ij} = \langle T(\mathbf{v}_j), \mathbf{v}_i \rangle$. ■

How can we construct an orthonormal basis? What we want to achieve is to take any basis $A = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and come up with $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ such that B is an orthonormal basis and $\operatorname{span} A = \operatorname{span} B$.

We introduce the Gram-Schmidt process.

Theorem 6.6 (Gram-Schmidt process)**.** *Given $A = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, we can construct an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ for* $\operatorname{span} A$ *with the following algorithm:*

1. *Choose*
$$\mathbf{v}_i = \frac{1}{\|\mathbf{u}_1\|}\mathbf{u}_1$$
2. *Let $\mathbf{v}_2' = \mathbf{u}_2 - \langle \mathbf{u}_2, \mathbf{v}_1 \rangle \mathbf{v}_1$. Now define $\mathbf{v}_2 = \frac{\mathbf{v}_2'}{\|\mathbf{v}_2'\|}$.*
3. *Suppose that we have already found an orthonormal set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ for some $1 \le k \le n-1$ and $\operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} = \operatorname{span} S$. Then let*
$$\mathbf{v}_{k+1}' = \mathbf{u}_{k+1} - \sum_{j=1}^{k} \langle \mathbf{u}_{k+1}, \mathbf{v}_j \rangle \mathbf{v}_j$$
4. *Repeat until we get n orthonormal vectors.*

Proof. The construction is an inductive one. $\{\mathbf{v}_1\}$ is an orthonormal set, and $\operatorname{span}\{\mathbf{v}_1\} = \operatorname{span}\{\mathbf{u}_1\}$.

$\mathbf{v}_2' \perp \mathbf{v}_1$:

$$\begin{aligned}\langle \mathbf{v}_2', \mathbf{v}_1 \rangle &= \langle \mathbf{u}_2 - \langle \mathbf{u}_2, \mathbf{v}_1 \rangle \mathbf{v}_1, \mathbf{v}_1 \rangle \\ &= \langle \mathbf{u}_2, \mathbf{v}_1 \rangle - \langle \mathbf{u}_2, \mathbf{v}_1 \rangle \langle \mathbf{v}_1, \mathbf{v}_1 \rangle \\ &= 0\end{aligned}$$

Thus $\{\mathbf{v}_1, \mathbf{v}_2\}$ is orthonormal and $\operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2\} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$, since $\mathbf{v}_2'$ is only a linear combination of $\mathbf{v}_1$ and $\mathbf{u}_2$.

For $1 \le i \le k$,

$$\begin{aligned}\langle \mathbf{v}_{k+1}', \mathbf{v}_i \rangle &= \langle \mathbf{u}_{k+1}, \mathbf{v}_i \rangle - \sum_{j=1}^{k} \langle \mathbf{u}_{k+1}, \mathbf{v}_j \rangle \underbrace{\langle \mathbf{v}_j, \mathbf{v}_i \rangle}_{\delta_{ij}} \\ &= \langle \mathbf{u}_{k+1}, \mathbf{v}_i \rangle - \langle \mathbf{u}_{k+1}, \mathbf{v}_i \rangle \\ &= 0\end{aligned}$$

So $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$ is an orthonormal set, and with a similar reasoning as above, we can show that $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\} = \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_{k+1}\}$. ■

The formulas seem to be plucked out of thin air, but we shall see how they are chosen later on.

Corollary 6.6.1. *Every finite dimensional inner product space has an orthonormal basis.*

Example 6.5. Take the space of all real polynomials, $P(\mathbb{R})$. Define the inner product for any $p, q \in P(\mathbb{R})$ as

$$\langle p, q\rangle = \int_{-1}^{1} p(t)q(t)\,\mathrm{d}t$$

Performing the Gram-Schmidt process on the basis $\{1, x, x^2, \ldots\}$ gives us the Legendre polynomials. To list a few:

$$P_0(x) = 1 \qquad P_1(x) = x \qquad P_2(x) = \frac{1}{2}(3x^2 - 1) \qquad P_3(x) = \frac{1}{2}(5x^3 - 3x)$$

□

What happens if we apply the Gram-Schmidt process to a linearly dependent set?

6.6 Orthogonal complements

Definition 6.6. Let V be an inner product space and W a subspace of V. The *orthogonal complement* of W, denoted $W^\perp$ is the set

$$W^\perp = \{\mathbf{v} \in V \mid \forall \mathbf{w} \in W : \langle \mathbf{v}, \mathbf{w}\rangle = 0\}$$

□

Example 6.6. In $\mathbb{R}^2$, if W is the x-axis then $W^\perp$ is the y-axis. □

Example 6.7. In $\mathbb{R}^3$, if W is the x-axis, then $W^\perp$ is the yz-plane. □

Lemma 6.7. *Let W be a subspace of an inner product space V. Then*

i. $W^\perp$ is a subspace of V.

ii. $W \cap W^\perp = \{\mathbf{0}\}$

Proof.

i. $\mathbf{0} \in W^\perp$ since $\langle \mathbf{0}, \mathbf{w}\rangle = 0$ for any $\mathbf{w} \in W$. Take $\mathbf{w}_1, \mathbf{w}_2 \in W^\perp$, any any scalars a_1, a_2, then

$$\begin{aligned}\langle a_1\mathbf{w}_1, a_2\mathbf{w}_2\rangle &= a_1\langle \mathbf{w}_1, \mathbf{w}_2\rangle + a_2\langle \mathbf{w}_1, \mathbf{w}_2\rangle \\ &= 0\end{aligned}$$

So $a_1\mathbf{w}_1 + a_2\mathbf{w}_2 \in W^\perp$. By Thm. 3.2, $W^\perp$ is a subspace.

ii. Take any $\mathbf{v} \in W \cap W^\perp$. Then

$$\begin{aligned}\|\mathbf{v}\|^2 &= \langle \underbrace{\mathbf{v}}_{\in W}, \underbrace{\mathbf{v}}_{\in W^\perp}\rangle \\ &= 0\end{aligned}$$

Therefore $\mathbf{v} = \mathbf{0}$.

■

Theorem 6.8. *If W is a finite dimensional subspace of an inner product space V, then*

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

Proof. From Lemma 6.7, $W \cap W^{\perp} = \{\mathbf{0}\}$. Using Thm. 3.16, we only need to show that $V = W + W^{\perp}$.

Take any $\mathbf{v} \in V$ and let $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ be an orthonormal basis for W. Define a vector $\mathbf{w}$ as

$$W \ni \mathbf{w} = \sum_{i=1}^{k} \langle \mathbf{v}, \mathbf{w}_i \rangle \mathbf{w}_i$$

Consider $\mathbf{v} - \mathbf{w}$ and its inner product with any of the basis elements:

$$\begin{aligned}
\langle \mathbf{v} - \mathbf{w}, \mathbf{w}_j \rangle &= \langle \mathbf{v}, \mathbf{w}_j \rangle - \langle \mathbf{w}_i, \mathbf{w}_j \rangle \\
&= \langle \mathbf{v}, \mathbf{w}_j \rangle - \sum_{i=1}^{k} \langle \mathbf{v}, \mathbf{w}_i \rangle \underbrace{\langle \mathbf{w}_i, \mathbf{w}_j \rangle}_{\delta_{ij}} \\
&= 0
\end{aligned}$$

Any vector $\mathbf{w} \in W$ can be expressed as a linear combination of $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$. It follows that $\mathbf{v} - \mathbf{w} \in W^{\perp}$. Then $\mathbf{w} = \underbrace{\mathbf{w}}_{\in W} + (\underbrace{\mathbf{v} - \mathbf{w}}_{\in W^{\perp}})$. ■

Thm. 6.8 is not true if W is not finite dimensional.

6.7 Orthogonal projections

Definition 6.7. Let W be a finite dimensional subspace of an inner product space V. Then

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

For every $\mathbf{v} \in V$, there is an unique $\mathbf{w} \in W$ and $\mathbf{w}' \in W^{\perp}$ such that $\mathbf{v} = \mathbf{w} + \mathbf{w}'$. Define a linear operator $\text{proj}_W : W \to W$, called the *orthogonal projection* of V on W as

$$\text{proj}_W(\mathbf{v}) = \mathbf{w}$$

□

Let $T = \text{proj}_W$. Then $T^2 = T$, $\mathcal{R}(T) = W$, and $\ker(T) = W^{\perp}$. Orthogonal projections are associated with inner product spaces. They may not exist for any arbitrary vector space. Also, they are associated to the direct sum decomposition $V = W \oplus W^{\perp}$.

Look at the proof Thm. 6.8. If $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ is an orthonormal basis of W, then we have an explicit formula for the orthogonal projection operator:

$$\text{proj}_W(\mathbf{v}) = \mathbf{w} = \sum_{i=1}^{k} \langle \mathbf{v}, \mathbf{w}_i \rangle \mathbf{w}_i$$

Also note that it was not assumed that V is finite dimensional. $W^\perp$ may not finite dimensional. However, if V is finite dimensional, then $W^\perp$ is also finite dimensional, and we may talk about $\text{proj}_{W^\perp}$, and it is given by

$$\text{proj}_{W^\perp} = \mathbf{w}' = \mathbf{v} - \mathbf{w} = (\mathbf{I}_V - \text{proj}_W)(\mathbf{v})$$

Let us revisit the Gram-Schmidt process, Thm. 6.6. Let $W_0 = \{\mathbf{0}\}$ and for every $1 \leq j \leq n-1$, let

$$W_j = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_j\}.$$

For some $1 \leq k \leq n$, suppose we have already found $k-1$ vectors that makes up our orthonormal basis. We can find another vector that is orthogonal to all the $k-1$ vectors we already have by projecting $\mathbf{u}_k$ on the orthogonal complement of the subspace spanned by the $k-1$ vectors.

$$\mathbf{v}_j' = \text{proj}_{W_{k-1}^\perp}(\mathbf{u}_k)$$

And we see that after normalization, this gives us the formula for the Gram-Schmidt process.

6.8 Best approximations

In $\mathbb{R}^3$, given a plane W and a point P, the point on W closest to P is given by drawing a perpendicular line from P to W. In vector notation, if we let $\mathbf{v} = \vec{OP}$, the perpendicular line from P to W is given by $\mathbf{v} - \text{proj}_W(\mathbf{v})$. For any other point Q on W, with $\mathbf{w} = \vec{OQ}$, $\|\mathbf{v} - \text{proj}_W(\mathbf{v})\| < \|\mathbf{v} - \mathbf{w}\|$. We say that $\text{proj}_W(\mathbf{v})$ is the best approximation to $\mathbf{v}$ by vectors in W, since it is the closest we can get to $\mathbf{v}$ while still being in W. We can generalize this to other inner product spaces.

Theorem 6.9. *If W is a finite dimensional subspace of an inner product space V and $\mathbf{v} \in V$, then*

$$\|\mathbf{v} - \text{proj}_W(\mathbf{v})\| < \|\mathbf{v} - \mathbf{w}\|$$

for every vector $\mathbf{w} \in W$ such that $\mathbf{w} \neq \text{proj}_W(\mathbf{v})$.

Proof. Let there be a $\mathbf{w} \in W$ such that $\mathbf{w} \neq \text{proj}_W(\mathbf{v})$. Then $\|\text{proj}_W(\mathbf{v}) - \mathbf{w}\| > 0$.

Observe that

$$\mathbf{v} - \mathbf{w} = \underbrace{\mathbf{v} - \text{proj}_W(\mathbf{v})}_{\in W^\perp} + \underbrace{\overbrace{\text{proj}_W(\mathbf{v})}^{\in W} - \overbrace{\mathbf{w}}^{\in W}}_{\in W}$$

Now we can apply Pythagoras' theorem, and

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v} - \text{proj}_W(\mathbf{v})\|^2 + \|\text{proj}_W(\mathbf{v}) - \mathbf{w}\|^2 > \|\mathbf{v} - \text{proj}_W(\mathbf{v})\|^2$$

■

Now consider the linear system $A\mathbf{x} = \mathbf{y}$. If it is consistent, then there is a solution $\mathbf{x}$ such that $A\mathbf{x} - \mathbf{y} = \mathbf{0}$, and $\|A\mathbf{x} - \mathbf{y}\| = 0$. However, if there is no solution, then can we find some $\mathbf{x}'$ such that we minimize $\|A\mathbf{x}' - \mathbf{y}\|$? We call $\mathbf{x}'$ the *least squares solution* of $A\mathbf{x} = \mathbf{y}$.

Let A contain the columns $\mathbf{a}_1, \dots, \mathbf{a}_n$ and $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$. So the space $W = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$ is the column space of A. Now we want to find $\mathbf{x}'$ such that $\|A\mathbf{x}' - \mathbf{y}\| \leq \|A\mathbf{x} - \mathbf{y}\|$. By Thm. 6.9, $A\mathbf{x}' = \text{proj}_W(\mathbf{y})$. This might be quiet a lengthy computation if we do not have an orthonormal basis for W.

We have an alternative approach. If $A\mathbf{x}' = \text{proj}_W(\mathbf{y})$, then $W^\perp \ni \mathbf{y} - \text{proj}_W(\mathbf{y}) = \mathbf{y} - A\mathbf{x}'$. Thus, $\forall \mathbf{v} \in \mathbb{R}^n, \langle A\mathbf{v}, \mathbf{y} - A\mathbf{x}' \rangle = 0$. We can express the inner product in $\mathbb{R}^n$ (the dot product) as a matrix multiplication instead: $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T\mathbf{v}$. Then

$$\begin{aligned} \langle A\mathbf{v}, \mathbf{b} - A\mathbf{x}' \rangle &= \mathbf{v}^T A^T(\mathbf{y} - A\mathbf{x}') \\ &= \langle \mathbf{v}, A^T(\mathbf{y} - A\mathbf{x}') \rangle \end{aligned}$$

This is only true if $A^T(\mathbf{y} - A\mathbf{x}') = \mathbf{0}$. Hence $A^T\mathbf{y} = A^TA\mathbf{x}'$, and we only need to solve this much simpler problem. We call this the *associated normal system* of $A\mathbf{x} = \mathbf{y}$. We arrive at the following theorem.

Theorem 6.10. *For any real linear system $A\mathbf{x} = \mathbf{y}$, the associated normal system*

$$A^TA\mathbf{x}' = A^T\mathbf{y}$$

is consistent, and all its solutions are least squares solutions of $A\mathbf{x} = \mathbf{y}$.

We may apply this to least squares analysis. Given a set of data points $(x_1, y_1), \dots, (x_n, y_n)$, we want to find a straight line $y = ax + b$ that minimizes the sum of square of errors $E = \sum_{i=1}^n (y_i - ax_i - b)^2$. In other words, we are finding least squares solution of $A\mathbf{x} = \mathbf{y}$, where

$$A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

and $E = \|A\mathbf{x} - \mathbf{y}\|^2$.

6.9 Adjoint of a linear operator

Lemma 6.11. *Let V be a finite dimensional inner product space over $\mathbb{F}$. If $f : V \to \mathbb{F}$ is a linear functional, then there exists a unique vector $\mathbf{u} \in V$ such that $\forall \mathbf{v} \in V$,*

$$f(\mathbf{u}) = \langle \mathbf{v}, \mathbf{u} \rangle$$

Proof. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be an orthonormal basis for V. Then for $\mathbf{v} \in V$,

$$\mathbf{v} = \sum_{i=1}^n \langle \mathbf{v}, \mathbf{v}_i \rangle \mathbf{v}_i$$

and therefore

$$
\begin{aligned}
f(\mathbf{v}) &= \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{v}_i \rangle f(\mathbf{v}_i) \\
&= \langle \mathbf{v}, \sum_{i=1}^{n} \overline{f(\mathbf{v}_i)} \mathbf{v}_i \rangle
\end{aligned}
$$

And we can let $\mathbf{u} = \sum_{i=1}^{n} \overline{f(\mathbf{v}_i)} \mathbf{v}_i$.

Suppose there is another vector $\mathbf{u}' \in V$ such that $f(\mathbf{v}) = \langle \mathbf{v}, \mathbf{u}' \rangle$, then

$$
\begin{aligned}
\|\mathbf{u} - \mathbf{u}', \mathbf{u} - \mathbf{u}'\|^2 &= \langle \mathbf{u} - \mathbf{u}', \mathbf{u} - \mathbf{u}' \rangle \\
&= \langle \mathbf{u} - \mathbf{u}', \mathbf{u} \rangle - \langle \mathbf{u} - \mathbf{u}', \mathbf{u}' \rangle \\
&= f(\mathbf{u} - \mathbf{u}') - f(\mathbf{u} - \mathbf{u}') \\
&= 0
\end{aligned}
$$

So $\mathbf{u} = \mathbf{u}'$. ■

Theorem 6.12. *Let $T : V \to V$ be a linear operator on a finite dimensional inner product space V. Then there exists an unique linear operator $T^* : V \to V$ such that*

$$\langle T(\mathbf{u}), \mathbf{v} \rangle = \langle \mathbf{u}, T^*(\mathbf{v}) \rangle$$

and we call T^* the adjoint *of T.*

Proof. Let $\mathbf{v} \in V$, and define the linear functional $f : V \to \mathbb{F}$ as

$$f(\mathbf{u}) = \langle T(\mathbf{u}), \mathbf{v} \rangle$$

for any $\mathbf{u} \in V$. By Lem. 6.11 there is an unique vector $\mathbf{v}' \in V$ such that $f(\mathbf{u}) = \langle \mathbf{u}, \mathbf{v}' \rangle$. Now let us set $T^*(\mathbf{v}) = \mathbf{v}'$. It can be verified that T^* is a linear operator. Also,

$$\langle \mathbf{u}, T^*(\mathbf{v}) \rangle = \langle \mathbf{u}, \mathbf{v}' \rangle = f(\mathbf{u}) = \langle T(\mathbf{u}), \mathbf{v} \rangle$$

Suppose there is another linear operator S that fulfils

$$\langle T(\mathbf{u}), \mathbf{v} \rangle = \langle \mathbf{u}, S(\mathbf{v}) \rangle$$

Then $\langle \mathbf{u}, S(\mathbf{v}) \rangle = \langle \mathbf{u}, T^*(\mathbf{v}) \rangle$. Since this is true for all $\mathbf{u}$,

$$
\begin{aligned}
\langle S(\mathbf{v}), S(\mathbf{v}) \rangle &= \langle S(\mathbf{v}), T^*(\mathbf{v}) \rangle \\
\langle S(\mathbf{v}), S(\mathbf{v}) - T^*(\mathbf{v}) \rangle &= 0
\end{aligned}
$$

so $S(\mathbf{v}) = T^*(\mathbf{v})$ for all $\mathbf{v}$.

■

Theorem 6.13. *Let $T : V \to V$ be a linear operator on a finite dimensional inner product space V and $\mathcal{B}$ be an orthonormal basis of V. Then*

$$[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}^*$$

where $[T]_{\mathcal{B}}^*$ denotes the conjugate transpose of $[T]_{\mathcal{B}}$.

Proof. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, $[T]_{\mathcal{B}} = (a_{ij})$, and $[T]^*_{\mathcal{B}} = (b_{ij})$. From Cor. 6.5.1,

$$\begin{aligned} b_{ij} &= \langle T^*(\mathbf{v}_j), \mathbf{v}_i \rangle \\ &= \overline{\langle \mathbf{v}_i, T^*(\mathbf{v}_j) \rangle} \\ &= \overline{\langle T(\mathbf{v}_i), \mathbf{v}_j \rangle} \\ &= \overline{a_{ij}} \end{aligned}$$

■

Theorem 6.14. *Let T, T_1, T_2 be linear operators on a finite dimensional inner product space V. Then*

i. $(T_1 + T_2)^* = T_1^* + T_2^*$.

ii. $(cT)^* = \overline{c}T^*$.

iii. $(T_1T_2)^* = T_2^*T_1^*$.

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

Proof. Let $\mathcal{B}$ be an orthonormal basis of V, and any vectors $\mathbf{u}, \mathbf{v} \in V$.

i. $[(T_1 + T_2)^*]_{\mathcal{B}} = [T_1 + T_2]^*_{\mathcal{B}} = [T_1]^*_{\mathcal{B}} + [T_2]^*_{\mathcal{B}}$.

ii. $\langle \mathbf{u}, (cT)^*(\mathbf{v}) \rangle = \langle cT(\mathbf{u}), \mathbf{v} \rangle = c\langle T(\mathbf{u}), \mathbf{v} \rangle = c\langle \mathbf{u}, T^*(\mathbf{v}) \rangle = \langle \mathbf{u}, \overline{c}T^*(\mathbf{v}) \rangle$.

iii. $\langle \mathbf{u}, (T_1T_2)^*(\mathbf{v}) \rangle = \langle T_1T_2(\mathbf{u}), \mathbf{v} \rangle = \langle T_2(\mathbf{u}), T_1^*(\mathbf{v}) \rangle = \langle \mathbf{u}, T_2^*(T_1^*(\mathbf{v})) \rangle$

iv. $\langle T(\mathbf{u}), \mathbf{v} \rangle = \langle \mathbf{u}, T^*(\mathbf{v}) \rangle = \langle (T^*)^*(\mathbf{u}), \mathbf{v} \rangle$

■

6.10 Normal and self-adjoint operators

Recall that a linear operator $T : V \to V$ is diagonalizable if V has a basis $\mathcal{B}$ consisting of eigenvectors of T. Then $[T]_{\mathcal{B}}$ is a diagonal matrix with diagonal entries corresponding to the eigenvalues of T.

If V is an inner product space over $\mathbb{R}$ or $\mathbb{C}$, then we may ask if $\mathcal{B}$ is orthonormal. If such an orthonormal basis exists, then we say that T is *orthogonally diagonalizable.* Now which linear operators on an inner product space are orthogonally diagonalizable? We will answer this question in this section.

Definition 6.8. A linear operator $T : V \to V$ on a finite dimensional inner product space V is called *self-adjoint* if $T = T^*$. □

Lemma 6.15. *All eigenvalues of a self-adjoint operator on a finite dimensional complex vector space are real.*

Proof. Let $T : V \to V$ be a self-adjoint operator on the complex inner product space V , and $\mathbf{v}$ be an eigenvector of T corresponding to the eigenvalue λ. Then

$$\lambda\langle \mathbf{v}, \mathbf{v}\rangle = \langle T(\mathbf{v}), \mathbf{v}\rangle = \langle \mathbf{v}, T(\mathbf{v})\rangle = \overline{\lambda}\langle \mathbf{v}, \mathbf{v}\rangle$$

Since $\mathbf{v} \neq 0$, $\langle \mathbf{v}, \mathbf{v}\rangle \neq 0$ and the result follows. ■

Definition 6.9. Let $T : V \to V$ be a self-adjoint operator on a finite dimensional inner product space V over a field $\mathbb{F}$. If $\mathcal{B}$ is an orthonormal basis of V and $A = [T]_{\mathcal{B}}$, then

- If $\mathbb{F} = \mathbb{R}$, $A = A^* = A^T$ and A is symmetric.
- If $\mathbb{F} = \mathbb{R}$, then $A = A^*$ and we call A a *Hermitian* matrix.

□

By Lemma 6.15, the eigenvalues of a Hermitian matrix are real.

Corollary 6.15.1. *A self-adjoint operator on a finite dimensional real inner product space has at least one eigenvalue.*

Proof. Let $T : V \to V$ be a self-adjoint operator on a real inner product space V. Let $\mathcal{B}$ be an orthonormal basis of V, and let $A = [T]_{\mathcal{B}}$. Then A is a real symmetric matrix.

However we may treat A as a complex matrix first, in which case it is Hermitian, and also all its complex eigenvalues (recall that they must exist) must be real. They form the roots of the characteristic polynomial of A, and hence the same roots are also the (real) eigenvalues of T. ■

Theorem 6.16. *A linear operator on a finite dimensional real inner product space is orthogonally diagonalizable iff it is self-adjoint.*

Proof.

($\Longrightarrow$) TODO

($\Longleftarrow$) Given T is self-adjoint, we want to show that T is orthogonally diagonalizable. We perform induction on $\dim V = n$. The statement is clearly true for $n = 1$.

Suppose the statement is true for some $n = k$. Consider the case for some where $\dim V = k+1$. Then by Cor. 6.15.1, T has an eigenvector $\mathbf{v}$ corresponding to some eigenvalue λ. Let $W = \operatorname{span}\{\mathbf{v}\}$.

We claim that $W^\perp$ is T-invariant. For any $\mathbf{u} \in W^\perp$,

$$\langle T(\mathbf{u}), \mathbf{v}\rangle = \langle \mathbf{u}, T(\mathbf{v})\rangle = \lambda\langle \mathbf{u}, \mathbf{v}\rangle = 0$$

so $T(\mathbf{u}) \in W^\perp$.

The restriction of T to $W^\perp$ is a linear operator $T_{W^\perp}$ on $W^\perp$, which is also self-adjoint. Since $\dim W^\perp = k$ (this can be shown from $V = W \oplus W^\perp$), by the inductive hypothesis $T_{W^\perp}$ is orthogonally diagonalizable. Thus $W^\perp$ has an orthonormal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ consisting of the eigenvectors of $T_{W^\perp}$. Thus $\{\mathbf{v}, \mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthonormal basis for V consisting of the eigenvectors of T. ■

Definition 6.10. A linear operator T on a finite dimensional inner product space is called *normal* if it commutes with its adjoint, i.e. $TT^* = T^*T$. □

Every self-adjoint operator is clearly normal, but a normal operator is not necessarily self-adjoint.

Theorem 6.17. *A linear operator on a finite dimensional complex inner product space is orthogonally diagonalizable iff it is normal.*

Proof. TODO ■

6.11 Unitary operators

Definition 6.11. A linear operator $T : V \to V$ on a finite dimensional inner product space V is called *unitary* if T is invertible and $T^{-1} = T^*$. □

Theorem 6.18. *Let $T : V \to V$ be a linear operator on a finite dimensional inner product space V. Then the following are equivalent:*

i. *T is unitary.*

ii. $\forall \mathbf{u}, \mathbf{v} \in V, \langle T(\mathbf{u}), T(\mathbf{v})\rangle = \langle \mathbf{u}, \mathbf{v}\rangle$.

iii. *T sends an orthonormal basis of V to an orthonormal basis.*

iv. $\forall \mathbf{v} \in V, \|T(\mathbf{v})\| = \|\mathbf{v}\|$

Proof. Let $\mathbf{v}, \mathbf{u} \in V$ and let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be an orthonormal basis of V.

($(i) \implies (ii)$) $\langle T(\mathbf{u}), T(\mathbf{v})\rangle = \langle \mathbf{u}, T^*(T(\mathbf{v}))\rangle = \langle \mathbf{u}, \mathbf{v}\rangle$.

($(ii) \implies (iii)$) Since T is invertible it is bijective and we are assured that $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$ is also a basis. All we need to check is orthogonality. For $1 \le i, j \le n$, it follows that $\langle T(\mathbf{v}_i), T(\mathbf{v}_j)\rangle = \langle \mathbf{v}_i, \mathbf{v}_j\rangle = \delta_{ij}$.

($(iii) \implies (iv)$) Since both $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$ are orthonormal bases, if $\mathbf{v} = \sum_{i=1}^n a_i \mathbf{v}_i$ and then $T(\mathbf{v}) = \sum_{i=1}^n a_i T(\mathbf{v}_i)$, we have that $\|\mathbf{v}\|^2 = \sum_{i=1}^n a_i^2 = \|T(\mathbf{v})\|^2$

($(iv) \implies (i)$) TODO ■

It might be noted that an unitary operator on a complex inner product space is orthogonally diagonalizable since it is normal. However, the same need not be true if it is on a real inner product space instead.

We may extend some of these notions to matrices. Notice that since $TT^* = T^*T\mathbf{I}_V$, we have that for an orthonormal basis $\mathcal{B}$ of V, $[TT^*]_{\mathcal{B}} = [T^*T]_{\mathcal{B}} = \mathbf{I}_n$. Hence if $A = [T]_{\mathcal{B}}$, $AA^* = A^*A = \mathbf{I}_n$.

Definition 6.12. Let A be an $n \times n$ matrix. Then

- A is called *orthogonal* if $AA^T = A^TA = \mathbf{I}_n$.

- A is called *unitary* if $AA^* = A^*A = \mathbf{I}_n$.

$\square$

Corollary 6.18.1. *Let $T : V \to V$ be an unitary operator on a finite dimensional inner product space V over a field $\mathbb{F}$. Let $\mathcal{B}$ be an orthonormal basis of V and let $A = [T]_{\mathcal{B}}$. Then,*

i. If $\mathbb{F} = \mathbb{R}$, A is a real orthogonal matrix.

ii. If $\mathbb{F} = \mathbb{C}$, A is an unitary matrix.

Let A be an $n \times n$ real matrix. Denote the columns of A as $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and consider the inner product of two columns, $\langle \mathbf{a}_i, \mathbf{a}_j \rangle = \sum_{k=1}^{n} a_{ki}a_{kj}$. Notice that if we let $B = A^TA = (b_{ij})$, then $b_{ij} = \langle \mathbf{a}_i, \mathbf{a}_j \rangle$.

From this we see that $A^TA = \mathbf{I}_n$ iff $\langle \mathbf{a}_i, \mathbf{a}_j \rangle = \delta_{ij}$, i.e. $\{\mathbf{a}_i, \ldots, \mathbf{a}_n\}$ is an orthonormal basis of $\mathbb{R}^n$. What about unitary matrices?