MAT 3144 Linear Algebra I

Joe Wells
Virginia Tech

Spring 2024[1]
Last Updated: April 28, 2024

---

[1] This courses is using Sheldon Axler's *Linear Algebra Done Right*, 4th Edition. The chapter/section titles have been retained, but otherwise internal numbering of theorems, examples, etc. will likely disagree with the course text.

Contents

Preface

Example 0.1: Fields (of characteristic 0)

Examples of fields, which we generally denote $\mathbb{K}$:

- [rational numbers]
$$\mathbb{Q} = \left\{ \frac{p}{q} \ : \ p, q \in \mathbb{Z},\ q \neq 0 \right\}$$
- [quadratic extension of $\mathbb{Q}$]
$$\mathbb{Q}(\sqrt{d}) = \left\{ a + b\sqrt{d} \ : \ a, b \in \mathbb{Q},\ d \in \mathbb{Z} \right\}$$
- [real numbers]
$$\mathbb{R}$$
- [complex numbers]
$$\mathbb{C} = \left\{ a + bi \ : \ a, b \in \mathbb{R}\, i^2 = -1 \right\}$$

Remark. Fields are much more general objects, but the above are sufficiently interesting for this course. For the duration of these notes, we restrict our attention only to (infinite) fields of characteristic 0 and leave the finite/nonzero characteristic cases to a course in abstract algebra/algebraic number theory.

Chapter 1

Vector Spaces

1B Definition of Vector Space

Definition: $\mathbb{K}$-vector space

Given a field $\mathbb{K}$, a **$\mathbb{K}$-vector space** is a set V of objects called **vectors**, endowed with two operations:

- **(vector) addition**, denoted $+$
- **scalar multiplication**, denoted (no symbol)

satisfying the following properties: For all vector $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ and for all **scalars** $k, \ell \in \mathbb{K}$:

1. [closure of addition] $\boldsymbol{u} + \boldsymbol{v} \in V$
2. [commutativity of addition] $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$
3. [associativity of addition] $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$
4. [existence of zero] There is some vector $\mathbf{0} \in V$, called the **zero vector** so that $\boldsymbol{v} + \mathbf{0} = \boldsymbol{v}$ for all $\boldsymbol{v} \in V$.
5. [existence of additive inverses] For each $\boldsymbol{v}$ in V, there is some vector $-\boldsymbol{v}$ for which $\boldsymbol{v} + (-\boldsymbol{v}) = \mathbf{0}$.
6. [closure of scalar multiplication] $k\boldsymbol{v} \in V$
7. [associativity of scalar multiplication] $(k\ell)\boldsymbol{v} = k(\ell\boldsymbol{v})$
8. [distributivity] $k(\boldsymbol{u} + \boldsymbol{v}) = k\boldsymbol{u} + k\boldsymbol{v}$
9. [distributivity] $(k + \ell)\boldsymbol{u} = k\boldsymbol{u} + \ell\boldsymbol{u}$
10. [existence of a multiplicative identity] If $1_{\mathbb{K}}$ is the multiplicative identity in $\mathbb{K}$, then $1_{\mathbb{K}}\boldsymbol{v} = \boldsymbol{v}$.

Remark. Unless otherwise specified, V will always refer to a $\mathbb{K}$-vector space.

Examples of Vector Spaces

Example 1B.1

Let $\mathbb{K}^n$ be the set of ordered n-tuples of $\mathbb{K}$ values. That is,

$$\mathbb{K}^n = \{(k_1, k_2, \ldots, k_n) \ : \ k_i \in \mathbb{K} \text{ for } i = 1, \ldots, n\}$$

Define vector addition via

$$(k_1, k_2, \ldots, k_n) + (k_1', k_2', \ldots, k_n') = (k_1 + k_1', k_2 + k_2', \ldots, k_n + k_n')$$

and scalar multiplication via

$$\lambda(k_1, k_2, \ldots, k_n) = (\lambda k_1, \lambda k_2, \ldots, \lambda k_n).$$

Show that $\mathbb{K}^n$ is a $\mathbb{K}$-vector space.

Proof. We verify each of the vector space axioms. Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{K}^n$ and $k, \ell \in \mathbb{K}$ be arbitrary.

1. [closure of addition] NOTES INCOMPLETE
2. [commutativity of addition] NOTES INCOMPLETE
3. [associativity of addition] NOTES INCOMPLETE
4. [existence of zero vector] NOTES INCOMPLETE
5. [existence of additive inverses] NOTES INCOMPLETE
6. [closure of scalar multipliction] NOTES INCOMPLETE
7. [associativity of scalar multiplication] NOTES INCOMPLETE
8. [distributivity] NOTES INCOMPLETE
9. [distributivity] NOTES INCOMPLETE
10. [multiplicative identity] NOTES INCOMPLETE

□

Example 1B.2: Space of $\mathbb{K}$-valued functions

Let S be any nonempty set, and let $\mathbb{K}^S$ denote the set of functions $f : S \to \mathbb{K}$. Define vector addition via

$$(f+g)(x) = f(x) + g(x) \qquad \text{where } f, g, \in \mathbb{K}^S,$$

and scalar multiplication via

$$(\lambda f)(x) = \lambda f(x) \qquad \text{where } f \in \mathbb{K}^s \text{ and } \lambda \in \mathbb{K},$$

for all $x \in S$. Show that $\mathbb{K}^S$ is a vector space.

Proof. We verify each of the vector space axioms. Let $f, g, h \in \mathbb{K}^S$ and $k, \ell \in \mathbb{K}$ be arbitrary.

1. [closure of addition] NOTES INCOMPLETE
2. [commutativity of addition] NOTES INCOMPLETE
3. [associativity of addition] NOTES INCOMPLETE
4. [existence of zero vector] NOTES INCOMPLETE
5. [existence of additive inverses] NOTES INCOMPLETE
6. [closure of scalar multipliction] NOTES INCOMPLETE
7. [associativity of scalar multiplication] NOTES INCOMPLETE
8. [distributivity] NOTES INCOMPLETE
9. [distributivity] NOTES INCOMPLETE
10. [multiplicative identity] NOTES INCOMPLETE

□

Example 1B.3: Space of polynomials, $\mathcal{P}(\mathbb{K})$

Let $\mathcal{P}(\mathbb{K})$ be the set of *polynomial functions on* $\mathbb{K}$, that is, the set

$$\mathcal{P}(\mathbb{K}) = \left\{ p \in \mathbb{K}^{\mathbb{K}} \ : \ \begin{array}{l} \exists m \in \mathbb{N} \text{ and } a_1, \dots, a_m \in \mathbb{K} \text{ s.t.} \\ p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m \text{ for all } x \in \mathbb{K} \end{array} \right\}$$

Define vector addition via

$$(p+q)(x) = p(x) + q(x) \qquad \text{where } p, q \in \mathcal{P}(\mathbb{K}),$$

and scalar multiplication via

$$(\lambda p)(x) = \lambda p(x) \qquad \text{where } p \in \mathcal{P}(\mathbb{K}) \text{ and } \lambda \in \mathbb{K},$$

for all $x \in \mathbb{K}$. Show that $\mathcal{P}(\mathbb{K})$ is a vector space.

We remark that this function addition is the usual polynomial addition.

Proof. We verify each of the vector space axioms. Let $p, q, r \in \mathcal{P}(\mathbb{K})$ and $\lambda, \mu \in \mathbb{K}$ be arbitrary.

1. [closure of addition] NOTES INCOMPLETE
2. [commutativity of addition] NOTES INCOMPLETE
3. [associativity of addition] NOTES INCOMPLETE
4. [existence of zero vector] NOTES INCOMPLETE
5. [existence of additive inverses] NOTES INCOMPLETE
6. [closure of scalar multipliction] NOTES INCOMPLETE
7. [associativity of scalar multiplication] NOTES INCOMPLETE
8. [distributivity] NOTES INCOMPLETE
9. [distributivity] NOTES INCOMPLETE
10. [multiplicative identity] NOTES INCOMPLETE

□

Example 1B.4: Space of degree-m polynomials, $\mathcal{P}_m(\mathbb{K})$

Fixing $m \in \mathbb{N}$, let $\mathcal{P}_m(\mathbb{K})$ be the set of *degree-m polynomial functions on* $\mathbb{K}$, that is, the set

$$\mathcal{P}_m(\mathbb{K}) = \left\{ p \in \mathbb{K}^{\mathbb{K}} \;:\; \begin{array}{l} \exists a_1, \ldots, a_m \in \mathbb{K} \text{ s.t.} \\ p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m \text{ for all } x \in \mathbb{K} \end{array} \right\}$$

Define vector addition via

$$(p+q)(x) = p(x) + q(x) \qquad \text{where } p, q \in \mathcal{P}_m(\mathbb{K}),$$

and scalar multiplication via

$$(\lambda p)(x) = \lambda p(x) \qquad \text{where } p \in \mathcal{P}_m(\mathbb{K}) \text{ and } \lambda \in \mathbb{K},$$

for all $x \in \mathbb{K}$. Show that $\mathcal{P}_m(\mathbb{K})$ is a vector space.

We remark that addition is the usual polynomial addition.

Proof. Fix $m \in \mathbb{N}$. We verify each of the vector space axioms. Let $p, q, r \in \mathcal{P}_m(\mathbb{K})$ and $\lambda, \mu \in \mathbb{K}$ be arbitrary.

1. [closure of addition] NOTES INCOMPLETE
2. [commutativity of addition] NOTES INCOMPLETE
3. [associativity of addition] NOTES INCOMPLETE

4. [existence of zero vector] NOTES INCOMPLETE
5. [existence of additive inverses] NOTES INCOMPLETE
6. [closure of scalar multipliction] NOTES INCOMPLETE
7. [associativity of scalar multiplication] NOTES INCOMPLETE
8. [distributivity] NOTES INCOMPLETE
9. [distributivity] NOTES INCOMPLETE
10. [multiplicative identity] NOTES INCOMPLETE

☐

Example 1B.5: The $(n-1)$-simplex

Let Δ^{n-1} be the following set of ordered n-tuples of *positive* real numbers:

$$\Delta^n = \left\{ (x_1, x_2, \dots, x_n) \;:\; \begin{array}{l} x_i \in \mathbb{R}, x_i > 0 \text{ for } i = 1, \dots, n+1 \\ \text{and } x_1 + x_2 + \cdots + x_n = 1 \end{array} \right\}.$$

Define vector addition via

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = \frac{(x_1 y_1, \dots, x_n y_n)}{\sum_{i=1}^n x_i y_i}$$

and scalar multiplication via

$$\lambda(x_1, \dots, x_n) = \frac{(x_1^\lambda, \dots, x_n^\lambda)}{\sum_{i=1}^n x_i^\lambda}.$$

Show that Δ^{n-1} is an $\mathbb{R}$-vector space.

Proof. Fix an integer $n > 0$. We verify each of the vector space axioms. Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \Delta^{n-1}$ and $\lambda, \mu \in \mathbb{R}$ be arbitrary.

1. [closure of addition] NOTES INCOMPLETE
2. [commutativity of addition] NOTES INCOMPLETE
3. [associativity of addition] NOTES INCOMPLETE
4. [existence of zero vector] NOTES INCOMPLETE
5. [existence of additive inverses] NOTES INCOMPLETE
6. [closure of scalar multipliction] NOTES INCOMPLETE
7. [associativity of scalar multiplication] NOTES INCOMPLETE
8. [distributivity] NOTES INCOMPLETE
9. [distributivity] NOTES INCOMPLETE
10. [multiplicative identity] NOTES INCOMPLETE

☐

Proposition 1B.6: Unique additive identity

Let V be a $\mathbb{K}$-vector space. The following properties hold.

1. The zero vector, $\mathbf{0} \in V$, is unique.
2. For every $\boldsymbol{v} \in V$, the additive inverse is unique.
3. For every $\boldsymbol{v} \in V$, $0\boldsymbol{v} = \mathbf{0}$.

4. For every $\lambda \in \mathbb{K}$, $\lambda\mathbf{0} = \mathbf{0}$.
5. For every $\boldsymbol{v} \in V$, $(-1)\boldsymbol{v} = -\boldsymbol{v}$ (the additive inverse of $\boldsymbol{v}$).

Proof.

(a) Suppose that $\mathbf{0}$ and $\mathbf{0}'$ are both additive identities. Then

$$\begin{aligned} \mathbf{0} &= \mathbf{0} + \mathbf{0}' && (\mathbf{0}' \text{ is additive identity}) \\ &= \mathbf{0}' + \mathbf{0} && ((\text{commutativity of } +) \\ &= \mathbf{0}' && (\mathbf{0}' \text{ is additive identity}) \end{aligned}$$

(b) Suppose that $\boldsymbol{w}$ and $\boldsymbol{w}'$ are both additive inverses for the same vector $\boldsymbol{v}$. Then

$$\begin{aligned} \boldsymbol{w} &= \boldsymbol{w} + \mathbf{0} && (\mathbf{0} \text{ is additive identity}) \\ &= \mathbf{0} + \boldsymbol{w} && (\text{commutativity of } +) \\ &= (\boldsymbol{v} + \boldsymbol{w}') + \boldsymbol{w} && (\boldsymbol{w}' \text{ is additive inverse}) \\ &= (\boldsymbol{w}' + \boldsymbol{v}) + \boldsymbol{w} && (\text{commutativity of } +) \\ &= \boldsymbol{w}' + (\boldsymbol{v} + \boldsymbol{w}) && (\text{associativity of } +) \\ &= \boldsymbol{w}' + \mathbf{0} && (\boldsymbol{w} \text{ is additive inverse}) \\ &= \boldsymbol{w}'. \end{aligned}$$

(c) Let $\boldsymbol{v} \in V$ be arbitrary and let $\boldsymbol{w}$ denote the additive identity of $0\boldsymbol{v}$. Then

$$\begin{aligned} \mathbf{0} &= \boldsymbol{w} + 0\boldsymbol{v} \\ &= \boldsymbol{w} + (0+0)\boldsymbol{v} \\ &= (\boldsymbol{w} + 0\boldsymbol{v}) + 0\boldsymbol{v} \\ &= \mathbf{0} + 0\boldsymbol{v} \\ &= 0\boldsymbol{v}. \end{aligned}$$

(d) Let $\lambda \in \mathbb{K}$ be arbitrary. Then

$$\lambda\mathbf{0} = \lambda(\mathbf{0} + \mathbf{0}) = \lambda\mathbf{0} + \lambda\mathbf{0}.$$

Letting $\boldsymbol{w}$ be the additive inverse of $\boldsymbol{\lambda}\mathbf{0}$, we have

$$\begin{aligned} \mathbf{0} &= \lambda\mathbf{0} + \boldsymbol{w} \\ &= \lambda(\mathbf{0} + \mathbf{0}) + \boldsymbol{w} \\ &= \lambda\mathbf{0} + \lambda\mathbf{0} + \boldsymbol{w} \\ &= \lambda\mathbf{0} + \mathbf{0} = \lambda\mathbf{0}. \end{aligned}$$

(e) Let $\boldsymbol{v} \in V$ be arbitrary. Then

$$\begin{aligned} \boldsymbol{v} + (-1)\boldsymbol{v} &= 1\boldsymbol{v} + (-1)\boldsymbol{v} \\ &= [1 + (-1)]\boldsymbol{v} \\ &= 0\boldsymbol{v} = \mathbf{0} \end{aligned}$$

whence $(-1)\boldsymbol{v} = -\boldsymbol{v}$.

□.

1C Subspaces

Definition: vector subspace

A subset U of V is called a **(vector) subspace** if is a vector space with the same vector addition, scalar multiplication, and additive identity as V.

Example 1C.1

The set of vectors

$$\big\{(x,y,0) \in \mathbb{R}^3 \;:\; x,y \in \mathbb{R}\big\} \qquad\qquad \text{(the } xy\text{-plane)}$$

is subspace of $\mathbb{R}^3$ (considered as an $\mathbb{R}$-vector space).

Proof. Let U be the subset of $\mathbb{R}^3$ described above. Let $\boldsymbol{u} = (x_1, y_1, 0)$, $\boldsymbol{v} = (x_2, y_2, 0)$, $\boldsymbol{w} = (x_3, y_3, 0)$ be arbitrary elements of U, and let $k, \ell \in \mathbb{R}$ be arbitrary. Then

1. [closure of addition]

   $$\boldsymbol{u} + \boldsymbol{v} = (x_1, y_1, 0) + (x_2, y_2, 0) = (x_1 + x_2, y_1 + y_2, 0).$$

   Since $x_1 + x_2 \in \mathbb{R}$ and $y_1 + y_2 \in \mathbb{R}$, then $\boldsymbol{u} + \boldsymbol{v} \in U$.
2. [commutativity of addition] NOTES INCOMPLETE
3. [associativity of addition] NOTES INCOMPLETE
4. [existence of zero vector] $\mathbf{0} = (0,0,0)$ is the additive identity in $\mathbb{R}^3$. Since 0 is a real number, then $(0,0,0) \in U$ also. Moreover,

   $$\boldsymbol{u} + \mathbf{0} = (x_1, y_1, 0) + (0,0,0) = (x_1 + 0, y_1 + 0, 0) = (x_1, y_1, 0) = \boldsymbol{u}.$$
5. [existence of additive inverses] NOTES INCOMPLETE
6. [closure of scalar multipliction] NOTES INCOMPLETE
7. [associativity of scalar multiplication] NOTES INCOMPLETE
8. [distributivity]

   $$\begin{aligned} k(\boldsymbol{u} + \boldsymbol{v}) &= k\Big((x_1, y_1, 0) + (x_2, y_2, 0)\Big) \\ &= k(x_1 + x_2, y_1 + y_2, 0) \\ &= \Big(k(x_1 + x_2), k(y_1 + y_2), 0\Big) \\ &= (kx_1 + kx_2, ky_1 + ky_2, 0) && \text{(distributivity of real numbers)} \\ &= (kx_1, ky_1, 0) + (kx_2, ky_2, 0) \\ &= k(x_1, y_1, 0) + k(x_2, y_2, 0) = k\boldsymbol{u} + k\boldsymbol{v}. \end{aligned}$$
9. [distributivity] NOTES INCOMPLETE
10. [multiplicative identity] NOTES INCOMPLETE

□

Theorem 1C.2: 3-Step Subspace Test

A subset $U \subseteq V$ is a subspace of V if and only if the following conditions hold:

- $\mathbf{0} \in U$. [contains the additive identity]
- For all $\boldsymbol{u_1}, \boldsymbol{u_2} \in U$, $\boldsymbol{u_1} + \boldsymbol{u_2} \in U$. [closed under $+$]
- For all $\boldsymbol{u} \in U$ and $k \in \mathbb{K}$, $k\boldsymbol{u} \in U$. [closed scalar multiplication]

Proof. Let V be a vector space and $U \subseteq V$.

($\Rightarrow$) Suppose that U is a vector space. Then it clearly satisfies the above criteria.

($\Leftarrow$) Suppose that U satisfies the above criteria. By assumption, U immediately satisfies properties (1), (6), and (4) [respectively] of the definition of a vector space. The existence of the additive inverse follows from the assumption that $k\boldsymbol{u} \in U$ for all k – in particular, take $k = -1$. The remaining properties all follow almost immediately from the fact that $\boldsymbol{u} \in U \implies \boldsymbol{u} \in V$, and all vectors in V satisfy these properties. To be super-duper explicit with the vector space axioms:

1. [closure of addition] This is by assumption.
2. [commutativity of addition] Let $\boldsymbol{u}, \boldsymbol{v} \in U$. Since U is a subset of V, then $\boldsymbol{u}, \boldsymbol{v} \in V$. Since V is a vector space, then $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$.
3. [associativity of addition] Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in U$. Since U is a subset of V, then $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$. Since V is a $\mathbb{K}$-vector space, then $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$.
4. [existence of zero vector] This is by assumption.
5. [existence of additive inverses] Since U is closed under scalar multiplication, for each vector $\boldsymbol{u} \in U$, the vector $-1\boldsymbol{u} \in U$. From a previous proposition, we know that $-1\boldsymbol{u}$ is the additive identity of $\boldsymbol{u}$.
6. [closure of scalar multipliction] This is by assumption.
7. [associativity of scalar multiplication] Let $k, \ell \in \mathbb{K}$ and $\boldsymbol{u} \in U$. Since U is a subset of B, then $\boldsymbol{u} \in V$. Since V is a $\mathbb{K}$-vector space, then $(k\ell)\boldsymbol{u} = k(\ell\boldsymbol{u})$.
8. [distributivity] Let $k \in \mathbb{K}$ and let $\boldsymbol{u}, \boldsymbol{v} \in U$. Since U is a subset of V, then $\boldsymbol{u}, \boldsymbol{v} \in V$. Since V is a $\mathbb{K}$-vector space, then $k(\boldsymbol{u} + \boldsymbol{v}) = k\boldsymbol{u} + k\boldsymbol{v}$.
9. [distributivity] Let $k, \ell \in \mathbb{K}$ and let $\boldsymbol{u} \in U$. Since U is a subset of V, then $\boldsymbol{u} \in V$. Since V is a $\mathbb{K}$-vector space, then $(k + \ell)\boldsymbol{u} = k\boldsymbol{u} + \ell\boldsymbol{u}$.
10. [multiplicative identity] NOTES INCOMPLETE

$\square$.

Example 1C.3

For any $m \in \mathbb{N}$, $\mathcal{P}_m(\mathbb{K})$ is a subspace of $\mathcal{P}(\mathbb{K})$.

We know that $\mathcal{P}_m(\mathbb{K})$ is a $\mathbb{K}$-vector space. Indeed, the zero vector in $\mathcal{P}_m(\mathbb{K})$ is the same as for $\mathcal{P}(\mathbb{K})$ (namely, the zero polynomail $p(x) = 0$). As well, the definitions of addition and scalar multiplication agree in both $\mathcal{P}_m(\mathbb{K})$ and $\mathcal{P}(\mathbb{K})$.

Example 1C.4

Is the set of vectors

$$U = \left\{(x, y, 0) \ : \ x, y \in \mathbb{Q}(\sqrt{3})\right\}$$

a subspace of $\mathbb{R}^3$ (which is an $\mathbb{R}$-vector space)?

U is not a subspace since it is not closed under scalar multiplication. In particular, $(1,0,0) \in U$, but

$$\pi(1,0,0) = (\pi,0,0)$$

and $\pi \notin \mathbb{Q}(\sqrt{3})$.

Definition: sums of subspaces

Let $U_1, \ldots, U_m$ be subspaces of V. The **sum** of $U_1, \ldots, U_m$, denoted $U_1 + \cdots + U_m$ is the set of all possible sums of elements from $U_1, \ldots, U_m$. More precisely,

$$U_1 + \cdots + U_m := \{u_1 + \cdots + u_m \ : \ u_j \in U_j \text{ for each } j = 1, \ldots, m\}.$$

One may also write $\displaystyle\sum_{i=1}^{m} U_i$.

Example 1C.5

Let U_1, U_2 be the following subspaces of $\mathbb{K}^4$.

$$U_1 = \{(x,x,y,y) \ : \ \text{for all } x,y \in \mathbb{K}\}$$
$$U_2 = \{(x,x,x,y) \ : \ \text{for all } x,y \in \mathbb{K}\}$$

Show that

$$U_1 + U_2 = \{(x,x,y,z) \ : \ \text{for all } x,y,z \in \mathbb{K}\}$$

Recall that showing two sets X and Y are equal means that we need to show that $X \subseteq Y$ and $X \supseteq X$.

Proof. To see that these sets are equal, we employ the usual tactic of proving that each set is a subset of the other. For notational simplicity, we let W be the set $\{(x,x,y,z) \ : \ \text{for all } x,y,z \in \mathbb{K}\}$.

$\boldsymbol{U_1 + U_2 \subseteq W}$. Let $(a,a,b,b) + (c,c,c,d) \in U_1 + U_2$. The sum of these vectors is

$$(a,a,b,b) + (c,c,c,d) = (a+c, a+c, b+c, b+d),$$

and since the first two components are equal, then $(a+c, a+c, b+c, b+d) \in W$.

$\boldsymbol{U_1 + U_2 \supseteq W}$. Let $(a,a,b,c) \in W$. Clearly one can write

$$(a,a,b,c) = (a,a,b,b) + (0,0,0,c-b),$$

and it is easily seen that $(a,a,b,b) \in U_1$ and $(0,0,0,c-b) \in U_2$.

□

Proposition 1C.6

Given subspaces $U_1, \ldots, U_m$ of V, the sum $U_1 + \cdots + U_m$ is the smallest (by set containment) subspace of V containing each U_i.

Proof.

Subspace. It is straightforward to verify that $U_1 + \cdots + U_m$ is a subspace using the 3-Step Subspace Test. That is contains each U_i is also immediate: for any $\boldsymbol{u_i} \in U_i$,

$$\mathbf{0} + \cdots + \boldsymbol{u_i} + \cdots + \mathbf{0} \in U_1 + \cdots + U_i + \cdots + U_m.$$

Smallest Subspace. Suppose now that W is a subspace satisfying

$$U_1 \cup \cdots \cup U_m \subseteq W \subseteq U_1 + \cdots + U_m.$$

Since W contains each subspace U_1 and W is a subspace itself, it must also contain all sums of elements from each U_i, hence it must also contain $U_1 + \cdots + U_m$. Thus

$$W \supseteq U_1 + \cdots + U_m$$

and therefore $W = U_1 + \cdots + U_m$.

□.

Definition: direct sum of subspaces

Let $U_1, \ldots, U_m$ be subspaces of a $\mathbb{K}$-vector space V. The sum $U_1 + \cdots + U_m$ is called a **direct sum** if and only if, for every $\boldsymbol{v} \in U_1 + \cdots + U_m$, there are unique $\boldsymbol{u_j}$'s so that

$$\boldsymbol{v} = \boldsymbol{u_1} + \cdots + \boldsymbol{u_m}$$

We denote the direct sum with

$$U_1 \oplus \cdots \oplus U_m.$$

One may also write $\displaystyle\bigoplus_{i=1}^{m} U_i$.

Proposition 1C.7

Let $U_1, \ldots, U_m$ be subspaces of V and let $\boldsymbol{u_j} \in U_j$, for each $j = 1, \ldots, m$. The following are equivalent:

1. $U_1 + \cdots + U_m$ is a direct sum.
2. If $\mathbf{0} = \boldsymbol{u_1} + \cdots + \boldsymbol{u_m}$, then $\boldsymbol{u_1} = \cdots = \boldsymbol{u_m} = \mathbf{0}$
3. For every i, the subspaces U_i and $\displaystyle\sum_{j \neq i}^{m} U_j$ have only the zero vector in common.

Proof.

(1 ⇒ 2). This is immediate from the definition of a direct sum.

(2 ⇒ 3). Let $\boldsymbol{u_i}$ be a vector common to both U_i and $\sum_{j\neq i}^{m} U_j$. Then, for $j \neq i$, there are vectors $\boldsymbol{u_j} \in U_j$ so that

$$\boldsymbol{u_i} = \boldsymbol{u_1} + \cdots + \boldsymbol{u_{i-1}} + \boldsymbol{u_{i+1}} + \cdots + \boldsymbol{u_m}.$$

This rearranges to

$$\mathbf{0} = \boldsymbol{u_1} + \cdots + \boldsymbol{u_{i-1}} + (-\boldsymbol{u_i}) + \boldsymbol{u_{i+1}} + \cdots + \boldsymbol{u_m}$$

and by assumption of Property (2), it must be that each $\boldsymbol{u_j} = \mathbf{0}$. In particular, $\boldsymbol{u_i} = \mathbf{0}$, so U_i and $\sum_{j\neq i}^{m} U_j$ have only the zero vector in common.

(3 ⇒ 1). Let $\boldsymbol{v} \in U_1 + \cdots + U_n$ and suppose for each i there are $\boldsymbol{u_i}, \boldsymbol{w_i} \in U_i$ so that

$$\boldsymbol{u_1} + \cdots + \boldsymbol{u_m} = \boldsymbol{v} = \boldsymbol{w_1} + \cdots + \boldsymbol{w_m}.$$

Subtracting and rearranging slightly, for each i we can write

$$\boldsymbol{u_i} - \boldsymbol{w_i} = \sum_{j\neq i} (\boldsymbol{w_j} - \boldsymbol{u_j})$$

and thus $(\boldsymbol{u_i} - \boldsymbol{w_i}) \in \sum_{j\neq i}^{m} U_j$. Since $(\boldsymbol{u_i} - \boldsymbol{w_i}) \in U_i$ as well, then by assumption of Property (3), we must have that $\boldsymbol{u_i} - \boldsymbol{w_i} = \mathbf{0}$, hence $\boldsymbol{u_i} = \boldsymbol{w_i}$ for every i. It follows that the decomposition of $\boldsymbol{v}$ is unique.

□.

Corollary 1C.8

Let U_1, U_2 be subspaces of V. Then $U_1 + U_2$ is a direct sum if and only if $U_1 \cap U_2 = \{\mathbf{0}\}$.

Example 1C.9: The Difference Between a Sum and Direct Sum

Let U_1, U_2, U_3 be the following subspaces of $\mathbb{R}^3$:

$$\begin{aligned} U_1 &= \big\{(x, y, 0) \in \mathbb{R}^3 \ : \ x, y \in \mathbb{R}\big\} \\ U_2 &= \big\{(0, y, z) \in \mathbb{R}^3 \ : \ x, y \in \mathbb{R}\big\} \\ U_3 &= \big\{(0, 0, z) \in \mathbb{R}^3 \ : \ x, y \in \mathbb{R}\big\} \end{aligned}$$

(a) Show that $\mathbb{R}^3 = U_1 + U_2$, but that this is not a direct sum.
(b) Show that $\mathbb{R}^3 = U_1 \oplus U_3$.

(a) $U_1 + U_2$ is a subspace of $\mathbb{R}^3$, so one only has to show that $U_1 + U_2 \supseteq \mathbb{R}^3$ to obtain equality. Indeed, note that $\mathbb{R}^3 \ni (x, y, z) = (x, y, 0) + (0, 0, z) \in U_1 + U_2$. To see that it is not a direct sum, one could use the corollary (any nonzero vector $(0, y, 0)$ is common to U_1 and U_2), but it can also be seen using uniqueness of sums. For any $(x, y, z) \in \mathbb{R}^3$ and any $t \in \mathbb{R}$, see that $(x, y, z) = (x, ty, 0) + (0, (1-t)y, z) \in U_1 + U_2$ so there is not a unique choice of vectors in U_1 and U_2 which sum to (x, y, z).

(b) Once again $U_1 + U_2$ is a subspace of $\mathbb{R}^3$, so one only has to show that $U_1 + U_3 \supseteq \mathbb{R}^3$ to obtain equality. indeed, note that $\mathbb{R}^3 \ni (x, y, z) = (x, y, 0) + (0, 0, z) \in U_1 + U_3$. To see that this is a direct sum, observe that the intersection of U_1 and U_3 is only the zero vector.

The purpose of a sum is to describe a decomposition of a vector space in terms of its subspaces. A direct sum is such a decomposition where the subspaces are independent of one another. That said, such subspaces are not typically unique.

Example 1C.10

Let U be the subspace of $\mathbb{K}^3$ given by

$$U = \left\{(k_1, k_2, k_1 + k_2) \in \mathbb{K}^3 \ : \ k_1, k_2 \in \mathbb{K}\right\}.$$

Let (a, b, c) be *any* vector for which $c \neq a + b$ and define the subspace

$$W = \left\{(at, bt, ct) \in \mathbb{K}^3 \ : \ t \in \mathbb{K}\right\}.$$

Prove that $\mathbb{K}^3 = U \oplus W$.

Proof. Suppose

$$(0, 0, 0) = (k_1, k_2, k_1 + k_2) + (at, bt, ct) \in U + W.$$

This equality yields the following systems

$$\begin{cases} k_1 + at &= 0 \\ k_2 + bt &= 0 \\ k_1 + k_2 + ct &= 0 \end{cases} \qquad \Longrightarrow \qquad \begin{cases} k_1 &= -at \\ k_2 &= -bt \\ -at - bt + ct &= 0 \end{cases}$$

Since $c \neq a + b$, then this last equation is 0 if and only if $t = 0$, whence $(at, bt, ct) = (0, 0, 0)$. It follows that $(k_1, k_2, k_3) = (0, 0, 0)$ and therefore there is a unique decomposition of $\mathbf{0}$. ☐

Section 1C Exercises

1. Let U_1, U_2 be arbitrary subspaces of an arbitrary $\mathbb{K}$-vector space, V. Prove or disprove each of the following.
    (a) $U_1 \cap U_2$ is a subspace of V.
    (b) $U_1 \cup U_2$ is a subspace of V.
2. Given two vectors $\mathbf{u} = (x_1, \ldots, x_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$ in $\mathbb{R}^n$, recall that the *dot product* is defined as

$$\mathbf{u} \cdot \mathbf{v} = x_1 y_1 + \cdots + x_n y_n.$$

Let U and N be the following subspaces of $\mathbb{R}^3$:

$$U = \{(x, y, x+y) \ : \ x, y \in \mathbb{R}\}$$
$$N = \left\{\mathbf{v} \in \mathbb{R}^3 \ : \ \mathbf{u} \cdot \mathbf{v} = 0 \text{ for every } \mathbf{u} \in U\right\}$$

    (a) Find the missing vector entry that makes the following statement true (no proof needed):

$$N = \left\{(x, x, \underline{\quad}) \in \mathbb{R}^3 \ : \ x \in \mathbb{R}\right\}.$$

    (b) Prove that $\mathbb{R}^3 = U \oplus N$.

3. Let $\mathcal{P}_m^{\text{odd}}(\mathbb{K})$ and $\mathcal{P}_m^{\text{even}}(\mathbb{K})$ be the following subsets of $\mathcal{P}_m(\mathbb{K})$:

$$\mathcal{P}_m^{\text{odd}}(\mathbb{K}) = \{a_0 + a_1 x + \cdots + a_m x^m \ : a_j = 0 \text{ whenever } j \text{ is not odd}\}$$
$$\mathcal{P}_m^{\text{even}}(\mathbb{K}) = \{a_0 + a_1 x + \cdots + a_m x^m \ : a_j = 0 \text{ whenever } j \text{ is not even}\}$$

In other words $\mathcal{P}_m^{\text{odd}}(\mathbb{K})$ is the collection of polynomials whose only nonzero terms are odd-degree, and $\mathcal{P}_m^{\text{even}}(\mathbb{K})$ is the collection of polynomials whose only nonzero terms are even-dgree.

    (a) Prove that $\mathcal{P}_m^{\text{odd}}(\mathbb{K})$ and $\mathcal{P}_m^{\text{even}}(\mathbb{K})$ are subspaces of $\mathcal{P}_m(\mathbb{K})$.
    (b) Prove that $\mathcal{P}_m(\mathbb{K}) = \mathcal{P}_m^{\text{odd}}(\mathbb{K}) \oplus \mathcal{P}_m^{\text{even}}(\mathbb{K})$.
    (c) Let V, W be $\mathbb{K}$-vector spaces with addition operations $\underset{V}{+}$ and $+_W$. The Cartesian product $V \times W$ is a vector space (called the *product space*) with addition given by

$$(\mathbf{v_1}, \mathbf{w_1}) + (\mathbf{v_2}, \mathbf{w_2}) = (\mathbf{v_1} +_V \mathbf{v_2}, \mathbf{w_1} +_W \mathbf{w_2})$$

and scalar multiplication given by

$$\lambda(\mathbf{v_1}, \mathbf{w_1}) = (\lambda \mathbf{v_1}, \lambda \mathbf{w_1})$$

for all $\lambda \in \mathbb{K}$ and all $\mathbf{v_1}, \mathbf{v_2} \in V$ and all $\mathbf{w_1}, \mathbf{w_2} \in W$. Suppose that X is a subspace of V and Y is a subspace of W. Prove or disprove the following: $X \times Y$ a subspace of $V \times W$.

Chapter 2

Finite-Dimensional Vector Spaces

2A Span and Linear Independence

Definition: linear combination

A **linear combination** of a finite set $S = \{\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}\}$ of vectors in V is a vector of the form

$$a_1\boldsymbol{v_1} + \cdots + a_m\boldsymbol{v}_m$$

where $a_1, \ldots, a_m \in \mathbb{K}$.

Definition: span

The **span** of a finite set $S = \{\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}\}$ of vectors in V, denoted either $\operatorname{Span}(S)$ or $\operatorname{Span}(\boldsymbol{v_1}, \ldots, \boldsymbol{v_m})$ is the set of all linear combinations of these vectors. Explicitly,

$$\operatorname{Span}(S) = \operatorname{Span}(\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}) = \{a_1\boldsymbol{v_1} + \cdots + a_m\boldsymbol{v_m} \ : \ a_1, \ldots, a_m \in \mathbb{K}\}$$

When $S = \emptyset$, we define $\operatorname{Span}(\emptyset) = \{\mathbf{0}\}$.
Given a subset $W \subseteq V$, if we can find a set S of vectors in V for which $\operatorname{Span}(S) = X$, we say that S **spans** W, or that S is a **spanning set** for W.

Example 2A.1

Let $V = \mathbb{R}^3$ (considered as an $\mathbb{R}$-vector space). Describe $\operatorname{Span}(\boldsymbol{v_1}, \boldsymbol{v_2})$ where $\boldsymbol{v_1} = (0, 1, 0)$ and $\boldsymbol{v_2} = (0, 0, 1)$.

We have that

$$\begin{aligned}\operatorname{Span}(\boldsymbol{v_1}, \boldsymbol{v_2}) &= \{x(1, 0, 0) + y(0, 1, 0) \ : \ x, y \in \mathbb{R}\} \\ &= \{(x, y, 0) \ : \ x, y \in \mathbb{R}\}\end{aligned}$$

which you might affectionately refer to as the xy-plane.

Example 2A.2

Let $V = \mathcal{P}(\mathbb{K})$ (considered as a $\mathbb{K}$-vector space). Describe $\operatorname{Span}(p_0, p_1, p_2)$ where $p_0(x) = 1$, $p_1(x) = x$, and $p_2(x) = 2 - x$.

$$\begin{aligned}\operatorname{Span}(p_0,p_1,p_2) &= \{k_1(1)+k_2(x)+k_3(2-x)\ :\ k_1,k_2,k_3\in\mathbb{K}\}\\ &= \{(k_1+2k_3)+(k_2-k_3)x\ :\ k_1,k_2,k_3\in\mathbb{K}\}\end{aligned}$$

and making the substitution $a_0 = k_1+2k_2$, $a_1 = k_2-k_3$

$$= \{(a_0+a_1x\ :\ a_0,a_1\in\mathbb{K}\}$$

which we might refer to as $\mathcal{P}_1(\mathbb{K})$.

Proposition 2A.3

Given any finite set S of vectors of V, $\operatorname{Span}(S)$ is the smallest (by set containment) subspace of V containing S.

The proof of this claim is very similar to the proof that the sum of subspaces is the smallest subspace containing all of the component subspaces.

Proof. Let $S=\{\boldsymbol{v_1},\dots,\boldsymbol{v_n}\}$.

Subspace. It is straightforward to see that $\operatorname{Span}(S)$ is a subspace by way of the 3-Step Subspace Test. Indeed, $\mathbf{0}\in\operatorname{Span}(S)$ as

$$\mathbf{0} = 0\boldsymbol{v_1}+\cdots+0\boldsymbol{v_n}.$$

To see that it is closed under addition and scalar multiplication, let $k_1,\dots,k_n,\ell_1,\dots,\ell_n,\lambda$ be scalars. Then

$$\begin{aligned}&\lambda\left(k_1\boldsymbol{v_1}+\cdots+k_n\boldsymbol{v_n}\right)+\left(\ell_1\boldsymbol{v_1}+\cdots+\ell_n\boldsymbol{v_n}\right)\\ &=(\lambda k_1+\ell_1)\boldsymbol{v_1}+\cdots+(\lambda k_n+\ell_n)\boldsymbol{v_n}\end{aligned}$$

Smallest Subspace. Suppose W is any subspace containing S. Then by closure of addition/scalar multiplication, W contains all of the linear combinations of vectors in S, hence $\operatorname{Span}(S)\subseteq W$.

$\square$.

Definition: finite-/infinite-dimensional vector space

A vector space V is called **finite-dimensional** if there is a finite set S of vectors in V for which $V=\operatorname{Span}(S)$. V is called **infinite-dimensional** otherwise.

Example 2A.4

Show that $\mathbb{R}^n$ is finite-dimensional.

Proof. For each $i=1,\dots,n$, let $\boldsymbol{e_i}=(e_1,e_2,\dots,e_n)$ where $e_j=\begin{cases}1 & \text{when } j=i\\ 0 & \text{otherwise}\end{cases}$. We then have that

$$\boldsymbol{v} = (v_1, \dots, v_n) = v_1\boldsymbol{e_1} + \cdots + v_n\boldsymbol{e_n}$$

hence

$$\mathbb{R}^n = \operatorname{Span}(\boldsymbol{e_1}, \dots, \boldsymbol{e_n})\,.$$

□

Example 2A.5

Show that $\mathcal{P}(\mathbb{K})$ is infinite-dimensional.

Proof. Let $S = \{p_1(x), \dots, p_n(x)\}$ be any finite set of polynomials. Since polynomials have finite degree by definition, then we may write

$$m := \max\{\deg(p_1), \dots, \deg(p_n)\}$$

and $m < \infty$. But now the polynomial $q(x) = x^{m+1} \notin \operatorname{Span}(S)$, so $\mathcal{P}(\mathbb{K})$ cannot be finite-dimensional. □

Definition: linear dependence/independence

Let $S = \{\boldsymbol{v_1}, \dots, \boldsymbol{v_m}\}$ be a finite set of vectors in V.
S is called **linearly independent** if and only if, for all $k_1, \dots, k_m \in \mathbb{K}$ for which

$$k_1\boldsymbol{v_1} + \cdots + k_m\boldsymbol{v_m} = \boldsymbol{0} \qquad \Longrightarrow \qquad k_1 = \cdots = k_m = 0.$$

S is called **linearly dependent** if it is not linearly independent. That is, there are coefficients $k_1, \dots, k_m \in \mathbb{K}$, not all zero, for which

$$k_1\boldsymbol{v_1} + \cdots + k_m\boldsymbol{v_m} = \boldsymbol{0}$$

.

Example 2A.6

Show that $S = \{(1,0,0),\ (0,1,0)\}$ is a linearly independent set of vectors in $\mathbb{R}^3$.

Let $x, y \in \mathbb{R}$ and suppose

$$\boldsymbol{0} = (0,0,0) = x(1,0,0) + y(0,1,0) = (x,y,0).$$

Clearly, $x = y = 0$, hence S is a linearly independent set of vectors.

Example 2A.7

Show that $S = \{1 + x, 1 + x^2\}$ is a linearly independent set of vectors in $\mathcal{P}_2(\mathbb{K})$. Recall that the “zero vector” is the zero polynomial, $p(x) = 0$.

Let $k_1, k_2 \in \mathbb{K}$ and suppose

$$0 = k_1(1+x) + k_2(1+x^2).$$

This rearranges to

$$0 = (k_1 + k_2) + k_1 x + k_2 x^2.$$

Since two polynomials are equal if and only if their corresponding coefficients are equal, then we must have

$$k_1 + k_2 = 0, \qquad k_1 = 0, \qquad k_2 = 0.$$

Thus S is a linearly independent set.

Example 2A.8

Is $S = \{1+x, 1+x+x^2, 1+x+3x^2\}$ a linearly independent set of vectors in $\mathcal{P}_2(\mathbb{K})$?

As before, let $k_1, k_2, k_3 \in \mathbb{K}$ and suppose

$$0 = k_1(1+x) + k_2(1+x+x^2) + k_3(1+x+3x^2)$$

This rearranges to

$$0 = (k_1 + k_2 + k_3) + (k_1 + k_2 + k_3)x + (k_2 + 3k_3)x^2$$

which results in the system

$$\begin{cases} k_1 + k_2 + k_3 & = 0 \\ k_1 + k_2 + k_3 & = 0 \\ k_2 + 3k_3 & = 0. \end{cases}$$

Choose $k_1 = -(k_2 + k_3)$ and $k_2 = -3k_3$. Then for any value of k_3, we will have a valid solution to this system. In particular, pick $k_3 = 1$ (from which it follows that $k_2 = -3$ and $k_3 = -(-3+1) = -(-2) = 2$. We see that at least one of (in fact, all three of) k_1, k_2, k_3 are nonzero, but

$$0 = 2(1+x) - 3(1+x+x^2) + (1+x+3x^2).$$

Therefore S is not linearly independent (i.e. it is a linearly dependent set).

Lemma 2A.9: Linear Dependence Lemma

Let $S = \{\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}\}$ be a finite set of vectors in V. The following are equivalent.

1. S is linearly dependent.
2. There is some j satisfying $1 \leq j \leq m$ for which

$$\boldsymbol{v_j} \in \operatorname{Span}(\boldsymbol{v_1}, \ldots, \boldsymbol{v_{j-1}}, \boldsymbol{v_{j+1}}, \ldots, \boldsymbol{v_m}).$$

3. There is some j satisfying $1 \leq j \leq m$ for which

$$\operatorname{Span}(\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}) = \operatorname{Span}(\boldsymbol{v_1}, \ldots, \boldsymbol{v_{j-1}}, \boldsymbol{v_{j+1}}, \ldots, \boldsymbol{v_m}).$$

Proof.

(1 ⇔ 2) By definition, S is linearly dependent if and only if there is at least one nonzero k_j (with $1 \leq j \leq m$) satisfying

$$k_1\boldsymbol{v_1} + \cdots + k_m\boldsymbol{v_m} = \mathbf{0}.$$

(WLOG, take $j = 1$.) Then the equation above holds if and only if

$$\boldsymbol{v_1} = -\frac{k_2}{k_1}\boldsymbol{v_2} - \cdots - \frac{k_m}{k_1}\boldsymbol{v_m},$$

(Aside: Since $\mathbb{K}$ is a field, we can freely divide by nonzero field elements, and $k_1 \neq 0$).

Now $\boldsymbol{v_1}$ is a linear combination of $\boldsymbol{v_2}, \ldots, \boldsymbol{v_m}$ (as above) if and only if $\boldsymbol{v_1} \in \operatorname{Span}(\boldsymbol{v_2}, \ldots, \boldsymbol{v_m})$.

(2 ⇔ 3) (WLOG, take $j = 1$.) $\boldsymbol{v_1} \in \operatorname{Span}(\boldsymbol{v_2}, \ldots, \boldsymbol{v_m})$ if and only if there are scalars $k_2, \ldots, k_m \in \mathbb{K}$ for which

$$\boldsymbol{v_1} = k_2\boldsymbol{v_2} + \cdots + k_m\boldsymbol{v_m}.$$

Now observe that

$$\begin{aligned}
&\operatorname{Span}(\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}) \\
&= \{c_1\boldsymbol{v_1} + \cdots + c_m\boldsymbol{v_m} \ : \ c_1, \ldots, c_m \in \mathbb{K}\} \\
&= \{c_1(k_2\boldsymbol{v_2} + \cdots + k_m\boldsymbol{v_m}) + c_2\boldsymbol{v_2} + \cdots + c_m\boldsymbol{v_m} \ : \ c_1, \ldots, c_m \in \mathbb{K}\} \\
&= \{(c_1k_2 + c_2)\boldsymbol{v_2} + \cdots + (c_1k_m + c_m)\boldsymbol{v_m}) \ : \ c_1, \ldots, c_m \in \mathbb{K}\} \qquad (2.1) \\
&= \operatorname{Span}(\boldsymbol{v_2}, \ldots, \boldsymbol{v_m})
\end{aligned}$$

where the equality in Equation (2.1) holds because each of the c_j's (and hence $c_1k_j + c_j$) are arbitrary.

□.

Corollary 2A.10

If S is any nonempty set of vectors and $\mathbf{0} \in S$, then S is linearly dependent.

Theorem 2A.11

Given two sets of vectors in V,

$$S_u = \{\boldsymbol{u_1}, \ldots, \boldsymbol{u_m}\} \text{ and}$$
$$S_w = \{\boldsymbol{w_1}, \ldots, \boldsymbol{w_n}\},$$

if S_u is linearly independent and $V = \operatorname{Span}(S_v)$, then $m \leq n$.

The strategy is that we're going to iteratively create a new spanning set by removing $\boldsymbol{w_i}$ vectors from S_w and replacing them with $\boldsymbol{u_j}$ vectors.

Proof. Let S_u and S_w be as defined above and define

$$B_0 := S_w$$

Since $V = \operatorname{Span}(S_w) = \operatorname{Span}(B_0)$, then every vector in S_u (and in particular, $\boldsymbol{u_1}$) must be a linear combination of vectors in B_0. That is, there are some scalars $k_1, \cdots k_n \in \mathbb{K}$, not all 0 (because Corollary 2A.10 implies that $\boldsymbol{u_1}$ cannot be zero) for which

$$\boldsymbol{u_1} = k_1\boldsymbol{w_1} + \cdots + k_n\boldsymbol{w_n}. \tag{2.2}$$

It follows from the Linear Dependence Lemma that

$$\operatorname{Span}(B_0) = \operatorname{Span}(\boldsymbol{w_1}, \ldots, \boldsymbol{w_n}) = \operatorname{Span}(\boldsymbol{u_1}, \boldsymbol{w_1}, \ldots, \boldsymbol{w_n}). \tag{2.3}$$

Up to relabeling the vectors in S_w, we may assume that $k_1 \neq 0$ in Equation (2.2), and thus we can rewrite

$$\boldsymbol{w_1} = \frac{1}{k_1}\boldsymbol{u_1} + \frac{k_2}{k_1}\boldsymbol{w_2} + \cdots + \frac{k_n}{k_1}\boldsymbol{w_n}.$$

We now have that $\boldsymbol{w} \in \operatorname{Span}(\boldsymbol{u_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_n})$, and by the Linear Dependence Lemma,

$$\operatorname{Span}(B_0) = \operatorname{Span}(\boldsymbol{u_1}, \boldsymbol{w_1}, \ldots, \boldsymbol{w_n}) = \operatorname{Span}(\boldsymbol{u_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_n}).$$

As such, we define

$$B_1 := \{\boldsymbol{u_1}\} \cup (B_0 - \{\boldsymbol{w_1}\})$$

and more generally, with appropriate relabeling in every step, for $j \geq 1$,

$$B_j := \{\boldsymbol{u_j}\} \cup (B_{j-1} - \{\boldsymbol{w_j}\}).$$

For each $j \geq 1$, we have that

$$\boldsymbol{u_j} \in \operatorname{Span}(\boldsymbol{u_1}, \ldots, \boldsymbol{u_{j-1}}, \boldsymbol{w_j}, \ldots, \boldsymbol{w_n}),$$

and by linear independence of the $\boldsymbol{u_i}$'s, it must be that

$$\boldsymbol{u_j} = 0\boldsymbol{u_1} + \cdots + 0\boldsymbol{u_{j-1}} + \ell_j\boldsymbol{w_j} + \cdots + \ell_n\boldsymbol{w_n}$$

where (after relabeling), $\ell_j \neq 0$. The same argument as above above gives us

$$\operatorname{Span}(B_0) = \cdots = \operatorname{Span}(B_j) = \operatorname{Span}(\boldsymbol{u_1}, \ldots, \boldsymbol{u_j}, \boldsymbol{w_{j+1}}, \ldots, \boldsymbol{w_n}),$$

so the only thing left to do is to ensure that this procedure doesn't terminate before m steps.

Seeking a contradiction, assume that it does terminate before m steps, i.e., that $n < m$. Then we have that

$$V = \operatorname{Span}(S_w) = \operatorname{Span}(B_0) = \cdots = \operatorname{Span}(B_n) = \operatorname{Span}(\boldsymbol{u_1}, \ldots, \boldsymbol{u_n})$$

and by linear independence of the $\boldsymbol{u_i}$'s, we have the following containments:

$$V = \operatorname{Span}(B_n) = \operatorname{Span}(\boldsymbol{u_1}, \ldots, \boldsymbol{u_n}) \subsetneq \operatorname{Span}(\boldsymbol{u_1}, \ldots, \boldsymbol{u_n}, \boldsymbol{u_{n+1}}, \ldots, \boldsymbol{u_m}) \subseteq V.$$

which is absurd. Therefore, $m \leq n$. $\square$.

Proposition 2A.12

If V is a finite-dimensional vector space and U is any subspace of V, then U is finite-dimensional.

Proof. If $U = \{\mathbf{0}\}$ then we're done. Otherwise, there is some nonzero vector $\boldsymbol{u_1} \in U$. Define the following sets:

$$\begin{aligned} S_1 &= \{\boldsymbol{u_1}\}\,, \text{ and for each } j \geq 2 \\ S_j &= S_{j-1} \cup \{\boldsymbol{u_k}\} \text{ where } \boldsymbol{u_j} \notin \operatorname{Span}(S_{j-1}) \end{aligned}$$

By construction, S_j is linearly independent for each j, and by Theorem 2A.11, $j \leq \dim(V) < \infty$, hence there some $j_0 \in \mathbb{N}$ for which $\operatorname{Span}(S_{j_0}) = U$, whence $\dim(U) = j_0$. ☐.

Corollary 2A.13

If V contains an infinite-dimensional subspace, then V is infinite-dimensional.

Section 2A Exercises

1. Let V be a $\mathbb{K}$-vector space and suppose there are subspaces U_i for which $V = U_1 \oplus \cdots \oplus U_n$. For each i, choose any nonzero vector $\boldsymbol{u_i} \in U_i$. Prove that $S = \{\boldsymbol{u_1}, \ldots, \boldsymbol{u_n}\}$ is a linearly independent set.

2. **Relationship between disjoint unions and direct sums.** Let S_1 and S_2 be (nonempty) pairwise disjoint sets of vectors and suppose $S_1 \cup S_2$ is linearly independent. Prove that

$$\operatorname{Span}(S_1 \cup S_2) = \operatorname{Span}(S_1) \oplus \operatorname{Span}(S_2).$$

3. Show that the vector space of continuous real-valued functions is infinite-dimensional.
   : Hint: Corollary 2A.13

4. Show that the set of functions $\{\cos(nx) \ : n \in \mathbb{N}\}$ is a linearly independent set.
   : Hint 1: Recall that linear combinations can only involve a finite number of vectors
   Hint 2: Recall the Maclaurin expansion $\cos(t) = 1 - \frac{1}{2}t^2 + \frac{1}{4}t^4 + \cdots + \frac{1}{(2k)!}t^{2k} + O(t^{2k+2})$

2B Bases

Definition: basis

A **basis** of a vector space V is a linearly independent set B of vectors in V which span V.

Example 2B.1

Show that $\{(1,0),\,(0,1)\}$ is a basis for $\mathbb{R}^2$.

Proof. **Span.** We need to show that $\operatorname{Span}\Big((1,0),\,(0,1)\Big) = \mathbb{R}^2$. Certainly we have that the span is a subset of $\mathbb{R}^2$, so we only need to show the reverse containment. Let $(x,y) \in \mathbb{R}^2$ be arbitrary. Then we see that

$$(x,y) = x(1,0) + y(0,1)$$

and since $x, y \in \mathbb{R}$, this is a linear combination of the vectors $(1,0)$ and $(0,1)$, hence $(x,y) \in \operatorname{Span}\Big((1,0),\,(0,1)\Big)$.

Linear Independence. Suppose that there are real numbers x, y for which

$$x(1,0) + y(0,1) = (0,0)$$

The left-hand side of this equation is equal to (x,y), and $(x,y) = (0,0)$ implies that $x = 0, y = 0$. Hence $\{(1,0),\,(0,1)\}$ is a linearly independent set.

□

Example 2B.2

Show that $\{1, x, x^2\}$ is a basis for $\mathcal{P}_2(\mathbb{K})$.

Proof. **Span.** We need to show that $\operatorname{Span}\Big(1, x, x^2\Big) = \mathcal{P}_2(\mathbb{K})$. Certainly we have that the span is a subset of $\mathcal{P}_2(\mathbb{K})$, so we only need to show the reverse containment. Let $k_0 + k_1 x + k_2 x^2 \in \mathcal{P}_2(\mathbb{K})$ be arbitrary. Then we see that

$$k_0 + k_1 x + k_2 x^2 = k_0(1) + k_1(x) + k_2(x^2)$$

which is a linear combination of the vectors 1, x, and x^2, $k_0 + k_1 x + k_2 x^2 \in \operatorname{Span}\Big((1,0),\,(0,1)\Big)$.

Linear Independence. Suppose that there are scalars k_0, k_1, k_2 for which

$$k_0 + k_1 x + k_2 x^2 = 0$$

Since two polynomials are equal if and only if their corresponding coefficients are equal, then it must be that $k_0 = k_1 = k_2 = 0$, hence this set is linearl independent.

□

Theorem 2B.3

A set $\mathcal{B} = \{\boldsymbol{b_1}, \dots, \boldsymbol{b_n}\}$ is a basis for V if and only if, for all $\boldsymbol{v} \in V$, there are unique scalars $k_1, \dots, k_n \in \mathbb{K}$ for which

$$\boldsymbol{v} = k_1\boldsymbol{b_1} + \cdots + k_n\boldsymbol{b_n}.$$

Proof. **(⇒)** Suppose that $\mathcal{B}$ is a basis for V. Then, by definition, it must span V, i.e., every vector $\boldsymbol{v} \in V$ is a linear combination of the $\boldsymbol{b_i}$'s. To see that this linear combination is unique, suppose there are scalars $k_1, \dots, k_n, \ell_1, \dots, \ell_n$ for which

$$\begin{aligned} \boldsymbol{v} &= k_1\boldsymbol{b_1} + \cdots + k_n\boldsymbol{b_n}, \text{ and} \\ \boldsymbol{v} &= \ell_1\boldsymbol{b_1} + \cdots + \ell_n\boldsymbol{b_n}. \end{aligned}$$

Subtracting, one gets

$$\mathbf{0} = (k_1 - \ell_1)\boldsymbol{b_1} + \cdots + (k_n - \ell_n)\boldsymbol{b_n}$$

and since $\mathcal{B}$ is linearly independent, $k_i - \ell_i = 0$ for each i, whence $k_i = \ell_i$.

(⇐) Suppose instead that, for every $\boldsymbol{v} \in V$, there is a unique choice of scalars $k_1, \dots, k_n$ so that

$$\boldsymbol{v} = k_1\boldsymbol{b_1} + \cdots + k_n\boldsymbol{b_n}.$$

Since every $\boldsymbol{v}$ can be written as a linear combination of vectors in $\mathcal{B}$, it must be that $V \subseteq \operatorname{Span}(\mathcal{B}) \subseteq V$ (where the second containment follows immediately from the fact that $\mathcal{B} \subset V$) and thus $\mathcal{B}$ is a spanning set. To see linear independence, take $\boldsymbol{v} = \mathbf{0}$ in the equation above

$$\mathbf{0} = k_1\boldsymbol{b_1} + \cdots + k_n\boldsymbol{b_n}.$$

We know that $k_1 = \cdots = k_n = 0$ gives one valid linear combination of the zero vector, and by assumption on uniqueness, this must be the only such linear combination. Hence $\mathcal{B}$ is a linearly independent set.

□

Theorem 2B.4: reducing/extending a basis

Let V be a finite-dimensional $\mathbb{K}$-vector space and let $S = \{\boldsymbol{v_1}, \dots, \boldsymbol{v_m}\}$ be a set of nonzero vectors in V.

Reduce a set to a basis. If S spans V, then there is a subset $S' \subseteq S$ which is a basis for V.

Extend a set to a basis. If S is linearly independent, then there is a superset $S' \supseteq S$ which is a basis for V.

1. Here's the strategy:
    - Take the spanning set S, and check for linear dependence.
    - If it is linearly independent, we're done. Otherwise, Lemma 2A.9 says that we can throw away some $\boldsymbol{v_i}$ without affecting the span.
    - Check for linear dependence of $S - \{\boldsymbol{v_i}\}$.

- If it is linearly independent, we're done. Otherwise, Lemma 2A.9 says that we can throw away some $\boldsymbol{v_j}$ without affecting the span.
- Check for linear dependence of $S - \{\boldsymbol{v_i}, \boldsymbol{v_j}\}$.
- ...Repeat this process until you have linear independence.

Proof. If S is linearly independent, then it is a basis. If S is not linearly independent, then there is some vector ($\boldsymbol{v_1}$, say) which is a linear combination of the remaining vectors, and by the Linear Dependence Lemma,

$$\operatorname{Span}(\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}) = \operatorname{Span}(\boldsymbol{v_2}, \ldots, \boldsymbol{v_m}).$$

Now we check linear independence of $\{\boldsymbol{v_2}, \ldots, \boldsymbol{v_m}\}$ and repeatedly apply Linear Dependence Lemma as necessary. As S contains only finitely-many vectors, this process must terminate, and the remaining set is linearly independent, hence a basis. □

2. Here's the strategy:
   - Find a spanning set for V, call it T, and look at the set $S \cup T$.
   - $S \cup T$ spans, but might not be linearly independent.
   - Apply the strategy from Part 1, making sure to only throw away vectors from T.

Proof. Suppose V is spanned by some set of vectors $\{\boldsymbol{w_1}, \ldots, \boldsymbol{w_n}\}$, and consider the set

$$\left\{ \underbrace{\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}}_{S}, \boldsymbol{w_1}, \ldots, \boldsymbol{w_n} \right\}$$

which contains S. Since the $\boldsymbol{w_i}$'s span V and the $\boldsymbol{v_i}$'s are all contained in V, this set spans V. However, it may not be linearly independent.

If it is not linearly independent, then any "dependencies" must be coming from the $\boldsymbol{w_i}$ vectors (because the $\boldsymbol{v_i}$'s are linearly independent[1]). We can repeatedly apply Linear Dependence Lemma as necessary to remove the appropriate $\boldsymbol{w_i}$ vectors until what remains is a linearly independent set and which still spans. That is, if we assume that we removed $n - k$ vectors, then up to relabeling, we would have

$$\left\{ \underbrace{\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}}_{S}, \boldsymbol{w_{m+1}}, \ldots, \boldsymbol{w_k} \right\}$$

This set is linearly independent, spans V, and contains S. Calling this set S' completes the proof. □

---

[1] To expand upon this, if there are some nonzero coefficients satisfying

$$\mathbf{0} = a_1\boldsymbol{v_1} + \cdots + a_m\boldsymbol{v_m} + b_1\boldsymbol{w_1} + \cdots + b_n\boldsymbol{w_n},$$

then it cannot be the case that all of the b_j coefficients are zero, otherwise this would result in a nontrivial linear combination of the $\boldsymbol{v_i}$ vectors summing to $\mathbf{0}$ and therefore violating the linear independence assumption.

Corollary 2B.5

For every subspace U of V, there is another subspace W for which $V = U \oplus W$.

Proof. Let S_U be a basis for U. Since S_U is linearly independent, applying the results of Theorem 2B.4, S_U can be extended to S, a basis for V. Taking $S_W = S - S_U$, one can take $W = \operatorname{Span}(S_W)$. It is straightforward to see that $U \cap W = \{\mathbf{0}\}$. □

Section 2B Exercises

1. Let $V = \mathbb{C}$ be a $\mathbb{C}$-vector space. Prove that any nonzero complex number z is a basis for V.
2. Let $V = \mathbb{C}$ be a $\mathbb{R}$-vector space. Find a basis for V.
3. Let $V = \mathcal{P}_3(\mathbb{R})$ be an $\mathbb{R}$-vector space and let U be the following subspace

$$U = \{p(x) \in \mathcal{P}_3(\mathbb{R}) \ : \ p'(7) = 0\}$$

where $p'(7)$ is the derivative of $p(x)$ evaluated at $x = 7$. Find a basis for U.

2C Dimension

Theorem 2C.1

Any two bases of a finite-dimensional vector space have the same cardinality.

Proof. Let B_1, B_2 be two basis for V. Since B_1 is linearly independent and $\operatorname{Span}(B_1) = \operatorname{Span}(B_2)$, then Theorem 2A.11 implies $|B_1| \leq |B_2|$. Similarly, since B_2 is linearly independent, then Theorem 2A.11 implies that $|B_2| \leq |B_1|$. Hence $|B_1| = |B_2|$. □

Because the cardinality of the basis is independent of the choice of basis, the following is well-defined:

Definition: dimension

The **dimension** of a finite-dimensional vector space, denoted $\dim V$, is the cardinality of any basis for V.

Example 2C.2

Determine the dimension of $\mathbb{K}^n$.

The set

$$\{(1,0,0,\ldots,0),\, (0,1,0,\ldots,0), \ldots, (0,0,\ldots,0,1)\}$$

is a basis for $\mathbb{K}^n$ and the cardinality of this set is n.

Example 2C.3

Determine the dimension of $\mathcal{P}_n(\mathbb{K})$.

The set

$$\{1, x, \ldots, x^n\}$$

is a basis for $\mathcal{P}_n(\mathbb{K})$ and the cardinality of this set is $n+1$.

Example 2C.4

Determine the dimension of the following subspace U of $\mathbb{K}^3$:

$$U = \{(x,y,z) \in \mathbb{K}^3 \ : \ x + y + z = 0\}.$$

Check that $U = \operatorname{Span}((1,0,1),(0,1,-1))$ and that these two vectors are linearly independent.

Theorem 2C.5

Let U be a subspace of V. Then $\dim U \leq \dim V$, with equality precisely when $U = V$.

Proof. If V is finite-dimensional, then so is U (see Proposition 2A.12). Since U is finite-dimensional, there is a finite set S which is a basis for U. Using Theorem 2B.4, this can be extended to a set $S' \supseteq S$ which is a basis for V. Hence we have

$$\dim(U) = \text{cardinality of } S \leq \text{cardinality of } S' = \dim(V)$$

and equality holds when S and S' have the same cardinality, i.e., are equal. In which case $U = \operatorname{Span}(S) = \operatorname{Span}(S') = V$. □

Theorem 2C.6

If U_1, U_2 are subspaces of a finite-dimensional vector space, then

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

The strategy of the proof is this: Find a basis $\{\boldsymbol{v_1}, \dots, \boldsymbol{v_n}\}$ for $V_1 \cap V_2$, extend it to a basis for V_1, and then extend it again to a basis for V_2.

$$\underbrace{\boldsymbol{u_1}, \dots, \boldsymbol{u_k}, \rlap{$\overbrace{\phantom{\boldsymbol{v_1}, \dots, \boldsymbol{v_m}, \boldsymbol{w_1}, \dots, \boldsymbol{w_n}}}^{\in V_2}$}\boldsymbol{v_1}, \dots, \boldsymbol{v_m}}_{\in V_1}, \boldsymbol{w_1}, \dots, \boldsymbol{w_n}$$

Proof. Using all of the notation above, let

$$S_u = \{\boldsymbol{u_1}, \dots, \boldsymbol{u_k}\}, \quad S_v = \{\boldsymbol{v_1}, \dots, \boldsymbol{v_m}\}, \quad \text{and } S_w = \{\boldsymbol{w_1}, \dots, \boldsymbol{w_n}\}.$$

First we show that $V_1 + V_2 = \operatorname{Span}(S_u \cup S_v \cup S_w)$.

First observe that

$$\begin{aligned} & V_1 = \operatorname{Span}(S_u \cup S_v) \subseteq \operatorname{Span}(S_u \cup S_v \cup S_w) \text{ and} \\ & V_2 = \operatorname{Span}(S_v \cup S_w) \subseteq \operatorname{Span}(S_u \cup S_v \cup S_w) \\ \Longrightarrow\ & V_1 \cup V_2 \subseteq \operatorname{Span}(S_u \cup S_v \cup S_w). \end{aligned}$$

Since $V_1 + V_2$ is the smallest subspace containing $V_1 \cup V_2$, then

$$V_1 + V_2 \subseteq \operatorname{Span}(S_u \cup S_v \cup S_w).$$

Now let $\boldsymbol{x} \in \operatorname{Span}(S_u \cup S_v \cup S_w)$. Then there exist scalars $\alpha_i, \beta_j, \gamma_\ell$ so that

$$\boldsymbol{x} = \sum_i \alpha_i \boldsymbol{u_i} + \sum_j \beta_j \boldsymbol{v_i} + \sum_\ell \gamma_\ell \boldsymbol{w_\ell} = \underbrace{\sum_i \alpha_i \boldsymbol{u_i} + \sum_j 0 \boldsymbol{v_i}}_{\in V_1} + \underbrace{\sum_j \beta_j \boldsymbol{v_i} + \sum_\ell \gamma_\ell \boldsymbol{w_\ell}}_{\in V_2}$$

hence $\boldsymbol{x} \in V_1 + V_2$. Thus $V_1 + V_2 = \operatorname{Span}(S_u \cup S_v \cup S_w)$.

Now we show that $S_u \cup S_v \cup S_w$ *is linearly independent.*

By construction in extending bases, $S_u \cup S_v$ and $S_v \cup S_w$ are linearly independent. Suppose there are scalars $\alpha_i, \beta_j, \gamma_\ell$ so that

$$\mathbf{0} = \underbrace{\sum_i \alpha_i \boldsymbol{u_i}}_{\in V_1} + \underbrace{\sum_j \beta_j \boldsymbol{v_i}}_{\in V_1 \cap V_2} + \underbrace{\sum_\ell \gamma_\ell \boldsymbol{w_\ell}}_{\in V_2}\,. \tag{2.4}$$

By rearranging this equation, we have

$$\sum_i \alpha_i \boldsymbol{u_i} = -\sum_j \beta_j \boldsymbol{v_i} - \sum_\ell \gamma_\ell \boldsymbol{w_\ell}$$

and thus $\sum_i \alpha_i \boldsymbol{u_i} \in \operatorname{Span}(S_v \cup S_W) = V_2$. It follows then that $\sum_i \alpha_i \boldsymbol{u_i} \in V_1 \cap V_2$, so it is some linear combination of the $\boldsymbol{v_i}$'s. We can thus find new scalars δ_j to rewrite Equation 2.4 as

$$\begin{aligned}\mathbf{0} &= \underbrace{\sum_i \alpha_i \boldsymbol{u_i}}_{\in V_1} + \underbrace{\sum_j \beta_j \boldsymbol{v_j}}_{\in V_1 \cap V_2} + \underbrace{\sum_\ell \gamma_\ell \boldsymbol{w_\ell}}_{\in V_2} \\ &= \underbrace{\sum_j \delta_j \boldsymbol{v_j}}_{\in V_1 \cap V_2} + \underbrace{\sum_\ell \gamma_\ell \boldsymbol{w_\ell}}_{\in V_2}\,.\end{aligned}$$

By linear independence of $S_v \cup S_w$, we must have $\gamma_\ell = 0$ for each ℓ. By a similar argument and linear independence of $S_u \cup S_v$, we must have $\alpha_i = 0$ for each i.

Equation 2.4 has now been reduced to

$$\mathbf{0} = \underbrace{\sum_j \beta_j \boldsymbol{v_i}}_{\in V_1 \cap V_2}$$

and by our assumption of linear independence, $\beta_j = 0$ for each j. Therefore $S_u \cup S_v \cup S_w$ is linearly independent. □

Corollary 2C.7

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

and by induction it follows that

Corollary 2C.8

$$\dim\left(\bigoplus_{i=1}^{m} U_i\right) = \sum_{i=1}^{m} \dim(U_i)$$

Chapter 3

Linear Maps

3A Vector Space of Linear Maps

Definition: linear map

Let V, W be $\mathbb{K}$-vector spaces. A **linear map** is a function $T : V \to W$ with the following properties:

(a) **additivity:** for all $v_1, v_2 \in V$, $T(v_1 + v_2) = T(v_1) + T(v_2)$.

(b) **homogeneity:** for all $\lambda \in \mathbb{K}$ and all $v \in V$, $T(\lambda v) = \lambda T(v)$.

Definition: Notation $\mathcal{L}(V, W)$ and $\mathcal{L}(V)$

The set of linear maps $T : V \to W$ is denoted $\mathcal{L}(V, W)$. When $V = W$, we simply write $\mathcal{L}(V)$ or $\operatorname{End}(V)$.

Remark. The notation $\operatorname{End}(V)$ refers to the *endomorphism algebra*. That $\operatorname{End}(V)$ is an algebra (i.e. a vector space with an appropriate multiplicative operation) will be more apparent later in this section.

3A.I Examples of Linear Maps

Example 3A.1: zero map

Show that the following map is a linear transformation.

$$\begin{gathered} T : V \to W \\ T(\boldsymbol{v}) = \mathbf{0} \end{gathered}$$

This is straightforward.

Example 3A.2: identity map

Show that the following map is a linear transformation.

$$\begin{gathered} T : V \to V \\ T(\boldsymbol{v}) = \boldsymbol{v} \end{gathered}$$

This is straightforward.

Example 3A.3: differentiation

Show that the following map is a linear transformation.

$$T : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$$
$$T(p(x)) = \frac{d}{dx}[p(x)]$$

This is straightforward, and follows from results from just about any first-year calculus course. Nevertheless, it might be nice to see it explicitly in terms of polynomials. For simplicity, we'll use summation notation.

Let $p(x) = \sum a_i x^i$ and $q(x) = \sum b_j x^j$ be polynomials and λ a scalar. Then

Additivity.

$$\begin{aligned} T(p(x)+q(x)) &= \frac{d}{dx}\left(\sum_{k\geq 0}(a_k+b_k)x^k\right) \\ &= \sum_{k\geq 1} k(a_k+b_k)x^{k-1} \\ &= \left(\sum_{k\geq 1} ka_kx^{k-1}\right) + \left(\sum_{k\geq 1} kb_kx^{k-1}\right) \\ &= \frac{d}{dx}\left(\sum_{k\geq 0} a_kx^k\right) + \frac{d}{dx}\left(\sum_{k\geq 0} b_kx^k\right) \\ &= T(p(x)) + T(q(x)) \end{aligned}$$

Homogeneity.

$$\begin{aligned} T(\lambda p(x)) &= \frac{d}{dx}\left(\sum_{k\geq 0}\lambda a_kx^k\right) \\ &= \sum_{k\geq 1} k\lambda a_kx^{k-1} \\ &= \lambda\left(\sum_{k\geq 1} ka_kx^{k-1}\right) \\ &= \lambda\frac{d}{dx}\left(\sum_{k\geq 0} a_kx^x\right) \\ &= \lambda T(p(x)) \end{aligned}$$

Example 3A.4: integration map

Show that the following map is a linear transformation.

$$T : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$$

$$T(p(x)) = \int_0^1 p(x)\,dx$$

This is straightforward, and follows from results from just about any first-year calculus course. Without appealing to outside results, the proof explicitly looks like that of the previous example.
We remark that

$$\int_0^1 p(x)\,dx = \int_0^1 \left(\sum_{k\geq 0} a_k x^k\right)\,dx = \int_0^1 a_0+\cdots+a_n x^n\,dx = a_0+\frac{a_1}{2}+\cdots+\frac{a^n}{n+1} = \sum_{k\geq 0}\frac{a_k}{k+1}.$$

so intuitively the output is just a finite sum of real numbers, so both conditions ought to be satisfied by the distributive property of real numbers.

Example 3A.5: multiplication by x^3

Show that the following map is a linear transformation.

$$\begin{aligned} T &: \mathcal{P}(\mathbb{K}) \to \mathcal{P}(\mathbb{K}) \\ T(p(x)) &= x^3 p(x)\,dx \end{aligned}$$

This is straightforward.

Lemma 3A.6: Linear Map Lemma

Let V, W be n-dimensional $\mathbb{K}$-vector spaces, with bases

$$\{\boldsymbol{v_1},\ldots,\boldsymbol{v_n}\} \qquad \text{and} \qquad \{\boldsymbol{w_1},\ldots,\boldsymbol{w_n}\},$$

respectively. Then there exists a unique linear map

$$\begin{aligned} T &: V \to W \\ T(\boldsymbol{v_j}) &= \boldsymbol{w_j} \qquad \text{for each } j = 1,\ldots,n. \end{aligned}$$

Proof strategy:

- Construction of such a T: just define it in the natural way.
- Uniqueness: Suppose there is another function S with the same properties, and see that $T(\boldsymbol{v}) = S(\boldsymbol{v})$ for every $\boldsymbol{v} \in V$.

Proof. Let $k_1,\ldots,k_n \in \mathbb{K}$.

Existence. Define the function

$$\begin{aligned} T &: V \to W \\ T(k_1\boldsymbol{v_1}+\cdots+k_n\boldsymbol{v_n}) &= k_1\boldsymbol{w_1}+\cdots+k_n\boldsymbol{w_n}. \end{aligned}$$

For each fixed i, taking $k_i = 1$ and $k_{j\neq i} = 0$, one achieves $T(\boldsymbol{v_i}) = \boldsymbol{w_i}$. To see that it is linear, we check both additivity and homogeneity at the same time. Let $\lambda \in \mathbb{K}$ and let $\boldsymbol{x}, \boldsymbol{y} \in V$, writing them as

$$\boldsymbol{x} = \sum_i \alpha_i \boldsymbol{v_i} \qquad \text{and} \qquad \boldsymbol{y} = \sum_i \beta_i \boldsymbol{v_i}.$$

Then

$$\begin{aligned} T(\lambda \boldsymbol{x} + \boldsymbol{y}) &= T\left(\lambda \sum_i \alpha_i \boldsymbol{v_i} + \sum_i \beta_i \boldsymbol{v_i}\right) \\ &= T\left(\sum_i (\lambda\alpha_i + \beta_i)\boldsymbol{v_i}\right) \\ &= \sum_i (\lambda\alpha_i + \beta_i)\boldsymbol{w_i} \\ &= \lambda\left(\sum_i \alpha_i \boldsymbol{w_i}\right) + \left(\sum_i \beta_i \boldsymbol{w_i}\right) \\ &= \lambda T(\boldsymbol{x}) + T(\boldsymbol{y}). \end{aligned}$$

Therefore T is linear.

Uniqueness. Suppose $S : V \to W$ is some other linear map satisfying $S(\boldsymbol{v_i}) = \boldsymbol{w_i}$ for each i. Then, for every vector $\boldsymbol{x} \in V$, it follows from linearity of S that

$$\begin{aligned} T(\boldsymbol{x}) &= c_1 T(\boldsymbol{v_1}) + \cdots + c_n T(\boldsymbol{v_n}) \\ &= c_1 \boldsymbol{w_1} + \cdots + c_n \boldsymbol{w_n} \\ &= c_1 S(\boldsymbol{v_1}) + \cdots + c_n S(\boldsymbol{v_n}) \\ &= S(\boldsymbol{x}) \end{aligned}$$

hence $T = S$.

□

Remark. The moral of this result is twofold: (1) a linear transformation is *uniquely defined* by where it sends the basis, and (2) given any two bases, there is a unique linear map which allows you to convert between bases.

3A.II Algebraic Operations on $\mathcal{L}(V, W)$

Theorem 3A.7: vector space of linear maps

Given two $\mathbb{K}$-vector spaces, V and W, $\mathcal{L}(V, W)$ is a vector space with the following addition and scalar multiplication operations:

addition. For all $T_1, T_2 \in \mathcal{L}(V, W)$ and for all $\boldsymbol{v} \in V$,

$$(T_1 + T_2)(\boldsymbol{v}) = T_1(\boldsymbol{v}) + T_2(\boldsymbol{v}).$$

scalar multiplication. For all $T \in \mathcal{L}(V, W)$, for all $\boldsymbol{v} \in V$, and for all $\lambda \in \mathbb{K}$,

$$(\lambda T)(\boldsymbol{v}) = \lambda\big(T(\boldsymbol{v})\big).$$

Note that most of the properties below require one to also check homogeneity and additivity. This proof is an exercise in overcoming tedium.

Proof. Let $T_1, T_2, T_3 \in \mathcal{L}(V, W)$ and $\lambda, \mu \in \mathbb{K}$. We verify each of the axioms of a vector space.

1. [closure of addition] Let $\boldsymbol{v_1}, \boldsymbol{v_2} \in V$ and $\lambda \in \mathbb{K}$ be arbitrary. Then
$$\begin{aligned}(T_1 + T_2)(\lambda\boldsymbol{v_1} + \boldsymbol{v_2}) &= T_1(\lambda\boldsymbol{v_1} + \boldsymbol{v_2}) + T_2(\lambda\boldsymbol{v_1} + \boldsymbol{v_2})\\ &= \lambda T_1(\boldsymbol{v_1}) + T_1(\boldsymbol{v_2}) + \lambda T_2(\boldsymbol{v_1}) + T_2(\boldsymbol{v_2})\\ &= \lambda\left[T_1(\boldsymbol{v_1}) + T_2(\boldsymbol{v_1})\right] + T_1(\boldsymbol{v_2}) + T_2(\boldsymbol{v_2})\\ &= \lambda\,(T_1 + T_2)(\boldsymbol{v_1}) + (T_1 + T_2)(\boldsymbol{v_2})\end{aligned}$$
hence $T_1 + T_2 \in \mathcal{L}(V, W)$.
2. [commutativity of addition] Left as an exercise for the reader
3. [associativity of addition] Left as an exercise for the reader
4. [existence of zero] Let Z be the zero map from Example 3A.1. Left as an exercise for the reader
5. [existence of additive inverses] Left as an exercise for the reader
6. [closure of scalar multiplication] Left as an exercise for the reader
7. [associativity of scalar multiplication] Left as an exercise for the reader
8. [distributivity] Left as an exercise for the reader
9. [distributivity] Left as an exercise for the reader
10. [existence of a multiplicative identity] Left as an exercise for the reader

□

Definition: product of linear maps

Given $\mathbb{K}$-vector spaces U, V, W, then for all $T \in \mathcal{L}(U, V)$ and for all $S \in \mathcal{L}(V, W)$, the **product** of S and T is the linear map $ST \in \mathcal{L}(U, W)$ given by

$$(ST)(\boldsymbol{u}) = S(T(\boldsymbol{u})).$$

In other words, this is just the usual composition of functions.

Proposition 3A.8: properties of products of linear maps

Let U, V, W, X be arbitrary $\mathbb{K}$-vector spaces.

1. [associativity] For all $T_3 \in \mathcal{L}(U, V)$, $T_2 \in \mathcal{L}(V, W)$, and $T_1 \in \mathcal{L}(W, X)$,
$$T_1(T_2T_3) = (T_1T_2)T_3.$$

2. [distributivity] For all $T_1, T_2 \in \mathcal{L}(U,V)$ and $S_1, S_2 \in \mathcal{L}(V,W)$,
$$(S_1 + S_2)T_1 = S_1T_1 + S_2T_1 \qquad \text{and} \qquad S_1(T_1 + T_2) = S_1T_1 + S_2T_2.$$
3. [identity]. Let $\mathrm{Id}_V \in \mathcal{L}(V)$ and $\mathrm{Id}_W \in \mathcal{W}$ be the identity maps on V and W (see Example 3A.2), respectively. For all $T \in \mathcal{L}(V,W)$,
$$\mathrm{Id}_W T = T\,\mathrm{Id}_V\,.$$

Proof. 1. [associativity] Let $T_3 \in \mathcal{L}(U,V)$, $T_2 \in \mathcal{L}(V,W)$, and $T_1 \in \mathcal{L}(W,X)$ be aribtrary. Also, let $\boldsymbol{u} \in U$ be arbitrary and define $\boldsymbol{v} = T_3(\boldsymbol{u})$, $\boldsymbol{w} = T_2(\boldsymbol{v})$, $\boldsymbol{x} = T_1(\boldsymbol{w})$. Then

$$\begin{aligned} &(T_2T_3)(\boldsymbol{u}) = T_2(T_3(\boldsymbol{u})) = T_2(\boldsymbol{v}) = \boldsymbol{w} \\ \implies &(T_1(T_2T_3))(\boldsymbol{u}) = T_1(\boldsymbol{w}) = \boldsymbol{x} \end{aligned}$$

and similarly

$$\begin{aligned} T_3(\boldsymbol{u}) &= \boldsymbol{v} \\ ((T_1T_2)T_3)(\boldsymbol{u}) &= (T_1T_2)(\boldsymbol{v}) = T_1(T_2(\boldsymbol{v})) = T_1(\boldsymbol{w}) = \boldsymbol{x}. \end{aligned}$$

2. [distributivity] Straightforward and left as an exercises for the reader.
3. [identity]. Straightforward and left as an exercise for the reader.

□

Example 3A.9

Let D be the differentiation map defined in Example 3A.3 and let T be the "product by x^3 map" in Example 3A.5. Explicitly,

$$\begin{aligned} D &: \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R}) & T &: \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R}) \\ D(p(x)) &= \frac{d}{dx}[p(x)] & T(p(x)) &= x^3p(x) \end{aligned}$$

Show that the linear map product is not a commutative operation by checking that

$$DT \neq TD.$$

We compute these explicitly. Let $p(x) \in \mathcal{P}(\mathbb{R})$ be arbitrary. Then

$$(DT)(p(x)) = D(x^3p(x)) = 3x^2p(x) + x^3p'(x)$$

and

$$(TD)(p(x)) = T(p'(x)) = x^3p'(x).$$

These functions are not equal for all polynomials $p(x)$, hence the functions DT and TD are not equal.

Section 3A Exercises

1. Show that the Linear Map Lemma is false when $\dim(V) > \dim(W)$.
2. Show that, for any linear map T, $T(\mathbf{0}) = \mathbf{0}$.
3. Give an example of a function $f : \mathbb{R}^2 \to \mathbb{R}$ which is homogeneous, but is not linear.
4. Give an example of a function $f : \mathbb{C} \to \mathbb{C}$ which is additive, but is not linear.

3B Null Spaces and Ranges

3B.I Null Space and Injectivity

Definition: null space

The **null space** or **kernel** of a linear map $T \in \mathcal{L}(V, W)$ is the set

$$\operatorname{Null}(T) := \{\boldsymbol{v} \in V \ : \ T(\boldsymbol{v}) = \boldsymbol{0}\}.$$

Example 3B.1

Let $T : \mathbb{R}^2 \to \mathbb{R}$ be the linear map given by $T((x, y)) = x + y$. Find $\operatorname{Null}(T)$.

By definition

$$\operatorname{Null}(T) = \big\{(x, y) \in \mathbb{R}^2 \ : \ x + y = 0\big\} = \big\{(t, -t) \in \mathbb{R}^2 \ : \ t \in \mathbb{R}\big\}$$

which is a one-dimensional subspace of $\mathbb{R}^2$ (intuitively - the line $y = -x$).

Example 3B.2

Let $T : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ be the linear map given by $T\big(p(x)\big) = \dfrac{d}{dx}\big[p(x)\big]$. Find $\operatorname{Null}(T)$.

Recall that the zero polynomial, $z(x) = 0$, is the "zero vector." So you're looking for all polynomials $p(x)$ for which

$$p'(x) = z(x) = 0$$

These are the constant polynomials.

Proposition 3B.3: the null space is a subspace

For any $T \in \mathcal{L}(V, W)$, $\operatorname{Null}(T)$ is a subspace of V.

Proof. Let $T \in \mathcal{L}(V, W)$ be arbitrary. We apply the 3-Step Subspace Test.

Contains the zero vector. Since $T(\boldsymbol{0}) = \boldsymbol{0}$, then $\boldsymbol{0} \in \operatorname{Null}(T)$.

Closure under addition/scalar multiplication. Suppose now that $\boldsymbol{v_1}, \boldsymbol{v_2} \in \operatorname{Null}(T)$ and λ is a scalar. Then

$$\begin{aligned} T(\lambda \boldsymbol{v_1} + \boldsymbol{v_2}) &= \lambda T(\boldsymbol{v_1}) + T(\boldsymbol{v_2}) && \text{(by assumption of linearity)} \\ &= \lambda \boldsymbol{0} + \boldsymbol{0} && \text{(by assumption of } \boldsymbol{v_1}, \boldsymbol{v_2} \in \operatorname{Null}(T)) \\ &= \boldsymbol{0} \end{aligned}$$

hence $\lambda \boldsymbol{v_1} + \boldsymbol{v_2} \in \operatorname{Null}(T)$.

□

Definition: nullity

The dimension of $\mathrm{Null}(T)$ is called the **nullity** of T.

Definition: injective

A linear map T is called **injective** or **one-to-one** precisely when, for all $\boldsymbol{v}, \boldsymbol{w}$

$$T(\boldsymbol{v}) = T(\boldsymbol{w}) \qquad \text{implies} \qquad \boldsymbol{v} = \boldsymbol{w},$$

or, equivalently

$$\boldsymbol{v} \neq \boldsymbol{w} \qquad \text{implies} \qquad T(\boldsymbol{v}) \neq T(\boldsymbol{w}).$$

Remark. This definition holds for all functions, in fact. The former description is typically easier to work with, but the latter description gives better insight into the behavior of such functions. The spirit of this definition is that it means the range is a "copy of" the domain. For vector spaces, this is made explicit in Corollary 3B.14.

Example 3B.4

Let $T : \mathbb{K}^2 \to \mathbb{K}^4$ be the linear map given by

$$T(x, y) = (x + y, 0, x - y, 0)$$

Show that T is injective.

Let $\boldsymbol{v} = (x_1, y_1)$ and $\boldsymbol{w} = (x_2, y_2)$. Suppose $T(\boldsymbol{v}) = T(\boldsymbol{w})$. Then

$$(x_1 + y_1, 0, x_1 - y_1, 0) = (x_2 + y_2, 0, x_2 - y_2, 0).$$

This yields the following system

$$\begin{cases} x_1 + y_1 &= x_2 + y_2 \\ x_1 - y_1 &= x_2 - y_2 \end{cases} \tag{3.1}$$

Adding the equations in the system 3.1 gives

$$2x_1 = 2x_2 \implies x_1 = x_2$$

and subtracting the equations in the system 3.1 gives

$$-2y_1 = -2y_2 \implies y_1 = y_2$$

hence $\boldsymbol{v} = \boldsymbol{w}$. Therefore T is injective.

Example 3B.5

Show that the derivative map (c.f. Example 3A.3)

$$T : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$$

$$T(p(x)) = \frac{d}{dx}[p(x)]$$

is *not* injective.

Counter-example: any two polynomials that differ only in the constant term will have the same derivative.

Theorem 3B.6

A linear map T is injective if and only if $\text{Null}(T) = \{\mathbf{0}\}$.

Proof. Let $\boldsymbol{v}, \boldsymbol{w} \in V$ and let $T \in \mathcal{L}(V, W)$.

($\Rightarrow$). Suppose T is injective and $\boldsymbol{v} \in \text{Null}(T)$. Then

$$T(\boldsymbol{v}) = \mathbf{0} = T(\mathbf{0})$$

and by injectivity, this implies that $\boldsymbol{v} = \mathbf{0}$, hence $\text{Null}(T) = \{\mathbf{0}\}$.

($\Leftarrow$). Conversely, suppose $\text{Null}(T) = \{\mathbf{0}\}$ and $T(\boldsymbol{v}) = T(\boldsymbol{w})$. We then have the following chain of implications:

$$\begin{aligned}
& T(\boldsymbol{v}) = T(\boldsymbol{w}) \\
\implies & T(\boldsymbol{v}) - T(\boldsymbol{w}) = \mathbf{0} \\
\implies & T(\boldsymbol{v} - \boldsymbol{w}) = \mathbf{0} && \text{(by linearity of } T\text{)} \\
\implies & \boldsymbol{v} - \boldsymbol{w} \in \text{Null}(T) \\
\implies & \boldsymbol{v} - \boldsymbol{w} = \mathbf{0} && \text{(since } \text{Null}(T) = \{\mathbf{0}\}\text{)} \\
\implies & \boldsymbol{v} = \boldsymbol{w}
\end{aligned}$$

whence T is injective.

☐

3B.II Range and Surjectivity

Definition: range

The **range** or **image** of a linear map $T \in \mathcal{L}(V, W)$ is the set

$$\text{Range}(T) := \{T(\boldsymbol{v}) \ : \ \boldsymbol{v} \in V\}.$$

Example 3B.7

Let $T : \mathbb{R}^2 \to \mathbb{R}$ be the linear map given by $T((x, y)) = x + y$. Find $\text{Range}(T)$.

Let $r \in \mathbb{R}$ be arbitrary and notice that $T(r, 0) = r$. It follows that $\text{Range}(T) = \mathbb{R}$.

Example 3B.8

Find Range(T) where T is the derivative map (c.f. Example 3A.3):

$$T : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$$
$$T(p(x)) = \frac{d}{dx}[p(x)]$$

Let $p(x) = a_n x^n + \cdots + a_1 x + a_0$ be any polynomial in $\mathcal{P}(\mathbb{R})$. Observe that the polynomial $P(x) = \frac{a_n}{n+1}x^{n+1} + \cdots + \frac{a_1}{2}x^2 + a_0 x$ has the property that

$$T(P(x)) = p(x).$$

Thus $\text{Range}(T) = \mathcal{P}(\mathbb{R})$.

Proposition 3B.9: the range is a subspace

For any $T \in \mathcal{L}(V, W)$, Range(T) is a subspace of W.

Proof. We apply the 3-Step Subspace Test.

Contains the zero vector. Since $T(\mathbf{0}) = \mathbf{0}$ for every linear map, then $\mathbf{0} \in \text{Range}(T)$.

Closure under addition/scalar multiplication. Let $\boldsymbol{w_1}, \boldsymbol{w_2} \in \text{Range}(T)$. By definition of the range, there are vectors $\boldsymbol{v_1}, \boldsymbol{v_2}$ for which $T(\boldsymbol{v_1}) = \boldsymbol{w_1}$ and $T(\boldsymbol{v_2}) = \boldsymbol{w_2}$. Also, let λ be a scalar. Then

$$\lambda \boldsymbol{w_1} + \boldsymbol{w_2} = \lambda T(\boldsymbol{v_1}) + T(\boldsymbol{v_2} = T(\lambda \boldsymbol{v_1} + \boldsymbol{v_2})$$

and therefore $\lambda \boldsymbol{w_1} + \boldsymbol{w_2} \in \text{Range}(T)$.

□

Definition: rank

The dimension of Range(T) is called the **rank** of T.

Definition: surjective

A linear map $T \in \mathcal{L}(V, W)$ is called **surjective** or **onto** precisely when, for all $\boldsymbol{w} \in W$,

$$T(\boldsymbol{v}) = \boldsymbol{w} \qquad \text{for some } \boldsymbol{v} \in V,$$

or, equivalently

$$\text{Range}(T) = W.$$

Remark. This definition holds for all functions, in fact. The former description is typically easier to work with, but the latter description gives better insight into the behavior of such functions.

Example 3B.10

Let $T : \mathbb{R}^2 \to \mathbb{R}$ be the linear map given by $T(x, y) = x + y$. Determine whether or not T is surjective.

Example 3B.11

Let T be the derivative map (c.f. Example 3A.3):

$$\begin{aligned} T &: \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R}) \\ T(p(x)) &= \frac{d}{dx}[p(x)] \end{aligned}$$

Determine whether or not T is surjective

Our work in Example 3B.8 shows that T is surjective.

Exercise 3B.12

Let T be the derivative map (c.f. Example 3A.3) be restricted to finite-degree polynomials:

$$\begin{aligned} T &: \mathcal{P}_m(\mathbb{R}) \to \mathcal{P}_m(\mathbb{R}) \\ T(p(x)) &= \frac{d}{dx}[p(x)] \end{aligned}$$

Determine whether or not T is surjective.

Theorem 3B.13: Fundamental Theorem of Linear Maps
(aka Rank–Nullity Theorem)

Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Then $\operatorname{rank}(T) < \infty$ and

$$\begin{aligned} \dim(V) &= \dim(\operatorname{Range}(T)) + \dim(\operatorname{Null}(T)) \\ &= \operatorname{rank}(T) + \operatorname{nullity}(T). \end{aligned}$$

Proof strategy:

- Let $\{\boldsymbol{n_1}, \ldots, \boldsymbol{n_k}\}$ be a basis for $\operatorname{Null}(T)$.
- Extend to a basis $\{\boldsymbol{n_1}, \ldots, \boldsymbol{n_k}, \boldsymbol{r_1}, \ldots, \boldsymbol{r_m}\}$ for V.
- Show that $\{T(\boldsymbol{r_1}), \ldots, T(\boldsymbol{r_m})\}$ are a basis for $\operatorname{Range}(T)$.
- Then

$$\dim(V) = k + m = \dim(\operatorname{Range}(T)) + \dim(\operatorname{Null}(T)) = \operatorname{rank}(T) + \operatorname{nullity}(T).$$

Proof. Suppose V is finite-dimensional and let $T \in \mathcal{L}(V, W)$ be arbitrary. Since V is finite-dimensional and Null(T) is a subspace of V, Null(T) is also finite dimensional. Let

$$\{\boldsymbol{n_1}, \ldots, \boldsymbol{n_k}\}$$

be a basis for Null(T). We can extend this to a basis for V (see Theorem 2B.4), written as

$$\{\boldsymbol{n_1}, \ldots, \boldsymbol{n_k}, \boldsymbol{r_1}, \ldots, \boldsymbol{r_m}\}$$

We claim that $\{T(\boldsymbol{r_1}), \ldots, T(\boldsymbol{r_m})\}$ is a basis for Range(T).

Span. That this set spans Range(T), let $\boldsymbol{y} \in$ Range(T). Then there is some $\boldsymbol{x} \in V$ for which $T(\boldsymbol{x}) = \boldsymbol{y}$. Writing down $\boldsymbol{x}$ as a linear combination of V's basis vectors,

$$\boldsymbol{x} = \alpha_1 \boldsymbol{n_1} + \cdots + \alpha_k \boldsymbol{n_k} + \beta_1 \boldsymbol{r_1} + \cdots + \beta_m \boldsymbol{r_m}$$

from which we see that

$$\begin{aligned} \boldsymbol{y} &= \alpha_1 T(\boldsymbol{n_1}) + \cdots + \alpha_k T(\boldsymbol{n_k}) + \beta_1 T(\boldsymbol{r_1}) + \cdots + \beta_m T(\boldsymbol{r_m}) \\ &= \boldsymbol{0} + \cdots + \boldsymbol{0} + \beta_1 T(\boldsymbol{r_1}) + \cdots + \beta_m T(\boldsymbol{r_m}) \end{aligned}$$

and thus every $\boldsymbol{y} \in$ Range(T) is a linear combination of $T(\boldsymbol{r_i})$ vectors.

Linear Independence. Suppose that there are scalars β_j, $j = 1, \ldots, m$ for which

$$\begin{aligned} \beta_1 T(\boldsymbol{r_1}) + \cdots + \beta_m T(\boldsymbol{r_m}) &= 0 \\ \implies T(\beta_1 \boldsymbol{r_1} + \cdots + \beta_m \boldsymbol{r_m}) &= 0 \end{aligned}$$

INCOMPLETE. The above shows this linear combination is actually in the null space. Hence the $\boldsymbol{r_j}$'s are a linear combination of $\boldsymbol{n_i}$'s, which can only happen when all of the coefficients are 0.

Now we have that

$$\begin{aligned} \dim(V) = m + k &= \dim(\text{Range}(T)) + \dim(\text{Null}(T)) \\ &= \text{rank}(T) + \text{nullity}(T). \end{aligned}$$

□

Combining this with Theorem 3B.6, one gets

Corollary 3B.14

If V is finite dimensional and $T \in \mathcal{L}(V, W)$, then T is injective if and only if $\dim(V) = \dim(\text{Range}(T))$.

Section 3B Exercises

1. Let V, W be finite-dimensional $\mathbb{K}$-vector spaces and let $T \in \mathcal{L}(V, W)$. Prove each of the following claims.
   (a) If $\dim(V) > \dim(W)$, then T cannot be injective.
   (b) If $\dim(V) < \dim(W)$, then T cannot be surjective.
2. Show that the converse of each statement in the previous problem is not true.
3. **Relationships to linear systems.** Let $T \in \mathcal{L}(\mathbb{K}^2)$ be the linear map

$$T(x_1, x_2) = (ax_1 + bx_2, cx_1 + dx_2)$$

for some scalars a, b, c, d. Suppose that $(k_1, k_2) \in \text{Range}(T)$. Prove that T is injective if and only if the system below has a unique solution:

$$\begin{cases} ax_1 + bx_2 = k_1 \\ cx_1 + dx_2 = k_2 \end{cases} .$$

3C Matrices

Definition: matrix

Let m, n be nonnegative integers. An $m \times n$ **matrix** A is a rectangular array of elements in $\mathbb{K}$ with m rows and n columns.

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ \vdots & \vdots & & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{pmatrix}.$$

We may write $A = [A_{i,j}]$ for short.

The set of all $m \times n$ matrices with entries in $\mathbb{K}$ is sometimes denoted $\mathbb{K}^{m \times n}$.

Definition: matrix of a linear map

Let V, W be finite-dimensional $\mathbb{K}$-vector spaces and let $T \in \mathcal{L}(V, W)$. Let

$$\begin{aligned} \mathcal{B} &= \{\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}\} \\ \mathcal{C} &= \{\boldsymbol{w_1}, \ldots, \boldsymbol{w_m}\} \end{aligned}$$

be bases for V and W, respectively.The **matrix of $\boldsymbol{T}$**, denoted $\mathcal{M}(T)$, is the $m \times n$ matrix $A = [A_{i,j}]$ whose entries are defined by

$$T(\boldsymbol{v_i}) = A_{1,i}\boldsymbol{w_1} + \cdots + A_{m,i}\boldsymbol{w_m}.$$

If the bases of V and W are not clear from context (but important enough to name), we write $\mathcal{M}(T, \mathcal{B}, \mathcal{C})$.

Visually

$$\begin{array}{c} \\ \boldsymbol{w_1} \\ \boldsymbol{w_2} \\ \\ \boldsymbol{w_m} \end{array} \begin{array}{c} \begin{array}{cccc} \boldsymbol{v_1} & \boldsymbol{v_2} & \quad & \boldsymbol{v_n} \end{array} \\ \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{pmatrix} \end{array}$$

Example 3C.1

Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear map given by

$$T(x, y, z) = (2x + 3y, 5x - 6y + 7z)$$

Find $\mathcal{M}(T)$ (using the standard bases for $\mathbb{R}^3$ and $\mathbb{R}^2$).

We have that

$$T(1, 0, 0) = (2, 5) = 2(1, 0) + 5(0, 1)$$

$$\begin{aligned} T(0,1,0) &= (3,-6) = 3(1,0) - 6(0,1) \\ T(0,0,1) &= (0,7) = 0(1,0) + 7(0,1) \end{aligned}$$

and therefore $\mathcal{M}(T) = \begin{pmatrix} 2 & 3 & 6 \\ 5 & -6 & 7 \end{pmatrix}$.

Example 3C.2

Let $D : \mathcal{P}_3(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ be the derivative map (c.f. Example 3A.3):

$$D(p(x)) = \frac{d}{dx}[p(x)].$$

Find $\mathcal{M}(D)$ using the standard polynomial basis $\{1, x, x^2, x^3, \ldots, x^n\}$

We have that

$$D(1)0 = 0 + 0x + 0x^2 D(x) \quad 1 = 1 + 0x + 0x^2 D(x^2) 2x = 0 + 2x + 0x^2 D(x^3) \quad 3x^2 = 0 + 0x + 3x^2$$

and therefore $\mathcal{M}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$.

Exercise 3C.3

Let $\{\boldsymbol{e_1}.\boldsymbol{e_2}, \boldsymbol{e_3}\}$ denote the standard basis for $\mathbb{R}^3$, and let $\varepsilon_1, \varepsilon_2, \varepsilon_3 \in \mathcal{L}(\mathbb{R}^3, \mathbb{R})$ be the linear maps given by

$$\varepsilon_i(x,y,z) = \boldsymbol{e_i} \cdot (x,y,z) \qquad \text{(the usual dot product).}$$

Find $\mathcal{M}(5\varepsilon_1 - 7\varepsilon_2 + 11\varepsilon_3)$

Definition: matrix sum

The **sum** of two $m \times n$ matrices, A and B, is the $m \times n$ matrix whose (i,j) entry is $A_{i,j} + B_{i,j}$. That is

$$\begin{aligned} &\begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ \vdots & \vdots & & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{pmatrix} + \begin{pmatrix} B_{1,1} & B_{1,2} & \cdots & B_{1,n} \\ \vdots & \vdots & & \vdots \\ B_{m,1} & B_{m,2} & \cdots & B_{m,n} \end{pmatrix} \\ &= \begin{pmatrix} A_{1,1}+B_{1,1} & A_{1,2}+B_{1,2} & \cdots & A_{1,n}+B_{1,n} \\ \vdots & \vdots & & \vdots \\ A_{m,1}+B_{m,1} & A_{m,2}+B_{m,2} & \cdots & A_{m,n}+B_{m,n} \end{pmatrix}. \end{aligned}$$

Proposition 3C.4

For any two linear maps $S, T \in \mathcal{L}(V, W)$

$$\mathcal{M}(S+T) = \mathcal{M}(S) + \mathcal{M}(T).$$

Proof. Let $\{\boldsymbol{v_1}, \dots, \boldsymbol{v_n}\}$ be the basis for V and $\{\boldsymbol{w_1}, \dots, \boldsymbol{w_m}\}$ the basis for W. Then, for each $\boldsymbol{v_j}$, we have

$$\begin{aligned} S(\boldsymbol{v_j}) &= A_{1,j}\boldsymbol{w_1} + \cdots + A_{m,j}\boldsymbol{w_j} \\ T(\boldsymbol{v_j}) &= B_{1,j}\boldsymbol{w_1} + \cdots + B_{m,j}\boldsymbol{w_j} \\ (S+T)(\boldsymbol{v_j}) &= (A_{1,j}+B_{1,j})\boldsymbol{w_1} + \cdots + (A_{m,j}+B_{m,j})\boldsymbol{w_j} \end{aligned}$$

Clearly then each (i,j)-entry of $\mathcal{M}(S) + \mathcal{M}(T)$ is equal to the (i,j)-entry of $\mathcal{M}(S+T)$, and therefore these two matrices are equal. ☐

Definition: scalar multiple of matrix

Let λ be a scalar and A an $m \times n$ matrix. The **scalar multiple** of A by λ is an $m \times n$ matrix λA whose (i,j)-entry is $\lambda A_{i,j}$. That is

$$\lambda \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ \vdots & \vdots & & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{pmatrix} = \begin{pmatrix} \lambda A_{1,1} & \lambda A_{1,2} & \cdots & \lambda A_{1,n} \\ \vdots & \vdots & & \vdots \\ \lambda A_{m,1} & \lambda A_{m,2} & \cdots & \lambda A_{m,n} \end{pmatrix}.$$

Proposition 3C.5

For any two linear maps $T \in \mathcal{L}(V, W)$ and any scalar λ,

$$\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T).$$

Proof. This proof is very straightforward and similar to the proof of Proposition 3C.4. ☐

Theorem 3C.6

$\mathbb{K}^{m,n}$ is a vector space of dimension mn.

Proof. Let $E_{i,j}$ denote the matrix that is zero everywhere except at entry (i,j).

$$E_{1,1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \quad E_{1,2} = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \quad \cdots \quad E_{m,n} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

It is completely straightforward to verify that $\mathbb{K}^{m,n}$ is a vector space with basis $\{E_{1,1}, E_{1,2}, \dots, E_{m,n}\}$. ☐

Definition: matrix product

Let A be an $m \times n$ matrix and B an $n \times p$ matrix. Then the **product** of A and B is an $m \times p$ matrix AB whose (i, j) entry is $\sum_{k=1}^{n} A_{i,k}B_{k,j}$.

Example 3C.7

Compute the following matrix product:

$$AB = \begin{pmatrix} 2 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

$$\begin{aligned} AB = \begin{pmatrix} 2 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} &= \begin{pmatrix} \sum_{k=1}^{2} A_{1,k}B_{k,1} & \sum_{k=1}^{2} A_{1,k}B_{k,2} & \sum_{k=1}^{2} A_{1,k}B_{k,3} \\ \sum_{k=1}^{2} A_{2,k}B_{k,1} & \sum_{k=1}^{2} A_{2,k}B_{k,2} & \sum_{k=1}^{2} A_{2,k}B_{k,3} \\ \sum_{k=1}^{2} A_{3,k}B_{k,1} & \sum_{k=1}^{2} A_{3,k}B_{k,2} & \sum_{k=1}^{2} A_{3,k}B_{k,3} \end{pmatrix} \\ &= \begin{pmatrix} (2)(1) + (0)(0) & 2(1) + 0(1) & 2(0) + 0(1) \\ & & \\ & & \end{pmatrix} = \begin{pmatrix} 2 & 2 & 0 \\ 0 & 2 & 2 \\ 1 & 2 & 1 \end{pmatrix} \end{aligned}$$

Example 3C.8

Let $S : \mathbb{R}^2 \to \mathbb{R}^3$ and $T : \mathbb{R}^3 \to \mathbb{R}^2$ be linear maps given by

$$S(x, y) = (2x, 2y, x + y) \qquad \text{and} \qquad T(x, y, z) = (x + y, y + z).$$

Find the matrix of the composition $\mathcal{M}(ST)$ (using the standard basis for $\mathbb{R}^n$).

INCOMPLETE

Theorem 3C.9

If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then

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

Proof. This is an exercise in symbol-pushing. ☐

Proposition 3C.10: More reasonable descriptions of matrix product

Let A be an $m \times n$ matrix and B an $n \times p$ matrix. Letting $\boldsymbol{a_i}$ denote row i of A (thought of as a $1 \times m$ matrix) and let $\boldsymbol{b_j}$ denote column j of B (thought of as an $n \times 1$ matrix). Then

1. The (i, j)-entry of AB is $\boldsymbol{a_i b_j}$ (or, more precisely, the $(1, 1)$ entry of this 1×1 matrix).
2. Column j of AB is given by $A\boldsymbol{b_j}$.
3. Row i of AB is given by $\boldsymbol{a_i}B$.

Proof. This proof is a straightforward from the definition of matrix multiplication. ☐

Visually,

1. $$AB = \begin{pmatrix} \boldsymbol{a_1 b_1} & \boldsymbol{a_1 b_2} & \cdots & \boldsymbol{a_1 b_n} \\ \boldsymbol{a_2 b_1} & \boldsymbol{a_2 b_2} & \cdots & \boldsymbol{a_2 b_n} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{a_m b_1} & \boldsymbol{a_1 b_2} & \cdots & \boldsymbol{a_m b_n} \end{pmatrix}$$

2. $$AB = \begin{pmatrix} & & \\ & A & \\ & & \end{pmatrix} \begin{pmatrix} | & | & & | \\ \boldsymbol{b_1} & \boldsymbol{b_2} & \cdots & \boldsymbol{b_n} \\ | & | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ A\boldsymbol{b_1} & A\boldsymbol{b_2} & \cdots & A\boldsymbol{b_n} \\ | & | & & | \end{pmatrix}$$

3. $$AB = \begin{pmatrix} — & \boldsymbol{a_1} & — \\ — & \boldsymbol{a_2} & — \\ & \vdots & \\ — & \boldsymbol{a_m} & — \end{pmatrix} \begin{pmatrix} & & \\ & B & \\ & & \end{pmatrix} = \begin{pmatrix} — & \boldsymbol{a_1}B & — \\ — & \boldsymbol{a_2}B & — \\ & \vdots & \\ — & \boldsymbol{a_m}B & — \end{pmatrix}.$$

Remark. This proposition is hinting at an important computational fact: Given a matrix A, the product XA is a function on the rows of A, and the product AX is a function on the columns of A.

Proposition 3C.11: More visual ways of thinking about matrix products

Let A be an $m \times n$ matrix and B an $n \times p$ matrix. Letting $\boldsymbol{a_j}$ denote column j of A (thought of as a $m \times 1$ matrix) and let $\boldsymbol{b_i}$ denote row i of B (thought of as an $1 \times p$ matrix). Then

1. The columns of AB are linear combinations of the columns of A (with coefficients given by the column entries of B).
2. The rows of AB are linear combinations of the rows of B (with coefficients given by the row entries of A).

Proof. This proof is a straightforward from the definition of matrix multiplication. ☐

Visually,

1. $$AB = \begin{pmatrix} | & | & & | \\ \boldsymbol{a_1} & \boldsymbol{a_2} & \cdots & \boldsymbol{a_n} \\ | & | & & | \end{pmatrix} \begin{pmatrix} & B_{1,j} & \\ & B_{2,j} & \\ & \vdots & \\ & B_{n,j} & \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \sum_{i=1}^{n} B_{i,1}\boldsymbol{a_i} & \sum_{i=1}^{n} B_{i,2}\boldsymbol{a_i} & \cdots & \sum_{i=1}^{n} B_{i,n}\boldsymbol{a_i} \\ | & | & & | \end{pmatrix}$$

2. $AB = \begin{pmatrix} A_{i,1} & A_{i,2} & \cdots & A_{i,n} \end{pmatrix} \begin{pmatrix} — & \boldsymbol{b_1} & — \\ — & \boldsymbol{b_2} & — \\ & \vdots & \\ — & \boldsymbol{b_n} & — \end{pmatrix} = \begin{pmatrix} — & \sum_{j=1}^{n} A_{1,j}\boldsymbol{b_j} & — \\ & \vdots & \\ — & \sum_{j=1}^{n} A_{m,j}\boldsymbol{b_j} & — \end{pmatrix}$

Definition: transpose

Let $A = [A_{i,j}]$ be an $m \times n$ matrix. The **transpose** of A, denoted A^t, is the $n \times m$ matrix whose (i, j) entry is the (j, i)-entry from A.

Example 3C.12

Find the transpose of $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$.

$A^t = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}$

Proposition 3C.13

For any $m \times n$ matrix A and $n \times p$ matrix B,

$$(AB)^t = B^t A^t.$$

Proof. This is straightforward. Just look at the (i, j)-entry of each matrix. □

Definition: column space, row space

Let A be an $m \times n$ matrix. The **column space** of A, denoted $\mathrm{Col}(A)$, is the span of the columns of A (thought of as a subspace of $\mathbb{K}^m$). The **row space** of A, denoted $\mathrm{Row}(A)$, is the span of the rows of A (through of as a subspace of $\mathbb{K}^n$). One could also think of $\mathrm{Row}(A) = \mathrm{Col}(A^T)$.

Observe that we can decompose a matrix A into a product of matrices BC where B is compried of the basis for $\mathrm{Col}(A)$:

$$\begin{aligned} A = \begin{pmatrix} | & | & | & | \\ \boldsymbol{a_1} & \boldsymbol{a_2} & \boldsymbol{a_3} & \boldsymbol{a_4} \\ | & | & | & | \end{pmatrix} &= \begin{pmatrix} | & | & | & | \\ \boldsymbol{b_1} & \boldsymbol{b_2} & (\alpha\boldsymbol{b_1} + \beta\boldsymbol{b_3} + \gamma\boldsymbol{b_3}) & \boldsymbol{b_3} \\ | & | & | & | \end{pmatrix} \\ &= \begin{pmatrix} | & | & | \\ \boldsymbol{b_1} & \boldsymbol{b_2} & \boldsymbol{b_3} \\ | & | & | \end{pmatrix} \begin{pmatrix} 1 & 0 & \alpha & 0 \\ 0 & 1 & \beta & 0 \\ 0 & 0 & \gamma & 1 \end{pmatrix} = BC. \end{aligned}$$

This will be a very useful construction in proving the next result. We also observe that C has linearly independent rows, since each row contains a 1 in some column which is otherwise all 0's (so no linear combination of the other rows can ever result in anything but a 0 in that column).

Theorem 3C.14: Row Rank = Column Rank

Let A be an $m \times n$ matrix. Then

$$\dim(\mathrm{Row}(A)) = \dim(\mathrm{Col}(A))$$

Proof. Let A be an $m \times n$ matrix. We employ the construction/decomposition immediately preceding this theorem. Let $\boldsymbol{a_j}$ denote the j^{th} column of A. We can reduce $\{\boldsymbol{a_1}, \dots, \boldsymbol{a_n}\}$ to a basis $\{\boldsymbol{b_1}, \dots, \boldsymbol{b_r}\}$ for $\mathrm{Col}(A)$. As such, for each $\boldsymbol{a_j}$, there are scalars $C_{1,j}, \dots, C_{r,j}$ for which

$$\boldsymbol{a_j} = C_{1,j}\boldsymbol{b_1} + \cdots + C_{r,j}\boldsymbol{b_r}$$

Let $B = \begin{pmatrix} | & & | \\ \boldsymbol{b_1} & \cdots & \boldsymbol{b_r} \\ | & & | \end{pmatrix}$ and $C = [C_{i,j}]$. We note that B is an $m \times r$ matrix and C is an $r \times n$ matrix. Then, by construction, $A = BC$ and it is easily seen that $\dim(\mathrm{Col}(B)) = r$. As well, C has linearly independent rows since each row contains a column where all but one entry is 0, hence $\dim(\mathrm{Row}(C)) = r$.

Every row of A is a linear combination of the rows of C, hence

$$\dim(\mathrm{Row}(A)) \leq \dim(\mathrm{Row}(C)) = r = \dim(\mathrm{Col}(A)).$$

Applying the same construction and argument to A^t, one achieves

$$\dim(\mathrm{Row}(A^t)) \leq \dim(\mathrm{Col}(A^t)).$$

Combining these results, one achieves

$$\dim(\mathrm{Row}(A)) \leq \dim(\mathrm{Col}(A)) = \dim(\mathrm{Row}(A^t)) \leq \dim(\mathrm{Col}(A^t)) = \dim(\mathrm{Row}(A))$$

and hence

$$\dim(\mathrm{Row}(A)) = \dim(\mathrm{Col}(A)).$$

$\square$

Definition: rank of a matrix

Given an $m \times n$ matrix A, the **rank** of A, denoted $\mathrm{rank}(A)$, is the dimension of the column space of A (or equivalently, the dimension of the row space of A).

Corollary 3C.15

$\mathrm{rank}(A) = \mathrm{rank}(A^t)$.

3D Invertibility and Isomorphisms

Definition: inverse

Let V, W be vector spaces and $T \in \mathcal{L}(V, W)$. T is called **invertible** if there exists $T^{-1} \in \mathcal{L}(W, V)$ so that $T^{-1}T = \mathrm{Id}_V$ and $TT^{-1} = \mathrm{Id}_W$ (the identity maps on V and W, respectively). T^{-1} is called the **inverse** of T.

Our use of the word “the” above is excusable, because

Proposition 3D.1

The inverse is unique.

Proof. Suppose S_1, S_2 are both inverses for T. Then

$$S_1 = S_1\mathrm{Id}_W = S_1(TS_2) = (S_1T)S_2 = \mathrm{Id}_V S_2 = S_2.$$

□

Example 3D.2

The map $T_\theta \in \mathcal{L}(R^2)$ given by

$$T_\theta(x, y) = (x\cos(\theta) - y\sin(\theta),\ x\sin(\theta) + y\cos(\theta)$$

is a rotation of the plane (counterclockwise) by an angle of θ. Find T^{-1}.

The inverse is gien by

$$(T_\theta)^{-1} = T_{-\theta}(x, y) = (x\cos(\theta) + y\sin(\theta),\ -x\sin(\theta) + y\cos(\theta)$$

We can check this explicitly:

$$\begin{aligned} T_{-\theta}T_\theta\begin{pmatrix} x \\ y \end{pmatrix} &= T_{-\theta}\begin{pmatrix} x\cos(\theta) - y\sin(\theta) \\ x\sin(\theta) + y\cos(\theta) \end{pmatrix} \\ &= \begin{pmatrix} x\cos(\theta)\cos(-\theta) - y\sin(\theta)\cos(-\theta) + x\sin(\theta)(-\sin(-\theta)) + y\cos(\theta)(-\sin(-\theta)) \\ x\cos(\theta)\sin(-\theta) - y\sin(\theta)\sin(-\theta) + x\sin(\theta)\cos(\theta) + y\cos(\theta)\cos(\theta) \end{pmatrix} \end{aligned}$$

Remembering that $\cos(-\theta) = \cos(\theta)$, $\sin(-\theta) = -\sin(\theta)$, and $\cos^2(\theta) + \sin^2(\theta) = 1$, the above expression simplifies to just $\begin{pmatrix} x \\ y \end{pmatrix}$, hence $T_{-\theta}T_\theta = \mathrm{Id}_{\mathbb{R}^2}$.

Theorem 3D.3

A linear map is invertible if and only if it is both one-to-one and onto.

This follows immediately from the following facts (established as exercises):

- $T \in \mathcal{L}(V, W)$ is one-to-one if and only if there exists $S \in \mathcal{L}(W, V)$ so that $ST = \mathrm{Id}_V$.

- $T \in \mathcal{L}(V, W)$ is onto if and only if there exists $S \in \mathcal{L}(W, V)$ so that $TS = \mathrm{Id}_W$.

We prove it concretely.

Proof. Suppose $T \in \mathcal{L}(V, W)$.

($\implies$) Suppose that T is invertible.

- To see that T is injective, let $\boldsymbol{u}, \boldsymbol{v} \in V$ and suppose $T(\boldsymbol{u}) = T(\boldsymbol{w})$. Then, since T^{-1} is a well-defined function, $T^{-1}T(\boldsymbol{u}) = T^{-1}T(\boldsymbol{v})$ and
$$\boldsymbol{u} = T^{-1}T(\boldsymbol{u}) = T^{-1}T(\boldsymbol{v}) = \boldsymbol{v}.$$
- To see that T is surjective, let $\boldsymbol{w} \in W$ and let $\boldsymbol{v} = T^{-1}(\boldsymbol{w})$. Then $T(\boldsymbol{v}) = TT^{-1}(\boldsymbol{w}) = \boldsymbol{w} \in \mathrm{Range}(T)$.

($\impliedby$) Suppose that T is both injective and surjective.

The strategy for this proof is that we're going to

1. define an "inverse" S explicitly,
2. verify that $ST = \mathrm{Id}_V$ and $TS = \mathrm{Id}_W$,
3. verify that S is indeed a linear map, making it the actual inverse.

Since T is surjective, then for every $\boldsymbol{w} \in W$ there is some $\boldsymbol{v} \in V$ so that $T(v) = \boldsymbol{w}$. By injectivity, this vector $\boldsymbol{v}$ is unique. As such, we define S to be the unique map so that $T \circ S(\boldsymbol{w}) = \boldsymbol{w}$. By construction, $T \circ S = \mathrm{Id}_W$. To see that $ST = \mathrm{Id}_V$, let $\boldsymbol{v} \in V$. Then by associativity of function composition,
$$T(S \circ T(\boldsymbol{v})) = (T \circ S)(T(\boldsymbol{v})) = \mathrm{Id}_W \circ T(\boldsymbol{v}) = T(\boldsymbol{v}).$$

Since T is injective, $S \circ T(\boldsymbol{v}) = \boldsymbol{v}$, hence $S \circ T = \mathrm{Id}_V$.

Finally, we verify that S is linear. Let k be a scalar and $\boldsymbol{v_1}, \boldsymbol{v_2} \in V$. Then
$$\begin{aligned} T(S(k\boldsymbol{v_1} + \boldsymbol{v_2})) &= k\boldsymbol{v_1} + \boldsymbol{v_2} \\ &= kT(S(\boldsymbol{v_1})) + T(S(\boldsymbol{v_2})) \\ &= T(kS(\boldsymbol{v_1} + S(\boldsymbol{v_2})) \end{aligned}$$

and since T is one-to-one,
$$S(k\boldsymbol{v_1} + \boldsymbol{v_2}) = kS(\boldsymbol{v_1}) + S(\boldsymbol{v_2}).$$

□

A straightforward application of the Fundamental Theorem of Linear Maps (AKA Rank–Nullity Theorem) yields the following:

Corollary 3D.4

Suppose V, W are $\mathbb{K}$-vector spaces of the same finite dimension. For any $T \in \mathcal{L}(V, W)$, the following are equivalent

1. T is invertible.
2. T is injective.
3. T is surjective.

The previous corollary does not apply to infinite-dimensional vector spaces.

Example 3D.5

Is the derivative map $D : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ invertible?

D is surjective, but not injective (since two polynomials which differ by a constant have the same derivative). Even if we tried to define an inverse, notice that

$$D(x^2 + 2) = D(x^2) = 2x$$

So any theoretical D^{-1} would satisfy

$$D(2x) = x^2 + 2 \qquad \text{and } D(2x) = x^2$$

which fails to be a well-defined function.

Example

Is the "multiplicatio by x^3 map" (c.f. Example 3A.5)

$$\begin{aligned} T : \mathcal{P}(\mathbb{K}) &\to \mathcal{P}(\mathbb{K}) \\ T(p(x)) &= x^3 p(x)\, dx \end{aligned}$$

invertible?

T is injective. Indeed, whenever $p(x)$ is not the zero polynomial, $\deg(x^3 p) \geq 3$, so $\text{Null}(T) = \{0\}$ (the zero polynomial). However T is not surjective, since there are no polynomials which map to x, for example. Even if we tried to define an inverse T^{-1}, notice that we could not define $T^{-1}(x)$.

However, if one can restrict to a finite-dimensional subspace of an infinite-dimensional vector space, then we can apply Corollary 3D.4.

Example 3D.6

For any polynomial $q(x) \in \mathcal{P}(\mathbb{R})$, there exists a polynomial $p(x) \in \mathcal{P}(\mathbb{R})$ so that $\frac{d^5}{dx}\big[(x^5 + 5x + 7)p(x)\big] = q(x)$.

Let $q(x)$ be a polynomial of degree n. Observe that if $p(x)$ has degreen m, then $(x^5 + 5x + 7)$ has degree $m + 5$, and then the fifth derivative reduces this to degree $(m + 5) - 5 = m$. We can therefore define the map

$$\begin{aligned} T : \mathcal{P}_n(\mathbb{R}) &\to \mathcal{P}_n(\mathbb{R}) \\ T(p(x)) &= \frac{d^5}{dx}\big[(x^5 + 5x + 7)p(x)\big] = q(x) \end{aligned}$$

which is linear (it is a composition of the linear maps: the derivative map, and the multiply-by-a-polynomial map).

Now, all polynomials whose fifth derivative is 0 are degree at most 4, and $(x^5 + 5x + 7)p(x)$ always has degree at least 5. Since $4 < 5$, there are no nonzero polynomials $p(x)$ for which $T(p(x)) = 0$. Therefore $\text{Null}(T) = \{\mathbf{0}\}$ and T is injective. By Corollary 3D.4, T is surjective, hence there is a polynomial $p(x)$ for which $T(p(x)) = q(x)$, as desired.

3D.I Isomorphic Vector Spaces

Definition: isomorphism,isomorphic

Let V and W be $\mathbb{K}$-vector spaces. If $T \in \mathcal{L}(V, W)$ is invertible, we say that T is an **isomorphism**. The vector spaces V and W are **isomorphic** if there exists an isomorphism $T : V \to W$ (or $S : W \to V$).

Remark. The term “isomorphism” is etymologically based in ancient Greek, where “iso” comes from the word meaning “the same” and “morphism” comes from the word meaning “shape/form.” An isomorphism is essentially just a relabeling of (basis) vectors.

Remark. For those who have seen isomorphisms in other algebraic contexts like groups/rings/fields, this is the exact same notion as (1) linear maps are the morphisms in the category of vector spaces and (2) invertible maps are necessarily bijections.

Theorem 3D.7

Let V, W be finite-dimensional $\mathbb{K}$-vector spaces. Then V and W are isomorphic if and only if $\dim(V) = \dim(W)$.

Proof.

(⇒) Suppose V and W are isomorphic. By definition, there is an isomorphism $T : V \to W$, which means T is both injective an surjective. Injectivitiy and surjectivity thus imply that

$$\text{Null}(T) = \{\mathbf{0}\} \qquad \text{and} \qquad \text{Range}(T) = W,$$

respectively. By the Rank–Nullity Theorem,

$$\dim(V) = \dim(\text{Null}(T)) + \dim(\text{Range}(T)) = 0 + \dim(W).$$

(⇐) Suppose now that $\dim(V) = \dim(W) = n$. Let $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}\}$ and $\{\boldsymbol{w_1}, \ldots, \boldsymbol{w_n}\}$ be bases for V and W, respectively.

Our strategy from here is going to be to explicitly define a map and verify that it is an isomorphism.

Just as in the Linear Map Lemma, define $T \in \mathcal{L}(W)$ by

$$T(\alpha_1 \boldsymbol{v_1} + \cdots + \alpha_n \boldsymbol{v_n}) = \alpha_1 \boldsymbol{w_1} + \cdots + \alpha_n \boldsymbol{w_n}.$$

T is a linear map.

($\boldsymbol{T}$ is injective.) Since the $\boldsymbol{w}_i$'s form a linearly independent set, the linear combination on the right is $\boldsymbol{0}$ precisely when each of the coefficients is 0, hence $\mathrm{Null}(T) = \{\boldsymbol{0}\}$ and thus T is injective.

($\boldsymbol{T}$ is surjective.) Since the $alpha_i$'s range over all $\mathbb{K}$-values, we must have that $\mathrm{Range}(T) = \mathrm{Span}(\{\boldsymbol{w_1}, \ldots, \boldsymbol{w_n}\}) = W$, hence T is surjective.

Therefore T is an isomorphism.

□

Lemma 3D.8

Suppose V, W are $\mathbb{K}$-vector spaces with $\dim(V) = n$ and $\dim(W) = m$ and bases $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}\}$ and $\{\boldsymbol{w_1}, \ldots, \boldsymbol{w_m}\}$, respectively. For each $i = 1, \ldots, n$ and $j = 1, \ldots, m$, define the maps

$$\varphi_{i,j}(\alpha_1 \boldsymbol{v_1} + \cdots + \alpha_n \boldsymbol{v_n}) = \alpha_j \boldsymbol{w_i}.$$

The set $\{\varphi_{1,1}, \ldots, \varphi_{n,m}\}$ is a basis for $\mathcal{L}(V, W)$.

Proof. Let $\boldsymbol{u} = \alpha_1 \boldsymbol{v_1} + \cdots + \alpha_n \boldsymbol{v_n}$ and note that

$$\varphi_{i,j}(\boldsymbol{u}) = \varphi_{i,j}(\alpha_1 \boldsymbol{v_1} + \cdots + \alpha_n \boldsymbol{v_n}) = \varphi_{i,j}(\alpha_j \boldsymbol{v_j}).$$

This observation is precisely what is applied to pass from Equation 3.2 to Equation 3.3.

($\{\boldsymbol{\varphi_{1,1}, \ldots, \varphi_{n,m}}\}$ spans $\boldsymbol{\mathcal{L}(V, W)}$). Let $T \in \mathcal{L}(V, W)$ be arbitrary. Then there are scalars $A_{i,j}$ for which

$$\begin{aligned}
T(\boldsymbol{u}) &= \alpha_1 T(\boldsymbol{v_1}) + \cdots + \alpha_n T(\boldsymbol{v_n}) \\
&= \alpha_1 \sum_{j=1}^{m} A_{j,1} \boldsymbol{w_j} + \cdots \alpha_n \sum_{j=1}^{m} A_{j,n} \boldsymbol{w_j} \\
&= \alpha_1 \sum_{j=1}^{m} A_{j,1} \varphi_{j,1}(\boldsymbol{v_1}) + \cdots + \alpha_n \sum_{j=1}^{m} A_{j,n} \varphi_{j,n}(\boldsymbol{v_n}) \\
&= \sum_{j=1}^{m} A_{1,j} \varphi_{j,1}(\alpha_1 \boldsymbol{v_1}) + \cdots + \sum_{j=1}^{m} A_{j,n} \varphi_{j,n}(\alpha_n \boldsymbol{v_n}) \qquad (3.2) \\
&= \sum_{j=1}^{m} A_{1,j} \varphi_{j,1}(\boldsymbol{u}) + \cdots + \sum_{j=1}^{m} A_{j,n} \varphi_{j,n}(\boldsymbol{u}) \qquad (3.3) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} A_{j,i} \varphi_{j,i}(\boldsymbol{u})
\end{aligned}$$

and therefore every T is a linear combination of the $\varphi_{i,j}$'s.

($\{\varphi_{i,1}, \ldots, \varphi_{m,n}\}$ is linearly independent) . Suppose there are scalars $A_{i,j}$ so that, for all $\boldsymbol{v} \in V$,

$$\sum_{i=1}^{n}\sum_{j=1}^{m} A_{j,i}\varphi_{j,i}(\boldsymbol{v}) = \mathbf{0}.$$

In particular, when $\boldsymbol{v} = \boldsymbol{v_1}$, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{m} A_{j,1}\varphi_{1,j}(\boldsymbol{v_1}) = A_{1,1}\varphi_{1,1}(\boldsymbol{v_1}) = \mathbf{0}$$

and therefore $A_{1,1} = 0$. Repeating this over all $i = 1, \ldots, n$ and $j = 1, \ldots, m$ implies that only the trivial linear combination can result in the constant zero function.

□

Example 3D.9

Let $\varphi_{i,j} \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$ be as in the last lemma. Find the matrices $\mathcal{M}(\varphi_{i,j})$ for each $i = 1, 2$ and $j = 1, 2$.

$$\begin{array}{ccc} \varphi_{1,1}(x,y) = (x,0,0) & \varphi_{1,2}(x,y) = (0,x,0) & \varphi_{1,3}(x,y) = (0,0,x) \\ \varphi_{2,1}(x,y) = (y,0,0) & \varphi_{2,2}(x,y) = (0,y,0) & \varphi_{2,3}(x,y) = (0,0,y) \end{array}$$

The matrices associated with these are

$$\mathcal{M}(\varphi_{1,1}) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \mathcal{M}(\varphi_{1,2}) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \mathcal{M}(\varphi_{1,3}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}$$

$$\mathcal{M}(\varphi_{2,1}) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \mathcal{M}(\varphi_{2,2}) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \mathcal{M}(\varphi_{2,3}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$$

These matrices are clearly a basis for $\mathbb{R}^{3,2}$.

Theorem 3D.10

If V, W are $\mathbb{K}$-vector spaces with $\dim(V) = n$ and $\dim(W) = m$, then $\mathcal{L}(V, W)$ is isomorphic to $\mathbb{K}^{m,n}$. Specifically, the "matrix for a linear transformation," $\mathcal{M}(\,\cdot\,)$ is the isomorphism (hence the somewhat clunky, function-y notation).

Proof. The claim that $\mathcal{L}(V, W)$ and $\mathbb{K}^{m,n}$ are isomorphic is immediate given that they have the same dimensions. It's only that $\mathcal{M}$ is an isomorphism that needs to be verified, and this is left as an exercise to the reader. □

Corollary 3D.11

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

3D.II Linear Maps and Matrix Multiplication

Definition: matrix of a vector

Let V be a finite-dimensional vector space with basis $\mathcal{B} = \{\boldsymbol{v_1}, \dots, \boldsymbol{v_n}\}$. For every $\boldsymbol{u} \in V$, the **matrix** of $\boldsymbol{u}$ (relative to the $\mathcal{B}$-basis), $\mathcal{M}(\boldsymbol{u}, \mathcal{B})$ (or just $\mathcal{M}(\boldsymbol{u})$ if the basis is understood) is the $n \times 1$ matrix

$$\mathcal{M}(\boldsymbol{u}, \mathcal{B}) = \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{pmatrix}$$

where $k_1, \dots, k_n \in \mathbb{K}$ satisfy

$$\boldsymbol{u} = k_1 \boldsymbol{v_1} + \cdots + k_n \boldsymbol{v_n}.$$

Example 3D.12

Let $V = \mathcal{P}_3(\mathbb{K})$ and consider the following two bases for V: $\mathcal{E} = \{1, x, x^2, x^3\}$ (the "standard" polynomial basis), and $\mathcal{B} = \{1 + x^2,\ x + x^3,\ x + 2x^3,\ x^2 + x^3\}$.

1. Find $\mathcal{M}(1 + x^2 + 3x^3,\ \mathcal{E})$.
2. Find $\mathcal{M}(1 + x^2 + 3x^3,\ \mathcal{B})$.

1. For simplicity, label the basis polynomials $e_1(x) = 1$, $e_2(x) = x$, $e_3(x) = x^2$, $e_4(x) = x^3$. Now it's easy to see that

$$1 + x^2 + 3x^3 = 1e_1(x) + 0e_2(x) + 1e_3(x) + 3e_4(x)$$

hence

$$\mathcal{M}(1 + x^2 + 3x^3) = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 3 \end{pmatrix}.$$

2. For simplicity, label the basis polynomials $b_1(x) = 1 + x^2$, $b_2(x) = x + x^3$, $b_3(x) = x + 2x^3$, $b_4(x) = x^2 + x^3$. Now we aim to find constants k_1, k_2, k_3, k_4 so that

$$1 + x^2 + 3x^3 = k_1 b_1(x) + k_2 b_2(x) + k_3 b_3(x) + k_4 b_4(x).$$

Expanding out the right-hand side (with a bit of rearranging), we get

$$1 + x^2 + 3x^3 = k_1 + (k_2 + k_3)x + (k_1 + k_4)x^2 + (k_2 + 2k_3 + k_4)x^3$$

and this leads us to see that $k_1 = 1$, $k_4 = 0$, and then (solving a small system) that $k_2 = -3$ and $k_3 = 3$. We thus have that

$$\mathcal{M}(1 + x^2 + 3x^3) = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \\ 3 \\ 0 \end{pmatrix}.$$

Theorem 3D.13: Linear maps as matrix multiplication

Let $T \in \mathcal{L}(V, W)$, where both V and W are finite-dimensional. Then, for all $\boldsymbol{v} \in V$, we have that

$$\mathcal{T}(\sqsubseteq) = \mathcal{M}(T)\mathcal{M}(\boldsymbol{v}).$$

Proof. Let $\{\boldsymbol{v_1}, \dots, \boldsymbol{v_n}\}$ be a basis for V and let $\boldsymbol{u} = \alpha_1\boldsymbol{v_1} + \cdots + \alpha_n\boldsymbol{v_n}$. Then for $T \in \mathcal{L}(V, W)$, we have

$$\begin{aligned}
T(\boldsymbol{u}) &= \alpha_1 T(\boldsymbol{v_1}) + \cdots + \alpha_n T(\boldsymbol{v_n}) \\
&= \alpha_1 \sum_{j=1}^{m} A_{j,1}\boldsymbol{w_j} + \cdots \alpha_n \sum_{j=1}^{m} A_{j,n}\boldsymbol{w_j}
\end{aligned}$$

from which it follows that

$$\begin{aligned}
\mathcal{M}(T(\boldsymbol{u})) &= \alpha_1 \mathcal{M}\left(\sum_{j=1}^{m} A_{j,1}\boldsymbol{w_j}\right) + \cdots \alpha_n \mathcal{M}\left(\sum_{j=1}^{m} A_{j,n}\boldsymbol{w_j}\right) \\
&= \alpha_1 \begin{pmatrix} A_{1,1} \\ \vdots \\ A_{m,1} \end{pmatrix} + \cdots + \alpha_n \begin{pmatrix} A_{1,n} \\ \vdots \\ A_{m,n} \end{pmatrix} \\
&= \mathcal{M}(T)\mathcal{M}(\boldsymbol{u}).
\end{aligned}$$

□

Example 3D.14

Verify Theorem 3D.13 using the following

$$\begin{aligned}
T &: \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_3(\mathbb{R}) \\
T(p(x)) &= (x-5)p(x)
\end{aligned}$$

(where each polynomial vector space is assumed to be using the standard basis).

INCOMPLETE Let $p(x) = a_0 + a_1x + a_2x^2$. Then we have that

$$\mathcal{M}(T)\mathcal{M}(p(x)) = \begin{pmatrix} -5 & 0 & 0 \\ 1 & -5 & 0 \\ 0 & 1 & -5 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} -5a_0 \\ a_0 - 5a_1 \\ a_1 - 5a_2 \\ a_2 \end{pmatrix}$$

Also,

$$\begin{aligned}
T(p(x)) &= (x-5)(a_0 + a_1x + a_2x^2) \\
&= a_0x + a_1x^2 + a_2x^3 - 5a_0 - 5a_1x - 5a_2x^2 \\
&= -5a_0 + (a_0 - 5a_1)x + (a_1 - 5a_2)x^2 + a_2x^3
\end{aligned}$$

$$\implies \mathcal{M}(T(p(x)) = \begin{pmatrix} -5a_0 \\ a_0 - 5a_1 \\ a_1 - 5a_2 \\ a_2 \end{pmatrix}.$$

Theorem 3D.15

Let V, W be finite-dimensional. For all linear transformations $T \in \mathcal{L}(V, W)$, $\operatorname{rank}(T) = \operatorname{rank}(\mathcal{M}(T))$.

Proof. INCOMPLETE – See Theorem 3.78 in the book. The observation is just that the matrix of any element in the range is a linear combination of the columns, so there is a correspondence between linearly independent columns and linearly independent range vectors. ☐

3D.III Change of Basis

Definition: identity matrix

The **identity matrix** is an $n \times n$ matrix $I = \left(I_{i,j}\right)$ where

$$I_{i,j} = \begin{cases} 1 & \text{when } i = j \\ 0 & \text{when } i \neq j. \end{cases}$$

Remark. The entries of the identity matrix are often written as $\delta_{i,j}$, the "**Kroenecker delta**."

Definition

An $n \times n$ matrix A is said to be **invertible** if there exists an $n \times n$ matrix A^{-1} for which $AA^{-1} = A^{-1}A = I$. The matrix A^{-1} is called the **inverse** of A.

Proposition 3D.16

The inverse of a matrix is unique.

Proof. Suppose $B - 1, B_2$ are both inverses for A. Then

$$B_1 = B_1 I = B_1(AB_2) = (B_1 A)B_2 = IB_2 = B_2.$$

☐

Given the correlation between matrix multiplication and linear maps, the following is immediate.

Theorem 3D.17

Let V, W be n-dimensional $\mathbb{K}$-vector spaces. Then $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, V)$ are inverse maps if and only if $\mathcal{M}(T)$ and $\mathcal{M}(S)$ are inverse matrices.

Let $\mathcal{B} = \{\boldsymbol{b_1}, \dots, \boldsymbol{b_n}\}$ and $\mathcal{C} = \{\boldsymbol{c_1}, \dots, \boldsymbol{c_n}\}$ be two bases for a finite-dimensional vector space, V. Consider the matrix whose j^{th} column is $\mathcal{M}(\boldsymbol{b_j}, \mathcal{C})$, the column of $\boldsymbol{b_j}$ written in the $\mathcal{C}$-basis.

Example 3D.18

Let $\mathcal{B} = \left\{ \underset{\boldsymbol{b_1}}{(1,0)}, \underset{\boldsymbol{b_2}}{(-1,1)} \right\}$ and $\mathcal{C} = \left\{ \underset{\boldsymbol{c_1}}{(1,1)}, \underset{\boldsymbol{c_2}}{(0,1)} \right\}$ be bases for $\mathbb{K}^2$.

1. Find the matrix $P = \begin{pmatrix} | & | \\ \mathcal{M}(\boldsymbol{b_1}, \mathcal{C}) & \mathcal{M}(\boldsymbol{b_2}, \mathcal{C}) \\ | & | \end{pmatrix}$.
2. Let $\boldsymbol{v} = (2,4)$. Find $\mathcal{M}(\boldsymbol{v}, \mathcal{B})$.
3. Find $\mathcal{M}((2,4), \mathcal{C})$.
4. What happens when you compute the matrix product $P\,\mathcal{M}(\boldsymbol{v}, \mathcal{B})$?

1. Observe that $\boldsymbol{b_1} = \boldsymbol{c_1} - \boldsymbol{c_2}$ and $\boldsymbol{b_2} = -\boldsymbol{c_1} + 2\boldsymbol{c_2}$. Hence $P = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$.
2. Observe that $\boldsymbol{v} = 6\boldsymbol{b_1} + 4\boldsymbol{b_2}$. Thus $\mathcal{M}(\boldsymbol{v}, \mathcal{C}) = \begin{pmatrix} 6 \\ 4 \end{pmatrix}$.
3. Observe that $\boldsymbol{v} = 2\boldsymbol{c_1} + 2\boldsymbol{c_2}$. Thus $\mathcal{M}(\boldsymbol{v}, \mathcal{C}) = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$.
4. $P\,\mathcal{M}(\boldsymbol{v}, \mathcal{B}) = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 6 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$.

We see that this matrix P in the last example represented a linear map $T : V \to V$ for which $T(\boldsymbol{v}) = \boldsymbol{v}$ (and with a little more thought, one can see that this holds for all $\boldsymbol{v}$). It must be then that T is the identity map.

Definition: change of basis matrix

Let V be a finite-dimensional $\mathbb{K}$-vector space with two bases $\mathcal{B} = \{\boldsymbol{b_1}, \dots, \boldsymbol{b_n}\}$ and $\mathcal{C} = \{\boldsymbol{c_1}, \dots, \boldsymbol{c_n}\}$. The **change of basis matrix** from the $\mathcal{B}$-basis to the $\mathcal{C}$-basis is given by

$$\mathcal{M}(\mathrm{Id}_V, \mathcal{B}, \mathcal{C}) = \begin{pmatrix} | & & | \\ \mathcal{M}(\boldsymbol{b_1}, \mathcal{C}) & \cdots & \mathcal{M}(\boldsymbol{b_n}, \mathcal{C}) \\ | & & | \end{pmatrix}$$

and this matrix has the following property: for all $\boldsymbol{v} \in V$,

$$\mathcal{M}(\mathrm{Id}_V, \mathcal{B}, \mathcal{C})\mathcal{M}(\boldsymbol{v}, \mathcal{B}) = \mathcal{M}(\boldsymbol{v}, \mathcal{C}).$$

Since matrix multiplication corresponds to composition of linear maps, it follows that

Proposition 3D.19

Let V be a finite-dimensional $\mathbb{K}$-vector space with two bases $\mathcal{B} = \{\boldsymbol{b_1}, \dots, \boldsymbol{b_n}\}$ and $\mathcal{C} = \{\boldsymbol{c_1}, \dots, \boldsymbol{c_n}\}$. Let W be a $\mathbb{K}$-vector space with basis $\mathcal{D} = \{\boldsymbol{d_1}, \dots, \boldsymbol{d_n}\}$. Write $P = \mathcal{M}(\mathrm{Id}_V, \mathcal{B}, \mathcal{C})$.

1. P is invertible and $P^{-1} = \mathcal{M}(\mathrm{Id}_V, \mathcal{C}, \mathcal{B})$.

2. Let $T \in \mathcal{L}(V)$. Writing $A_{\mathcal{B}} = \mathcal{M}(T, \mathcal{B})$ and $A_{\mathcal{C}} = \mathcal{M}(T, \mathcal{C})$,

$$A_{\mathcal{B}} = P^{-1} A_{\mathcal{C}} P$$

3. For any linear maps $S \in \mathcal{L}(W, V)$ and $T \in \mathcal{L}(V, W)$,

$$\mathcal{M}(\mathrm{Id}_V, \mathcal{B}, \mathcal{C})\ \mathcal{M}(S, \mathcal{D}, \mathcal{B}) = \mathcal{M}(S, \mathcal{D}, \mathcal{C})$$
$$\mathcal{M}(T, \mathcal{C}, \mathcal{D})\ \mathcal{M}(\mathrm{Id}_V, \mathcal{B}, \mathcal{C}) = \mathcal{M}(T, \mathcal{B}, \mathcal{D}).$$

Remark. The notation is clunky, but item 3 above just says that you can pre- or post-compose your linear map T with a change of basis to get the marix $\mathcal{M}(T)$ written in different bases.

3Z Elementary Row Operations and Reduced Row Echelon Form

INSTRUCTOR NOTE: This topic is entirely absent from the course text, so this section is an attempt to tie in some of the computational aspects of linear algebra. As such, this section is in a very raw/clunky state. That said, the target audience for this class is expected to be familiar already with computational linear algebra (from, say, a 2000-level class...), so hopefully said audience can be somewhat forgiving of notational discrepancies, mixed up indices, and the like.

Let $A \in \mathbb{K}^{m,n}$ be a rank r matrix with columns $\{\boldsymbol{a_1}, \dots, \boldsymbol{a_n}\}$. One can find a basis $\{\boldsymbol{a_{j_1}}, \dots, \boldsymbol{a_{j_r}}\}$ for $\operatorname{Col}(A)$ with the following properties:

1. $j_i < j_{i+1}$ for each $i = 1, \dots, r-1$, and
2. For every $\ell = 1, \dots, n$ where $j_i \leq \ell < j_{i+1}$,

$$\boldsymbol{a_\ell} = k_{j_1}\boldsymbol{a_{j_1}} + \dots + k_{j_i}\boldsymbol{a_{j_i}} + 0\boldsymbol{a_{j_{i+1}}} + \dots + 0\boldsymbol{a_{j_r}}$$

(In plain English, this says that the order that the vectors appear in your bases matches the order in which these columns appear in the matrix, and that every column in A can be written as a linear combination of basis columns only to its left.) For lack of a name in the literature, call this an **echelon basis**.

Extend this to an (ordered) basis $\mathcal{B} = \{\boldsymbol{a_{j_1}}, \dots, \boldsymbol{a_{j_r}}, \boldsymbol{v_{r+1}}, \dots, \boldsymbol{v_m}\}$ of $\mathbb{K}^m$, and for lack of any better name, call $\mathcal{B}$ an **echelon basis**. Finally, let $\mathcal{M}(A, \mathcal{B})$ denote the $m \times n$ matrix

$$\begin{pmatrix} | & & | \\ \mathcal{M}(\boldsymbol{a_1}, \mathcal{B}) & \cdots & \mathcal{M}(\boldsymbol{a_1}, \mathcal{B}) \\ | & & | \end{pmatrix}.$$

Remark. This is actually a change-of-basis on the codomain of a linear map $T : V \to W$. In a future iteration of these notes, it can – and should – be changed to reflect this and to tie it in even better with the previous section.

Example 3Z.1

Let $A = \begin{array}{c} \begin{array}{ccccc} \boldsymbol{a_1} & \boldsymbol{a_2} & \boldsymbol{a_3} & \boldsymbol{a_4} & \boldsymbol{a_5} \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 3 \\ -1 & 0 & -2 & 2 & 5 \\ 0 & 2 & -2 & -1 & 1 \end{pmatrix} \end{array}$ and let $\mathcal{B} = \{\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{a_4}\}$. Find $\mathcal{M}(A, \mathcal{B})$.

Observe that

$$\boldsymbol{a_3} = 2\boldsymbol{a_1} - \boldsymbol{a_2} + 0\boldsymbol{a_4}$$

and

$$\boldsymbol{a_5} = \boldsymbol{a_1} + 2\boldsymbol{a_2} + 3\boldsymbol{a_3}.$$

Then

$$\mathcal{M}(A,\mathcal{B}) = \begin{pmatrix} 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & -1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 3 \end{pmatrix}.$$

Definition: Reduced Row Echelon Form

Let A and $\mathcal{B}$ be as described above. The matrix $\mathcal{M}(A,\mathcal{B})$ is said to be the **(reduced row) echelon form** of A. The columns $j_1, \ldots, j_r$ in A are called **pivot columns**.

Some results fall out almost immediately.

Theorem 3Z.2

Let V and W be $\mathbb{K}$-vector spaces of dimension n, m, respectively. Let $T \in \mathcal{L}(V,W)$ and write $A = \mathcal{M}(T)$ (an $m \times n$ matrix). Finally, let $\mathcal{B}$ be the echelon basis for A.

1. $\operatorname{rank}(T)$ is the number of pivot columns in $\mathcal{M}(A,\mathcal{B})$, which is also the number of nonzero rows in $\mathcal{M}(A,\mathcal{B})$.
2. When $m = n$, T is invertible if and only if $\mathcal{M}(A,\mathcal{B}) = I$, the $n \times n$ identity matrix.

It turns out that there is a very procedural way to find the reduced row echlon

3Z.I Elementary Operations

Definition: elementary matrices

Let $\mathcal{B} = \{\boldsymbol{b_1}, \ldots, \boldsymbol{b_n}\}$ be a basis for $\mathbb{K}^n$. Consider the following bases $\mathcal{C}$ obtained by...

1. ...swapping $\boldsymbol{b_i}$ and $\boldsymbol{b_j}$.
$$\mathcal{C} = \{\boldsymbol{b_1}, \ldots, \boldsymbol{b_j}, \ldots, \boldsymbol{b_i}, \ldots, \boldsymbol{b_n}\}$$
2. ...scaling $\boldsymbol{b_i}$ by a nonzero scalar k.
$$\mathcal{C} = \{\boldsymbol{b_1}, \ldots, k\boldsymbol{b_i}, \ldots, \boldsymbol{b_n}\}$$
3. ...adding a multiple of $\boldsymbol{b_j}$ to $\boldsymbol{b_i}$.
$$\mathcal{C}\,\{\boldsymbol{b_1}, \ldots, \boldsymbol{b_i} + k\boldsymbol{b_j}, \ldots, \boldsymbol{b_n}\}$$

An **elementary matrix** is a matrix corresponding to one of the change-of-bases described above.

Example 3Z.3

Let $\boldsymbol{v} = \alpha_1\boldsymbol{b_1} + \cdots + \alpha_n\boldsymbol{b_n}$ and let E_1, E_2, E_3 denote the elemntary matrices from the previous definition. Observe the differences in $\mathcal{M}(\boldsymbol{v},\mathcal{B})$ and $\mathcal{M}(\boldsymbol{v},\mathcal{C})$ in each case.

1. Swapping $\boldsymbol{b_i}$ and $\boldsymbol{b_j}$.

$$\mathcal{M}(\boldsymbol{v},\mathcal{B}) = \left(\begin{array}{c} \alpha_1 \\ \vdots \\ \alpha_i \\ \vdots \\ \alpha_j \\ \vdots \\ \alpha_n \end{array}\right)\begin{array}{l} \\ \\ \text{row } i \\ \\ \text{row } j \\ \\ \\ \end{array} \qquad E_1\,\mathcal{M}(\boldsymbol{v},\mathcal{B}) = \mathcal{M}(\boldsymbol{v},\mathcal{C}) = \left(\begin{array}{c} \alpha_1 \\ \vdots \\ \alpha_j \\ \vdots \\ \alpha_i \\ \vdots \\ \alpha_n \end{array}\right)\begin{array}{l} \\ \\ \text{row } i \\ \\ \text{row } j \\ \\ \\ \end{array}$$

2. Scaling $\boldsymbol{b_i}$ by a nonzero scalar k.

$$\mathcal{M}(\boldsymbol{v},\mathcal{B}) = \left(\begin{array}{c} \alpha_1 \\ \vdots \\ \alpha_i \\ \vdots \\ \alpha_n \end{array}\right)\begin{array}{l} \\ \\ \text{row } i \\ \\ \\ \end{array} \qquad E_2\,\mathcal{M}(\boldsymbol{v},\mathcal{B}) = \mathcal{M}(\boldsymbol{v},\mathcal{C}) = \left(\begin{array}{c} \alpha_1 \\ \vdots \\ \frac{1}{k}\alpha_i \\ \vdots \\ \alpha_n \end{array}\right)\begin{array}{l} \\ \\ \text{row } i \\ \\ \\ \end{array}$$

3. Adding $k\boldsymbol{b_j}$ to $\boldsymbol{b_i}$.

$$\mathcal{M}(\boldsymbol{v},\mathcal{B}) = \left(\begin{array}{c} \alpha_1 \\ \vdots \\ \alpha_j \\ \vdots \\ \alpha_n \end{array}\right)\begin{array}{l} \\ \\ \text{row } j \\ \\ \\ \end{array} \qquad E_2\,\mathcal{M}(\boldsymbol{v},\mathcal{B}) = \mathcal{M}(\boldsymbol{v},\mathcal{C}) = \left(\begin{array}{c} \alpha_1 \\ \vdots \\ -k\alpha_i + \alpha_j \\ \vdots \\ \alpha_n \end{array}\right)\begin{array}{l} \\ \\ \text{row } j \\ \\ \\ \end{array}$$

We observe that multiplication by an elementary matrix amounts to doing one of the following things to the rows of a matrix: swapping two rows, scaling a row by a nonzero scalar, and adding a multiple of one row to another.

Definition: elementary row operation

An **elementary row operation** on an $m \times n$ matrix A is an operation on the rows of a matrix which is the result of multiplying A by an $m \times m$ elementary matrix E on the left (that is, the product EA).

Theorem 3Z.4

For any $m \times n$ matrix A, there is a sequence of elementary row operations taking A to its reduced row echelon form.

Proof. The proof is given by the **row reduction** algorithm below. □

Algorithm 3Z.5: Row Reduction

```
  /*ZEROING OUT LOWER-LEFT ENTRIES*/
$i \leftarrow 1$;   /*Row Number*/
$j \leftarrow 1$;   /*Column Number*/
while $i \leq m$ and $j \leq n$ do
    if Column $j$ contains nonzero entries then
        Use row swap to move nonzero entry to Row $i$;
        Use row addition to make entries below Row $i$ all zero;
        $i \leftarrow i + 1$;
        $j \leftarrow j + 1$;
    else
        $j \leftarrow j + 1$;
    end if
end while
  /*ZEROING OUT UPPER-RIGHT ENTRIES*/
$i \leftarrow m$;   /*Row Number*/
$j \leftarrow n$;   /*Column Number*/
while $i \geq 1$ and $j \geq 1$ do
    if Column $j$ contains a leading entry then
        Use row scaling to make leading entry 1;
        Use row addition to make entries above Row $i$ all zero;
        $i \leftarrow i - 1$;
        $j \leftarrow j - 1$;
    else
        $j \leftarrow j - 1$;
    end if
end while
```

Here is a visual of the Row Reduction Algorithm.

Step 1. Look at Rows $1 \ldots m$ in Column 1. If there are any nonzero entries, find it and move it to Row 1. Then use Row Addition to clear everything below it.

$$\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 3 & * & * & * & * \\ * & * & * & * & * \end{pmatrix} \longrightarrow \begin{pmatrix} 3 & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{pmatrix} \longrightarrow \begin{pmatrix} 3 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \end{pmatrix}$$

Step 2. Look at Rows $2 \ldots m$ in Column 2. There are no nonzero entries here, so we move onto the next column. We won't change the range of rows.

$$\begin{pmatrix} 3 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \end{pmatrix} \longrightarrow \begin{pmatrix} 3 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \end{pmatrix}$$

Steps 3 ... n. Repeat the above steps for all remaining columns. Now the matrix is in row echelon form.

$$\begin{pmatrix} 3 & * & * & * & * \\ 0 & 0 & 2 & * & * \\ 0 & 0 & 0 & 4 & * \\ 0 & 0 & 0 & 0 & 19 \end{pmatrix}$$

Step $n+1$. Look at Column n. If there is a leading entry, use Row Scaling to make that leading entry a 1, and then Row Addition to clear everything above it.

$$\begin{pmatrix} 3 & * & * & * & * \\ 0 & 0 & 2 & * & * \\ 0 & 0 & 0 & 4 & * \\ 0 & 0 & 0 & 0 & 19 \end{pmatrix} \longrightarrow \begin{pmatrix} 3 & * & * & * & * \\ 0 & 0 & 2 & * & * \\ 0 & 0 & 0 & 4 & * \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 3 & * & * & * & 0 \\ 0 & 0 & 2 & * & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Step $n+2$. Look at Column $n-1$. If there is a leading entry, use Row Scaling to make that leading entry a 1, and then Row Addition to clear everything above it.

$$\begin{pmatrix} 3 & * & * & * & 0 \\ 0 & 0 & 2 & * & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 3 & * & * & * & 0 \\ 0 & 0 & 2 & * & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 3 & * & * & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Steps $n+3$ … $2n$. Repeat the above steps for all remaining columns. Now the matrix is in reduced row echelon form.

$$\begin{pmatrix} 1 & * & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Remark. As a human, you can inject convenient steps into the algorithm above. You do not have to wait to scale until the end – you can clear fraction denominators/shrink large numbers at any point. You also have some choice when row swapping – if you can choose between multiple rows for a leading entry, pick one that already has a leading 1.

And now, I defer you to your 2000-level linear algebra course for all of the computational techniques you have previously learned.

Chapter 5

Eigenvalues, Eigenvectors, and Invariant Subspaces

5A Invariant Subspaces

Definition: linear operator

For any $\mathbb{K}$-vector space, V, a linear map $T \in \mathcal{L}(V)$ is called a **linear operator (on V)**.

Definition: invariant subspaces

Let V be a $\mathbb{K}$-vector space and $T \in \mathcal{L}(V)$. A subspace U of V is said to be in **invariant under $\boldsymbol{T}$** if $T(U) \subseteq U$ (that is, for every $\boldsymbol{u} \in U$, $T(\boldsymbol{u}) \in U$).

Example 5A.1: Trivial examples

For any vector space V and any operator $T \in \mathcal{L}(V)$, the subspaces $\{\mathbf{0}\}$ and V are an invariant subspaces of T.

Example 5A.2: Less-trivial examples

For any vector space V and any operator $T \in \mathcal{L}(V)$, the subspaces $\text{Null}(T)$ and $\text{Range}(T)$ are invariant subspaces.

For any $\boldsymbol{v} \in \text{Null}(T)$, $T(\boldsymbol{v}) = \mathbf{0} \in \text{Null}(T)$. As well, for any $\boldsymbol{v} \in \text{Range}(T)$, $T(\boldsymbol{v}) \in \text{Range}(T)$ (because T is a map from V to itself).

Example 5A.3: Slightly- interesting example

Let $T \in \mathcal{L}(\mathbb{K}^2)$ be given by $T(x, y) = (x + y, y)$. Show that the subspace $U = \{(x, 0) \in \mathbb{K}^2 : x \in \mathbb{K}\}$ is an invariant subspace.

Let $\boldsymbol{u} = (x, 0) \in U$. Then

$$T(\boldsymbol{u}) = T(x, 0) = (x + 0, 0) = (x, 0) \in U.$$

5A.I 1-Dimensional Invariant Subspaces

Fix some $\boldsymbol{v} \in V$ and consider the subspace $U = \operatorname{Span}(\boldsymbol{v}) = \{\lambda \boldsymbol{v} \ : \ \lambda \in \mathbb{K}\}$. If T is an operator on V and U is an invariant subspace, then it must be that

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

for some λ. This type of invariant subspace gets a special name.

Definition: eigenvectors, eigenvalues

Let V be a $\mathbb{K}$-vector space and $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbb{K}$ is called an **eigenvalue** if there is some nonzero vector $\boldsymbol{v}$ (called an **eigenvector**) for which $T(\boldsymbol{v}) = \lambda \boldsymbol{v}$.

Such a vector $\boldsymbol{v}$ is an element of $\operatorname{Null}(T - \lambda \operatorname{Id}_V)$, so we call this null space the **eigenspace** corresponding to λ.

Remark. The prefix *eigen–* is not a name, but is derived from German and means “own” as in the sense of characterizing an intrinsic property; a less literal translation would be along the lines of“special” or “characteristic.”

Remark. Eigenvectors are not unique: If $\boldsymbol{v}$ is an eigenvector, so is $k\boldsymbol{v}$ for any $k \in \mathbb{K}$:

$$T(k\boldsymbol{v}) = kT(\boldsymbol{v}) = k\lambda \boldsymbol{v} = \lambda(k\boldsymbol{v}).$$

Example 5A.4

Let $V = \mathbb{K}^2$ and T be the operator given by $T(x, y) = (x + y, y)$ (as in Example 5A.3). Find any eigenvalues/eigevectors for T.

Let $\boldsymbol{v} = (a, b)$. Then $\boldsymbol{v}$ is an eigenvector if and only if we can find some $\lambda \in \mathbb{K}$ for which

$$(a + b, b) = T(\boldsymbol{v}) = \lambda \boldsymbol{v} = \lambda(a, b).$$

This yields the following system of linear equations

$$\begin{cases} \lambda a &= a + b \\ \lambda b &= b \end{cases}$$

The second equation implies that $\lambda = 1$, and then the first equation implies that $b = 0$. So all eigenvectors of T are of the form $(x, 0)$, for some $x \in \mathbb{K}$.

Example 5A.5

Let $V = \mathbb{K}^2$ and let T be the operator given by $T(x, y) = (-y, x)$. Find any eigenvalues/eigenvectors for T when...

1. ...$\mathbb{K} = \mathbb{R}$
2. ...$\mathbb{K} = \mathbb{C}$

Let $\boldsymbol{v} = (a, b)$. Then $\boldsymbol{v}$ is an eigenvector if and only if we can find some $\lambda \in \mathbb{K}$ for which

$$(-b, a) = T(\boldsymbol{v}) = \lambda \boldsymbol{v} = \lambda(a, b).$$

This yields the following system of linear equations

$$\begin{cases} \lambda a &= -b \\ \lambda b &= a \end{cases}$$

Since x and y cannot both be zero, then in this case, *neither* can be zero. Substituting the second equation into the first, we have

$$\lambda(\lambda y) = -y$$

and so $\lambda^2 = -1$.

1. There are no real numbers whose squares are negative, so there are no eigenvalues and thusly no eigenvectors.
   This makes sense! A 1-dimensional invariant subspace of $\mathbb{R}^n$ is a subspace that is only stretched by some amount, no vector within in can change direction. A rotation, however, is a map that changes every vector's direction (unless that vector is $\mathbf{0}$ or the angle is an integer multiple of π).
2. There are two complex numbers that square to -1: namely $\pm i$. The corresponding invariant subspaces would then be $\operatorname{Span}((\pm i, 1))$, respectively.
   This also makes sense! Complex multiplication is weird and rotations can totally happen as a result of scaling by a complex number.

Theorem 5A.6

Let V be a $\mathbb{K}$-vector space and T an operator on V. Suppose that $\lambda_1, \cdots, \lambda_m$ are distinct eigenvalues with corresponding eigenvectors $\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}$. Then the set $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}\}$ is linearly independent.

Proof. We induct on m, taking our base case to be $m = 2$. Let k_1, k_2 be scalars for which

$$k_1\boldsymbol{v_1} + k_2\boldsymbol{v_2} = \mathbf{0}.$$

Applying $(T - \lambda_2 \operatorname{Id}_V)$ to both sides of this equation, one gets

$$\begin{aligned} k_1(\lambda_1 - \lambda_2)\boldsymbol{v_1} + k_2(\lambda_2 - \lambda_2)\boldsymbol{v_2} &= \mathbf{0} \\ \Longrightarrow \qquad k_1(\lambda_1 - \lambda_2)\boldsymbol{v_1} &= \mathbf{0} \end{aligned}$$

and since $\lambda_1 \neq \lambda_2$, it must be that $k_1 = 0$, whence $k_2 = 0$.

Suppose now that $\boldsymbol{v_1}, \ldots, \boldsymbol{v_{m-1}}$ are linearly independent and let $k_1, \ldots, k_m$ be scalars for which

$$k_1\boldsymbol{v_1} + \cdots + k_{m-1}\boldsymbol{v_{m-1}} + k_m\boldsymbol{v_m} = \mathbf{0}.$$

Applying $(T - \lambda_m \operatorname{Id}_V)$ to both sides of this equation, one gets

$$\begin{aligned} k_1(\lambda_1 - \lambda_m)\boldsymbol{v_1} + \cdots + k_{m-1}(\lambda_{m-1} - \lambda_m)\boldsymbol{v_{m-1}} + k_m(\lambda_m - \lambda_m)\boldsymbol{v_m} &= \mathbf{0} \\ \Longrightarrow \qquad k_1(\lambda_1 - \lambda_m)\boldsymbol{v_1} + \cdots + k_{m-1}(\lambda_{m-1} - \lambda_m)\boldsymbol{v_{m-1}} &= \mathbf{0}. \end{aligned}$$

Since $\boldsymbol{v_1}, \ldots, \boldsymbol{v_{m-1}}$ are linearly independent, we have that $k_i(\lambda_i - \lambda_m) = 0$ for all $i = 1, \ldots, m-1$. But since $\lambda_i - \lambda_m \neq 0$, then each $k_i = 0$. It follows then that $k_m = 0$ as well. □

If $\dim(V) = n$, then the cardinality of any set of linearly independent vectors is at most n, so Theorem 5A.6 yields

Proposition 5A.7

If $\dim(V) = n < \infty$, then any operator $T \in \mathcal{L}(V)$ has at most n distinct eigenvalues.

5A.II Polynomials Applied to Operators

Given that eigenvectors are nonzero, the equation $T(\boldsymbol{v}) = \lambda\boldsymbol{v}$ is equivalent to writing $\boldsymbol{v} \in \operatorname{Null}(T - \lambda \operatorname{Id}_V)$. Should we know the eigenvalue, finding the corresponding eigenspace (at least in the finite-dimensional case), is fairly procedural. The difficulty is finding these eigenvalues in the first place. Our goal is to come up with a systematic means of finding these eigenvalues and eigenvectors.

To do this, we introduce some new notation:

Notation: polynomial applied to an operator

Let V be a $\mathbb{K}$ vector space and $T \in \mathcal{V}$. For each positive integer m, we write

- T^m to mean $\underbrace{TT\cdots T}_{m \text{ times}}$
- T^0 to mean Id_V
- (and when T is invertible) T^{-m} to mean $\underbrace{T^{-1}T^{-1}\cdots T^{-1}}_{m \text{ times}}$.

For a polynomial $p \in \mathcal{P}(\mathbb{K})$

$$p(x) = k_0 + k_1 x + \cdots + k_m x^m$$

Then $p(T)$ is the operator in $\mathcal{L}(V)$ given by

$$p(T) = k_0 \operatorname{Id}_V + k_1 T + \cdots + k_m T^m$$

Example 5A.8

Suppose $D \in \mathcal{L}(\mathcal{P}(\mathbb{R}))$ is the derivative operator (c.f. Example 3A.3) given by $D(q) = \dfrac{dq}{dx}$. Let $p(x) = x^4 + 75x^3 - x + 1$. What is $p(D)(q)$?

$$p(D)(q) = D^4(q) + 75D^3(q) - D(q) + q = \frac{d^4q}{dx^2} + 75\frac{d^3q}{dx^3} - \frac{dq}{dx} + q$$

Recall that a product of polynomials p and q is not composition, but honest function multiplication:

$$(pq)(z) = p(z)q(z).$$

We obtain the following

Theorem 5A.9

Let V be a $\mathbb{K}$-vector space, $T \in \mathcal{L}(V)$, and $p, q \in \mathcal{P}(\mathbb{K})$. Then

1. $(pq)(T) = p(T)q(T) = q(T)p(T)$
2. $\text{Range}(p(T))$ is invariant under T
3. $\text{Null}(p(T))$ is invariant under T

The first item is completely straightforward to prove and is an exercise in bookkeeping, so we prove only the second two.

Proof. Let p be the polynomial

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n$$

We make the following observation: for any vector $\boldsymbol{u} \in V$,

$$\begin{aligned} T(p(T)(\boldsymbol{u})) &= T(a_0 \operatorname{Id}_V(\boldsymbol{u}) + a_1 T(\boldsymbol{u}) + \cdots + a_n T^n(\boldsymbol{u})) \\ &= a_0 T(\boldsymbol{u}) + a_1 T^2(\boldsymbol{u}) + \cdots + a_n T^{n+1}(\boldsymbol{u}) \\ &= a_0 \operatorname{Id}_V(T(\boldsymbol{u})) + a_1 T(T(\boldsymbol{u})) + \cdots + a_n T^n(T(\boldsymbol{u})) \\ &= p(T)(T(\boldsymbol{u})) \end{aligned}$$

2. We aim to show that $T(\text{Range}(p(T))) \subseteq \text{Range}(p(T))$.
   So, suppose that $\boldsymbol{v} \in \text{Range}(p(T))$. Then there exists $\boldsymbol{u} \in V$ so that $p(T)(\boldsymbol{u}) = \boldsymbol{v}$. Hence by our observation above,

   $$T(\boldsymbol{v}) = T(p(T)(\boldsymbol{u})) = p(T)(T(\boldsymbol{u}))$$

   and therefore $T(\boldsymbol{v}) \in \text{Range}(p(T))$.
3. We aim to show that $T(\text{Null}(p(T))) \subseteq \text{Null}(p(T))$.
   So, suppose that $\boldsymbol{u} \in \text{Null}(p(T))$. Then, following from our above observation, we have

   $$p(T)(T(\boldsymbol{v})) = T(p(T)(\boldsymbol{v})) = T(\mathbf{0}) = \mathbf{0}$$

   and therefore $T(\boldsymbol{v}) \in \text{Null}(p(T))$.

□

5B The Minimal Polynomial

Recall the following fact:

Theorem 5B.1: Fundamental Theorem of Algebra

Every non-constant polynomial with coefficients in $\mathbb{C}$ has at least one root in $\mathbb{C}$.

We will use this to prove the following major result:

Theorem 5B.2

Let V be a finite-dimensional $\mathbb{C}$-vector space. Then every operator $T \in \mathcal{L}(V)$ has at least one complex eigenvalue.

Proof. Let V be a $\mathbb{C}$-vector space with $\dim(V) = n < \infty$. Let $T \in \mathcal{L}(V)$ be any operator and $\boldsymbol{v} \in V$ a fixed nonzero vector. The set of vectors

$$\{\boldsymbol{v}, T(\boldsymbol{v}), \ldots, T^n(\boldsymbol{v})\}$$

cannot be linearly independent (it has $n+1$ vectors). Without loss of generality, we'll suppose that no proper subset of these vectors is linearly dependent (so in particular, $\{\boldsymbol{v}, T(\boldsymbol{v}), \ldots, T^{n-1}(\boldsymbol{v})\}$ is linearly independent). There are scalars $\alpha_0, \ldots, \alpha_n$ – not all zero – for which

$$\alpha_0 \boldsymbol{v} + \alpha_1 T(\boldsymbol{v}) + \cdots + \alpha_n T^n(\boldsymbol{v}) = \mathbf{0} \tag{5.1}$$

Let $p(x)$ be the polynomial in $\mathcal{P}_n(\mathbb{C})$

$$p(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n,$$

so that Equation 5.1 can be written as $p(T)(\boldsymbol{v}) = \mathbf{0}$. Without loss of generality, we'll assume that $p(x)$ is the smallest-degree polynomial for which this is true.

Now, following the Fundamental Theorem of Algebra, there is a complex number λ so that $p(\lambda) = 0$. We can thus find a polynomial q of degree $n-1$ for which

$$p(x) = (x - \lambda)q(x).$$

We now have that

$$p(T)(\boldsymbol{v}) = (T - \lambda \operatorname{Id}_V)q(T)\boldsymbol{v} = \mathbf{0}.$$

Since $q(T)$ is a linear combination of $\{\boldsymbol{v}, \ldots, T^{n-1}(\boldsymbol{v})\}$, which was assumed to be linearly independent, it must be that $(T - \lambda \operatorname{Id}_V)(\boldsymbol{v}) = \mathbf{0}$. Therefore

$$(T - \lambda \operatorname{Id}_V)(\boldsymbol{v}) = \mathbf{0} \quad \Longrightarrow \quad T(\boldsymbol{v}) - \lambda \operatorname{Id}_V(\boldsymbol{v}) = \mathbf{0} \quad \Longrightarrow \quad T(\boldsymbol{v}) = \lambda \boldsymbol{v}$$

and therefore $\boldsymbol{v}$ is an eigenvector for T corresponding to the eigenvalue λ. □

Definition: monic polynomial

A polynomial $a_nx^n + \cdots + a_1x + a_0$ is **monic** if $a_n = 1$.

Theorem 5B.3: Existence and Uniqueness of minimal polynomial

Suppose V is a finite-dimensional $\mathbb{K}$-vector space and $T \in \mathcal{L}(V)$. Then there is a unique monic polynomial $p \in \mathcal{P}(\mathbb{K})$ of smallest degree so that $p(T) = 0$.

Definition: minimal polynomial

The polynomial in the previous theorem is called the **minimal polynomial** of T.

Proof of Theorem 5B.3. Let $T \in \mathcal{L}(V)$

Monic. If $a_nx^n + \cdots + a_1x + a_0$ is any polynomial in $\mathcal{P}(\mathbb{K})$, then

$$x^n + \frac{a_{n-1}}{a_n}x^{n-1} + \cdots + \frac{a_1}{a_n}x + \frac{a_0}{a_n}$$

is contained in $\mathcal{P}(\mathbb{K})$ also. As such, we can always consider any candidate minimal polynomial to be monic.

Uniqueness. Suppose that p_1 is the minimal polynomial for T and p_2 is another polynomial for which $p_2(T) = 0$. By the division algorithm for polynomials, then there are polynomials q and r so that

1. $p_2(x) = q(x)p_1(x) + r(x)$, and
2. $0 \leq \deg(r) < \deg(p_1)$

Item 1 yields

$$0 = p_2(T) = q(T)p_1(T) + r(T) = 0 + r(T)$$

so must have that $r(T) = 0$. Since $p_1(x)$ is assumed to be of minimal degree, then Item 2 implies that $r(x) = 0$. It follows that p_1 divides p_2. If p_2 is also assumed to be minimal, then it is necessarily monic, hence $q(x) = 1$ and $p_1 = p_2$.

□

As a corollary of the uniqueness proof above,

Corollary 5B.4

A polynomial q is a polynomial multiple of the minimal polynomial if and only if $q(T) = 0$.

Let $\dim(V) = d$. Since $\mathcal{L}(V)$ has dimension d^2 and $\left\{\mathrm{Id}, T, T^2, \ldots, T^{d^2}\right\}$ is a set of $d^2 + 1$ vectors in $\mathcal{L}(V)$, then clearly it must be a linearly dependent set, and thus the degree of the minimal polynomial is less than d^2. Now observe that, for any $\boldsymbol{v} \in V$, the set $\left\{\mathrm{Id}(\boldsymbol{v}), T(\boldsymbol{v}), \ldots, T^d(\boldsymbol{v})\right\}$ is a set of $d+1$ vectors in V, and thus is clearly linearly dependent. It raises the question about whether the bound for the degree of the minimal polynomial can be lowered from d^2 to d. Indeed,

Theorem 5B.5

Let V be a finite-dimensional vector space, $T \in \mathcal{L}(V)$, and p the minimal polynomial for T. Then $\deg(p) \leq \dim(V)$.

Your book gives an elementary, inductive proof of this fact, but we'll wait until much later and appeal to the Cayley–Hamilton Theorem.

Theorem 5B.6

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then the zeroes of the minimal polynomial are precisely the eigenvalues of T.

Recall that, in the realm of logical statements, "precisely" means a biconditional statement.

Proof. Let p be the minimal polynomial for T.

(⇒) Suppose that λ is a root of p. Then there is a polynomial, q (of strictly smaller degree) for which

$$p(x) = (x - \lambda)q(x).$$

As p is the minimal polynomial of T, we must have that $p(T) = 0$, $q(T) \neq 0$. Let $\boldsymbol{v} \in V$ satisfying $q(T)\boldsymbol{v} \neq 0$. Then

$$\mathbf{0} = 0\boldsymbol{v} = p(T)\boldsymbol{v} = (T\boldsymbol{v} - \lambda \operatorname{Id} \boldsymbol{v})q(T)\boldsymbol{v}$$

from which it follows that $T\boldsymbol{v} = \lambda \operatorname{Id} \boldsymbol{v}$, and thus λ is an eigenvalue of T.

(⇒) Suppose now that λ is an eigenvalue for T, and $\boldsymbol{v}$ a corresponding eigenvector. Observe that, for every nonnegative integer m and every scalar k,

$$T^m(k\boldsymbol{v}) = k\lambda^m\boldsymbol{v} = kT^m(\boldsymbol{v}).$$

We thus have that, for any polynomial q,

$$q(T)\boldsymbol{v} = q(\lambda)\boldsymbol{v}.$$

In particular, when taking the minimal polynomial p,

$$\mathbf{0} = 0\boldsymbol{v} = p(T)\boldsymbol{v} = p(\lambda)\boldsymbol{v}$$

and thus $p(\lambda) = 0$, so λ is a root of p.

□

Finding the Minimal Polynomial

1. Pick a basis and write the operator as a matrix, A say, in that basis.
2. If A is the zero matrix, the minimal polynomial is just $p(x) = 1$. If A is not the zero matrix, then...
3. For each m with $1 \leq m \leq \dim(V)$, try to solve the linear system $x_0 I + x_1 A + \cdots + x_{m-1}A^{m-1} + A^m = 0$.
4. For the first m where you find a solution, use the solution x_i-values as your polynomial coefficients.

Example 5B.7

Find the minimal polynomial for $T \in \mathcal{L}(\mathbb{R}^2)$ given by $T(x, y) = (1x + 2y,\, 3x + 4y)$.

In the standard basis, $A := \mathcal{T} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

A is not the zero matrix, so we try to solve

$$x_1 I + A = 0,$$

but this has no solution because A is not a scalar multiple of I. So we try to solve

$$x_1 I + x_2 A + A^2 = 0 \qquad \Longrightarrow \qquad x_1 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + x_2 \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 7 & 10 \\ 15 & 22 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

This produces the system

$$\begin{cases} x_1 + x_2 = -7 \\ \qquad 2x_2 = -10 \\ \qquad 3x_2 = -15 \\ x_1 + 4x_2 = -22 \end{cases}$$

which is readily seen to have the solution $x_2 = -5$ and $x_1 = -2$. Thus the minimal polynomial is

$$p(x) = -2 - 5x + x^2.$$

When a monic polynomial has degree less than $\dim(V)$, it signals repeated eigenvalues. However, "most" (which has a precise, qualitative meaning) matrices have all distinct eigenvalues and thus have a minimal polynomial of maximum degree. As such, there is a slightly-faster method that one can employ in general and which has a high success rate. (Axler claims that it works on more than 99.999% of 4×4 matrices $A = (A_{i,j})$ with integer entries where $-100 \leq A_{i,j} \leq 100$).

Finding the Minimal Polynomial – Faster Method (but not guaranteed)

Let $\dim(V) = m$.

1. Pick a basis and write the operator as a matrix, A say, in that basis. Fix any nonzero vector, $\boldsymbol{v}$ (the standard basis vectors are usually easiest).
2. If A is the zero matrix, the minimal polynomial is just $p(x) = 1$. If A is not the zero matrix, then...
3. Try to solve the linear system $x_0\boldsymbol{v} + x_1 A\boldsymbol{v} + \cdots + x_{m-1}A^{m-1}\boldsymbol{v} + A^m\boldsymbol{v} = 0$.
4. If the solution is unique, use the solution x_i-values as your polynomial coefficients.

Example 5B.8

Using the fast technique with the standard basis vector $\boldsymbol{v} = \boldsymbol{e_1}$, find the minimal polynomial for $T \in \mathcal{L}(\mathbb{K}^4)$ where

$$\mathcal{M}(T) = \begin{pmatrix} 0 & -1 & -1 & 2 \\ 1 & 2 & -2 & -1 \\ 2 & 2 & 0 & 1 \\ 0 & -2 & 0 & 1 \end{pmatrix}$$

Let $A := \mathcal{M}(T)$. Then we see that

$$A\boldsymbol{e_1} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix}, \qquad A^3\boldsymbol{e_1} = A\begin{pmatrix} -3 \\ -2 \\ 2 \\ -2 \end{pmatrix} = \begin{pmatrix} -4 \\ -9 \\ -12 \\ 2 \end{pmatrix},$$

$$A^2\boldsymbol{e_1} = A\begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -3 \\ -2 \\ 2 \\ -2 \end{pmatrix}, \qquad A^4\boldsymbol{e_1} = A\begin{pmatrix} -4 \\ -9 \\ -12 \\ 2 \end{pmatrix} = \begin{pmatrix} 25 \\ 0 \\ -25 \\ 20 \end{pmatrix}.$$

So we try to solve the system

$$x_1\boldsymbol{e_1} + x_2A\boldsymbol{e_1} + x_3A^2\boldsymbol{e_1} + x_4A^3\boldsymbol{e_1} + A^4\boldsymbol{e_1} = \mathbf{0}$$

$$\longrightarrow \begin{cases} x_1 \qquad\quad - 3x_3 - \;\; 4x_4 = -25 \\ \qquad\; x_2 - 2x_3 - \;\; 9x_4 = 0 \\ \quad\; 2x_2 + 2x_3 - 12x_4 = 24 \\ \qquad\quad\;\; - 2x_3 + \;\; 2x_4 = -20 \end{cases}$$

Using Gaussian elimination,

$$\left(\begin{array}{cccc|c} 1 & 0 & -3 & -4 & 25 \\ 0 & 1 & -2 & -9 & 0 \\ 0 & 2 & 2 & -12 & -24 \\ 0 & 0 & -2 & 2 & 20 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 16 \\ 0 & 1 & 0 & 0 & 13 \\ 0 & 0 & 1 & 0 & -7 \\ 0 & 0 & 0 & 1 & 3 \end{array}\right)$$

We see that the minimal polynomial is

$$p(x) = 16 + 13x - 7x^2 + 3x^3 + x^4.$$

Generally-speaking, the choice of $\boldsymbol{v}$ is also very important and the technique will fail if $\boldsymbol{v}$ is an eigenvector.

Example 5B.9

Let $T \in \mathcal{L}(\mathbb{R}^4)$ where

$$\mathcal{T} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & -1 \\ 0 & 2 & -2 & -1 \\ 0 & -2 & -1 & 2 \end{pmatrix}.$$

Show that the fast technique fails to find the minimal polynomial with each of the following choices of $\boldsymbol{v}$:

1. $\boldsymbol{v} = (1, 0, 0, 0)$
2. $\boldsymbol{v} = (3, -2, -2, 6)$
3. $\boldsymbol{v} = (-7, 3, 18, 6)$
4. $\boldsymbol{v} = (-1, 2, 2, 2)$

5D Diagonalizable Operators

Definition: diagonal, diagonal matrix

An $n \times n$ matrix $A = (A_{i,j})$ is said to be **diagonal** if $A_{i,j} = 0$ whenever $i \neq j$. The matrix entries $\{A_{1,1}, A_{2,2}, \ldots, A_{n,n}\}$ are called the **diagonal** of A.

Remark. A diagonal matrix is defined by having 0 for non-diagonal entries. It is entirely possible that 0's can occur along the diagonal.

Definition: diagonalizable matrix

An operator $T \in \mathcal{L}(V)$ is called **diagonalizable** if there exists a basis $\mathcal{B}$ for which $\mathcal{M}(T, \mathcal{B})$ is a diagonal matrix.

Example 5D.1

Let $T \in \mathcal{L}(\mathbb{R}^2)$ be the operator given by

$$T(x, y) = (-x + 3y, 3x - y)$$

Find $\mathcal{M}(T)$ in both the standard basis, $\mathcal{E} = \{\boldsymbol{e_1} = (1, 0), \boldsymbol{e_2} = (0, 1)\}$, and the basis $\mathcal{B} = \{\boldsymbol{b_1} = (1, 1), \boldsymbol{b_2} = (1, -1)\}$.

We have that

$$T(\boldsymbol{e_1}) = (-1, 3) = -1\boldsymbol{e_1} + 3\boldsymbol{e_2} \qquad \text{and} \qquad T(\boldsymbol{e_2}) = (3, -1) = 3\boldsymbol{e_1} - 1\boldsymbol{e_2}$$

so

$$\mathcal{M}(T, \mathcal{E}) = \begin{pmatrix} -1 & 3 \\ 3 & -1 \end{pmatrix}.$$

We also have that

$$T(\boldsymbol{b_1}) = (2, 2) = 2\boldsymbol{b_1} + 0\boldsymbol{b_2} \qquad \text{and} \qquad T(\boldsymbol{b_2}) = (-4, 4) = 0\boldsymbol{b_1} - 4\boldsymbol{b_2}$$

and thus

$$\mathcal{M}(T, \mathcal{B}) = \begin{pmatrix} 2 & 0 \\ 0 & -4 \end{pmatrix}.$$

Therefore T is diagonalizable.

Notation: Eigenspace – $E(\lambda, T)$

Recall that the **eigenspace** of an operator T, corresponding to the eigenvalue λ, is $\text{Null}(T - \lambda I_d)$. We'll write $\boldsymbol{E(\lambda, T)}$ to denote this eigenspace.

Theorem 5D.2: sum of eigenspaces

Let V be a $\mathbb{K}$-vector space and $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T. Then the sum

$$E(\lambda_1, T) + \cdots + E(\lambda_m, T)$$

is actually a direct sum. Moreover,

$$\dim(E(\lambda_1, T)) + \cdots + \dim(E(\lambda_m, T)) \leq \dim(V).$$

Proof. Because eigenvectors corresponding to different eigenvalues are linearly independent, it follows immediately that the sum is a direct sum (we also proved this on the homework in the context of spans). Moreover, letting B_i be basis for each $E(\lambda_i, T)$, one has that $\bigcup_{i=1}^m B_i$ is a linearly independent set of vectors, and

$$\dim\left(\bigoplus_{i=1}^m E(\lambda_1, T)\right) = \sum_{i=1}^m \dim(E(\lambda_i, T)) = \text{cardinality}\left(\bigcup_{i=1}^m B_i\right) \leq \dim(V).$$

□

Theorem 5D.3: Diagonalization

Let V be a finite-dimensional $\mathbb{K}$-vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. The following are equivalent.

1. T is diagonalizable.
2. V has a basis of eigenvectors.
3. $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$
4. $\dim(V) = \dim(E(\lambda_1, T)) + \cdots + \dim(E(\lambda_m, T))$

Proof. Let $n = \dim(V)$.

(1 ⇔ 2) $\mathcal{M}(T)$ is a diagonal matrix if and only if there is some basis $\{\boldsymbol{b_1}, \ldots, \boldsymbol{b_n}\}$ for which

$$T(\boldsymbol{b_i}) = 0\boldsymbol{b_1} + \cdots + 0\boldsymbol{b_{i-1}} + \lambda_i\boldsymbol{b_i} + 0\boldsymbol{b_{i+1}} + \cdots + 0\boldsymbol{b_n} = \lambda_i\boldsymbol{b_i};$$

in other words, if and only if V has a basis of eigenvectors.

(2 ⇒ 3) Suppose V has a basis of eigenvectors. Then every vector in V is a linear combination of eigenvectors, hence

$$V = E(\lambda_1, T) + \cdots + E(\lambda_m, T),$$

and Theorem 5D.2 implies this is a direct sum.

(3 ⇒ 4) This follows immediately from Corollary 2C.8.

(3 ⇒ 2) Suppose $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$. In each $E(\lambda_i, T)$, there is a unique eigenvector $\boldsymbol{v_i}$ for which

$$\boldsymbol{v_1} + \cdots + \boldsymbol{v_m} = \boldsymbol{0}.$$

Now let B_i be a basis for each eigenspace $E(\lambda_i, T)$. Then each $\boldsymbol{v_i}$ is a unique linear combination of the basis vectors B_i, and thus the collection of all basis vectors $\bigcup_{i=1}^{m} B_i$ is a linearly independent set. That V equals this direct sum implies that this set spans V, and hence is a basis.

(4 ⇒ 2) Suppose $\dim(V) = \dim(E(\lambda_1, T)) + \cdots + \dim(E(\lambda_m, T))$. The linearly independent set $\bigcup_{i=1}^{m} B_i$ constructed in the previous part contains n-many vectors, which means it is a basis for an n-dimensional vector space.

□

Corollary 5D.4

If $\dim(V) = n$ and $T \in \mathcal{L}(V)$ has n distinct eigenvalues, then T is diagonalizable.

Example 5D.5

Let $T \in \mathcal{L}(\mathbb{Q}^3)$ be given by

$$T(x, y, z) = (3x + z, -x + 3y - z, 3z).$$

Is T diagonalizable?

We first look for eigenvalues by way of the minimal polynomial. In the standard basis, $A = \mathcal{M}(T) = \begin{pmatrix} 3 & 0 & 1 \\ -1 & 3 & -1 \\ 0 & 0 & 3 \end{pmatrix}$, so

$$k_1\boldsymbol{e_3} + k_2A\boldsymbol{e_3} + k_3A^2\boldsymbol{e_3} = -A^3\boldsymbol{e_3}$$

$$k_1\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + k_2\begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix} + k_3\begin{pmatrix} 6 \\ -7 \\ 9 \end{pmatrix} = \begin{pmatrix} -27 \\ 36 \\ -27 \end{pmatrix}$$

is a system which can be solved with Gaussian elimination:

$$\left(\begin{array}{ccc|c} 0 & 1 & 6 & -27 \\ 0 & -1 & -7 & 36 \\ 1 & 3 & 9 & -27 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & -27 \\ 0 & 1 & 0 & 27 \\ 0 & 0 & 1 & -9 \end{array}\right)$$

and thus the minimal polynomial is

$$-27 + 27x - 9x^2 + x^3 = (x - 3)^3$$

and our lone eigenvalue is 3.

We only need to count the number of linearly independent eigenvectors associated with $\lambda = 3$. This is quickly achieved by considering the nullity of $T - 3\,\mathrm{Id}$ – T is diagonalizable if and only if $T - 3\,\mathrm{Id}$ has nullity 3 (or equivalently rank 0). Since

$$(T - 3\,\mathrm{Id})(x, y, z) = (z, -x - z, 0)$$

then $\operatorname{rank}(T - 3\,\mathrm{Id}) > 0$ and therefore T is not diagonalizable.

The minimal polynomial degree is roughly inversely proportional to the number of linearly independent eigenvectors decreases (it's not a one-to-one correspondence, however). All of the matrices in the following examples (Examples 5D.6 to 5D.10) have only the eigenvalue 7, and only A_1 is diagonalizable.

Example 5D.6

Find the eigenvectors and minimal polynomial for $A_1 = \begin{pmatrix} 7 & & & \\ & 7 & & \\ & & 7 & \\ & & & 7 \end{pmatrix}$.

A_1 has eigenvectors $\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}, \boldsymbol{e_4}$. Also,

$$A_1 - 7I = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{pmatrix},$$

and thus the minimal polynomial is $p_1(x) = (x - 7)$.

Example 5D.7

Find the eigenvectors and minimal polynomial for $A_2 = \begin{pmatrix} 7 & 1 & & \\ & 7 & 0 & \\ & & 7 & 0 \\ & & & 7 \end{pmatrix}$.

A_2 has eigenvectors $\boldsymbol{e_1}, \boldsymbol{e_3}, \boldsymbol{e_4}$. Also,

$$A_2 - 7I = \begin{pmatrix} 0 & 1 & & \\ & 0 & 0 & \\ & & 0 & 0 \\ & & & 0 \end{pmatrix}, \qquad (A_2 - 7I)^2 = \begin{pmatrix} 0 & 0 & & \\ & 0 & 0 & \\ & & 0 & 0 \\ & & & 0 \end{pmatrix},$$

and thus the minimal polynomial is $p_2(x) = (x - 7)^2$.

Example 5D.8

Find the eigenvectors and minimal polynomial for $A_3 = \begin{pmatrix} 7 & 1 & & \\ & 7 & 0 & \\ & & 7 & 1 \\ & & & 7 \end{pmatrix}$.

A_3 has eigenvectors has eigenvectors $\boldsymbol{e_1}, \boldsymbol{e_3}$. Also,

$$A_3 - 7I = \begin{pmatrix} 0 & 1 & & \\ & 0 & 0 & \\ & & 0 & 1 \\ & & & 0 \end{pmatrix}, \qquad (A_3 - 7I)^2 = \begin{pmatrix} 0 & 0 & & \\ & 0 & 0 & \\ & & 0 & 0 \\ & & & 0 \end{pmatrix},$$

and thus the minimal polynomial is $p_3(x) = (x-7)^2$.

Example 5D.9

Find the eigenvectors and minimal polynomial for $A_4 = \begin{pmatrix} 7 & 1 & & \\ & 7 & 1 & \\ & & 7 & 0 \\ & & & 7 \end{pmatrix}$.

A_4 has eigenvectors $\boldsymbol{e_1}, \boldsymbol{e_4}$. Also

$$A_4 - 7I = \begin{pmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & 0 \\ & & & 0 \end{pmatrix}, \qquad (A_4 - 7I)^2 = \begin{pmatrix} 0 & 0 & 1 & \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & & & 0 \end{pmatrix}, \qquad (A_4 - 7I)^2 = \begin{pmatrix} 0 & 0 & 0 & \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & & & 0 \end{pmatrix},$$

and thus the minimal polynomial is $p_4(x) = (x-7)^3$.

Example 5D.10

Find the eigenvectors and minimal polynomial for $A_5 = \begin{pmatrix} 7 & 1 & & \\ & 7 & 1 & \\ & & 7 & 1 \\ & & & 7 \end{pmatrix}$.

A_5 has eigenvector $\boldsymbol{e_1}$. Also,

$$A_5 - 7I = \begin{pmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & 1 \\ & & & 0 \end{pmatrix}, \qquad (A_5 - 7I)^2 = \begin{pmatrix} 0 & 0 & 1 & \\ & 0 & 0 & 1 \\ & & 0 & 0 \\ & & & 0 \end{pmatrix},$$

$$(A_5 - 7I)^3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & & & 0 \end{pmatrix}, \qquad (A_5 - 7I)^4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & & & 0 \end{pmatrix}$$

and thus the minimal polynomial is $p_5(x) = (x-7)^4$.

Theorem 5D.11: Diagonalizability and Minimal Polynomial

An operator T is diagonalizable if and only if its minimal polynomial is $p(x) = (x - \lambda_1) \cdots (x - \lambda_m)$ where the λ_i's are distinct eigenvalues.

Proof. Let $T \in \mathcal{L}(V)$ be an operator with minimal polynomial p.

(⇒) Suppose that T is diagonalizable. Then V has a basis of eigenvectors, $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}\}$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ (not necessarily distinct). Up to relabeling, suppose that there is some $m \leq n$ for which $\lambda_1, \ldots, \lambda_m$ are distinct. Certainly one then has that

$$(T - \lambda_m \operatorname{Id})\boldsymbol{v_m} = 0$$

and since $\lambda_{m+1}, \ldots, \lambda_n$ are repeats of eigenvalues in $\lambda_1, \ldots, \lambda_m$, it must be that for every $j \in \{m+1, \ldots, n\}$ there is some $k \in \{1, \ldots, m\}$ for which $(T - \lambda_k \operatorname{Id})\boldsymbol{v_j} = 0$. Therefore

$$(T - \lambda_1 \operatorname{Id}) \cdots (T - \lambda_m \operatorname{Id})\boldsymbol{v_i} = \mathbf{0}$$

for every basis eigenvector $\boldsymbol{v_i}$ and the minimal polynomial is $(x - \lambda_1) \cdots (x - \lambda_m)$.

(⇐) [1] Suppose now that we can write $p(x) = (x - \lambda_1) \cdots (x - \lambda_m)$ where the λ_i's are distinct eigenvalues. By Theorem 5D.3, T is diagonalizable if and only if

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$$

The sum on the right is always a subset of V, so we only need to show that every $\boldsymbol{u} \in V$ can be written as a sum

$$\boldsymbol{u} = \boldsymbol{v_1} + \cdots + \boldsymbol{v_m} \tag{5.2}$$

where $\boldsymbol{v_j} \in E(\lambda_j, T) = \operatorname{Null}(T - \lambda_j \operatorname{Id})$. Looking at the partial fraction decomposition of $\frac{1}{p(x)}$, there are scalars $k_1, \ldots, k_m$ for which

$$\frac{1}{p(x)} = \frac{1}{(x - \lambda_1) \cdots (x - \lambda_m)} = \sum_{j=1}^{m} \frac{k_j}{x - \lambda_j}.$$

For each j, define the polynomial

$$q_j(x) = \frac{k_j\, p(x)}{x - \lambda_j}.$$

This can be rearranged to $(x - \lambda_j)q_j(x) = k_j p(x)$, so we see that, for every $\boldsymbol{u}$,

$$(T - \lambda_j \operatorname{Id})q_j(T)(\boldsymbol{u}) = k_j p(T)(\boldsymbol{u}) = 0\boldsymbol{u} = \mathbf{0}.$$

It follows that $q_j(T)(\boldsymbol{u}) \in \operatorname{Null}(T - \lambda_j \operatorname{Id}) = E(\lambda_j, T)$ and therefore we can take each $\boldsymbol{v_j} = q_j(T)(\boldsymbol{u})$ in Equation 5.2.

□

---

[1] Credit for this slick proof goes to Math.StackExchange user Faust.
`https://math.stackexchange.com/questions/2676557/prove-that-t-is-diagonalizable-if-and-only-if-the-minimal-polynomial-of-t-has-no`

5D.I Utility of Diagonal Operators

Observe that, if λ is an eigenvalue for an operator T and $\boldsymbol{v}$ is a corresponding eigenvector, then $T^k\boldsymbol{v} = \lambda^k\boldsymbol{v}$, and then linearity yields

$$T^k(\alpha_1\boldsymbol{v_1} + \cdots + \alpha_m\boldsymbol{v_m}) = \alpha_1\lambda_1^k\boldsymbol{v_1} + \cdots + \alpha_m\lambda_m^k\boldsymbol{v_m}.$$

Although outright computing T^k for large values of k can be computationally expensive[2], but computing images of specific vectors can be very fast if said vectors are linear combinations of eigenvectors.

Example 5D.12

The operator $T \in \mathcal{M}(\mathbb{R}^3)$ given by

$$T(x, y, z) = (3x, -3x + 2y - 3z, 2x + 5z)$$

is diagonalizable and has eigenvalues $2, 3, 5$. Find

$$T^{3141592}(-\pi^2, -\pi + \pi^3, \pi + \pi^2).$$

In the standard basis, $A := \mathcal{M}(T) = \begin{pmatrix} 3 & 0 & 0 \\ -3 & 2 & -3 \\ 2 & 0 & 5 \end{pmatrix}$, and the eigenvectors are quickly achieved with Gaussian elimination.

$$\big(A - 2I \mid \mathbf{0}\big) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad \text{and thus} \quad (0, 1, 0) \in \operatorname{Null}(T - 2\operatorname{Id}) = E(2, T)$$

$$\big(A - 3I \mid \mathbf{0}\big) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad \text{and thus} \quad (-1, 0, 1) \in \operatorname{Null}(T - 3\operatorname{Id}) = E(3, T)$$

$$\big(A - 5I \mid \mathbf{0}\big) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad \text{and thus} \quad (0, -1, 1) \in \operatorname{Null}(T - 5\operatorname{Id}) = E(5, T)$$

We see that

$$(-\pi^2, -\pi + \pi^3, \pi + \pi^2) = \pi^3(0, 1, 0) + \pi^2(-1, 0, 1) + \pi(0, -1, 1)$$

from which it follows that

$$\begin{aligned} &T^{3141592}(-\pi^2, -\pi + \pi^3, \pi + \pi^2) \\ &= T^{3141592}\Big(\pi^3(0, 1, 0) + \pi^2(-1, 0, 1) + \pi(0, -1, 1)\Big) \\ &= 2^{3141592} \cdot \pi^3(0, 1, 0) + 3^{3141592} \cdot \pi^2(-1, 0, 1) + 5^{3141592} \cdot \pi(0, -1, 1) \end{aligned}$$

---

[2] Directly applying the definition of the product of two $n \times n$ matrices is an $O(n^3)$-time algorithm. The best-known improvement on this result is from 2022, which provides a faster $O(n^{2.37188})$-time algorithm. However, this improved algorithm is known as a “galactic” algorithm, because the constants involved in it are so large that the algorithm is totally impractical.

Chapter 9

Multilinear Algebra and Determinants

9A Bilinear Forms and Quadratic Forms

9A.I Bilinear Forms

Definition: bilinear form

A **bilinear form** on V (a $\mathbb{K}$-vector space) is a function $\beta : V \times V \to \mathbb{K}$ so that, for every $\boldsymbol{v}$, the functions given by

$$\beta_1(\boldsymbol{x}) = \beta(\boldsymbol{x}, \boldsymbol{v}) \qquad \text{and} \qquad \beta_2(\boldsymbol{x}) = \beta(\boldsymbol{v}, \boldsymbol{x})$$

are linear functionals $\beta_1, \beta_2 : V \to \mathbb{K}$.

Remark. Bilinear forms generalize the idea of a dot product.

Example 9A.1: The canonical example

Show that the usual dot product on $\mathbb{R}^n$ is a bilinear form.

Although we haven't formally covered the dot product in this class, we'll assume that the reader is familiar with it and has seen the following properties: For all scalars k and all vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$:

$$(k\boldsymbol{u}) \cdot \boldsymbol{v} = k(\boldsymbol{u} \cdot \boldsymbol{v})$$
$$(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = (\boldsymbol{u} \cdot \boldsymbol{w}) + (\boldsymbol{v} \cdot \boldsymbol{w})$$

and

$$\boldsymbol{u} \cdot (k\boldsymbol{v}) = k(\boldsymbol{u} \cdot \boldsymbol{v})$$
$$\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} \cdot \boldsymbol{v}) + (\boldsymbol{u} \cdot \boldsymbol{w})$$

The bilinearity is immediate.

Example 9A.2

Let $A_{i,j}$ be scalars (where $i = 1, \ldots, n$ and $j = 1, \ldots, n$). Show that

$$\beta : \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}$$
$$\beta(\boldsymbol{u}, \boldsymbol{v}) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j} u_i v_j$$

is a bilinear form.

By staring at the above long enough, one sees that

$$\beta(\boldsymbol{u},\boldsymbol{v}) = \left(\mathcal{M}(\boldsymbol{u})^t A \mathcal{M}(\boldsymbol{v})\right)_{1,1}$$

where A is the $n \times n$ matrix $A = (A_{i,j})$. For this reason, and especially when $\boldsymbol{u}, \boldsymbol{v}$ are already column vectors, it's common to draw no distinction between $\mathbb{K}^{1,1}$ and $\mathbb{K}$ and simply write

$$\beta(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{u}^t A \boldsymbol{v}$$

Notation

The set of bilinear forms on V is denoted $V^{(2)}$.

Theorem 9A.3

$V^{(2)}$ is a vector space with the usual function addition/scalar multiplication.

Proof. This is an exercise to the reader. ☐

Definition

Suppose $\mathcal{B} = \{\boldsymbol{b_1}, \ldots, \boldsymbol{b_n}\}$ is a basis on V and $\beta \in V^{(2)}$. The **matrix of $\boldsymbol{\beta}$ with respect to the basis $\boldsymbol{\mathcal{B}}$** is the $n \times n$ matrix $\mathcal{M}(\beta, \mathcal{B})$ whose (i,j)-entry is given by $\beta(\boldsymbol{b_i}, \boldsymbol{b_j})$.

When the basis is clear from context, one just writes $\mathcal{M}(\beta)$.

Example 9A.4

Find the matrix for the standard dot product on $\mathbb{R}^n$, c.f. Example 9A.1.

$$\mathcal{M}(\beta) = \begin{pmatrix} \boldsymbol{e_1}\cdot\boldsymbol{e_1} & \boldsymbol{e_1}\cdot\boldsymbol{e_2} & \cdots & \boldsymbol{e_1}\cdot\boldsymbol{e_n} \\ \boldsymbol{e_2}\cdot\boldsymbol{e_1} & \boldsymbol{e_2}\cdot\boldsymbol{e_2} & \cdots & \boldsymbol{e_2}\cdot\boldsymbol{e_n} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{e_n}\cdot\boldsymbol{e_1} & \boldsymbol{e_n}\cdot\boldsymbol{e_2} & \cdots & \boldsymbol{e_n}\cdot\boldsymbol{e_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = I$$

Example 9A.5

Find the matrix for the bilinear form in Example 9A.2.

Let $\{\boldsymbol{e_1},\dots,\boldsymbol{e_n}\}$ be the standard basis for $\mathbb{K}^n$. Then

$$\mathcal{M}(\beta)=\begin{pmatrix}\beta(\boldsymbol{e_1},\boldsymbol{e_1}) & \beta(\boldsymbol{e_1},\boldsymbol{e_2}) & \cdots & \beta(\boldsymbol{e_1},\boldsymbol{e_n})\\ \beta(\boldsymbol{e_2},\boldsymbol{e_1}) & \beta(\boldsymbol{e_2},\boldsymbol{e_2}) & \cdots & \beta(\boldsymbol{e_2},\boldsymbol{e_n})\\ \vdots & \vdots & \cdots & \vdots\\ \beta(\boldsymbol{e_n},\boldsymbol{e_1}) & \beta(\boldsymbol{e_n},\boldsymbol{e_2}) & \cdots & \beta(\boldsymbol{e_n},\boldsymbol{e_n})\end{pmatrix}=\begin{pmatrix}A_{1,1} & A_{1,2} & \cdots & A_{1,n}\\ A_{2,1} & A_{2,2} & \cdots & A_{2,n}\\ \vdots & \vdots & & \vdots\\ A_{n,1} & A_{n,2} & \cdots & A_{n,n}\end{pmatrix}$$

Theorem 9A.6

$\dim V^{(2)}=(\dim V)^2$

We show that the map

$$\begin{aligned}\Phi : V^{(2)} &\to \mathbb{K}^{n,n}\\ \beta &\mapsto \mathcal{M}(\beta)\end{aligned}$$

is an isomorphism.

Proof. Let $\{\boldsymbol{e_1},\dots,\boldsymbol{e_n}\}$ be a basis for V.

Linearity. Let $\alpha,\beta\in V^{(2)}$ and $k\in\mathbb{K}$. Then

$$\mathcal{M}(k\alpha+\beta)_{i,j}=(k\alpha+\beta)(\boldsymbol{e_i},\boldsymbol{e_j})=k\alpha(\boldsymbol{e_i},\boldsymbol{e_j})+\beta(\boldsymbol{e_i},\boldsymbol{e_j})=k\mathcal{M}(\alpha)_{i,j}=\mathcal{M}(\beta)_{i,j}$$

Injective. Suppose $\alpha,\beta\in V^{(2)}$ satisfy

$$\mathcal{M}(\alpha)=\Phi(\alpha)=\Phi(\beta)=\mathcal{M}(\beta).$$

Then the (i,j)-entry of these matrices agree, hence $\alpha(\boldsymbol{e_i},\boldsymbol{e_j})=\beta(\boldsymbol{e_i},\boldsymbol{e_j})$ for all i,j, and since functionals are uniquely defined by their values on the basis, then it must be that $\alpha=\beta$.

Surjective. Abusing notation slightly as per Example 9A.2, for every $A\in\mathbb{K}^{n,n}$, we can define $\beta\in V^{(2)}$ via

$$\beta(\boldsymbol{u},\boldsymbol{v})=\boldsymbol{u}^tA\boldsymbol{v}$$

and as we saw in Example 9A.5,

$$\Phi(\beta)=\mathcal{M}(\beta)=A.$$

Therefore Φ is an isomorphism, and so $\dim(V^{(2)})=\dim(\mathbb{K}^{n,n})=n^2=\dim(V)^2$. □

Lemma 9A.7

Suppose $\beta\in V^{(2)}$ and $T\in\mathcal{L}(V)$. Define the following related bilinear forms:

$$\alpha(\boldsymbol{u},\boldsymbol{v})=\beta(T(\boldsymbol{u}),\boldsymbol{v})\qquad\text{and}\qquad\gamma(\boldsymbol{u},\boldsymbol{v})=\beta(\boldsymbol{u},T(\boldsymbol{v}))$$

Then, with respect to any basis,

$$\mathcal{M}(\alpha) = \mathcal{M}(T)^t\mathcal{M}(\beta) \qquad \text{and} \qquad \mathcal{M}(\gamma) = \mathcal{M}(\beta)\mathcal{M}(T).$$

Proof. Let $\{\boldsymbol{e_1}, \ldots, \boldsymbol{e_n}\}$ be basis for V. Then

$$\begin{aligned}
\mathcal{M}(\alpha)_{i,j} = \alpha(\boldsymbol{e_i}, \boldsymbol{e_j}) &= \beta(T\boldsymbol{e_i}, \boldsymbol{e_j}) \\
&= \beta\left(\sum_{k=1}^{n} \mathcal{M}(T)_{k,i}\boldsymbol{e_k}, \boldsymbol{e_j}\right) \\
&= \sum_{k=1}^{n} \mathcal{M}(T)_{k,i}\beta(\boldsymbol{e_k}, \boldsymbol{e_j}) \\
&= \sum_{k=1}^{n} \mathcal{M}(T)_{k,i}\mathcal{M}(\beta)_{k,j} \\
&= \sum_{k=1}^{n} \mathcal{M}(T^t)_{i,k}\mathcal{M}(\beta)_{k,j} \\
&= \left(\mathcal{M}(T)\mathcal{M}(\beta)\right)_{i,j}.
\end{aligned}$$

Also,

$$\begin{aligned}
\mathcal{M}(\gamma)_{i,j} = \gamma(\boldsymbol{e_i}, \boldsymbol{e_j}) &= \beta(\boldsymbol{e_i}, T\boldsymbol{e_j}) \\
&= \beta\left(\boldsymbol{e_i}, \sum_{k=1}^{n} \mathcal{M}(T)_{k,j}\boldsymbol{e_k}\right) \\
&= \sum_{k=1}^{n} \beta(\boldsymbol{e_i}, \boldsymbol{e_k})\mathcal{M}(T)_{k,j} \\
&= \sum_{k=1}^{n} \mathcal{M}(\beta)_{i,k}\mathcal{M}(T)_{k,j} \\
&= \left(\mathcal{M}(\beta)\mathcal{M}(T)\right)_{i,j}.
\end{aligned}$$

□

Theorem 9A.8: Change-of-basis

Suppose $\beta \in V^{(2)}$ and let $\mathcal{B}, \mathcal{C}$ be two bases for V. Write $B = \mathcal{M}(\beta, \mathcal{B})$ and $C = \mathcal{M}(\beta, \mathcal{C})$ and let A be the change-of-basis matrix

$$A = \mathcal{M}(\mathrm{Id}, \mathcal{B}, \mathcal{C}).$$

Then

$$B = A^tCA$$

Proof. Let $\mathcal{B} = \{\boldsymbol{b_1}, \ldots, \boldsymbol{b_n}\}$ and $\mathcal{C} = \{\boldsymbol{c_1}, \ldots, \boldsymbol{c_n}\}$, and define $T \in \mathcal{L}(V)$ via

$$T(\boldsymbol{c_i}) = \boldsymbol{b_i}, \qquad \text{for all } i = 1, \ldots, n$$

Observe that T is linear and invertible. We then have that

$$\begin{aligned} B_{i,j} &= \beta(\boldsymbol{b_i}, \boldsymbol{b_j}) \\ &= \beta(T(\boldsymbol{c_i}), T(\boldsymbol{c_j})) \\ &= \sum_{k=1}^{n} \mathcal{M}(T)^t_{i,k} \beta(\boldsymbol{c_k}, T(\boldsymbol{c_j})) && \text{(applyying Lemma 9A.7)} \\ &= \sum_{k=1}^{n} \sum_{\ell=1}^{n} \mathcal{M}(T)^t_{i,k} \beta(\boldsymbol{c_k}, \boldsymbol{c_\ell}) \mathcal{M}(T)_{\ell,j} && \text{(applyying Lemma 9A.7)} \\ &= (\mathcal{M}(T)^t C \mathcal{M}(T))_{i,j} \end{aligned}$$

and thus taking $\mathcal{M}(T) = A$, we have the desired result. □

9A.II Symmetric Bilinear Forms

Definition: symmetric bilinear form

A bilinear form $\beta \in V^{(2)}$ is called **symmetric** if, for all $\boldsymbol{u}, \boldsymbol{v} \in V$,

$$\beta(\boldsymbol{u}, \boldsymbol{v}) = \beta(\boldsymbol{v}, \boldsymbol{u})$$

The set of all symmetric bilinear forms is denoted $V^{(2)}_{sym}$.

Example 9A.9

Show that the dot product, Example 9A.1, is a symmetric bilinear form.

Example 9A.10

Show that there is a matrix A for which the bilinear form Example 9A.2 is *not* a symmetric bilinear form.

Consider $V = \mathbb{K}^2$ and let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Define

$$\beta(\boldsymbol{u}, \boldsymbol{v}) = \left(\mathcal{M}(\boldsymbol{u})^t \, A \, \mathcal{M}(\boldsymbol{v})\right)_{1,1}$$

(bilinearity is up to the reader to check) Then

$$\beta(\boldsymbol{e_1}, \boldsymbol{e_2}) = 2 \qquad \text{and} \qquad \beta(\boldsymbol{e_2}, \boldsymbol{e_1}) = 3.$$

Definition: symmetric matrix

A matrix A is called **symmetric** if $A = A^t$.

Theorem 9A.11

Let $\beta \in V^{(2)}$ where $\dim(V) = n$. The following are equivalent.

1. β is a symmetric bilinear form.
2. For every basis $\mathcal{B}$, $\mathcal{M}(\beta, \mathcal{B})$ is a symmetric matrix.
3. For some basis $\mathcal{B}$, $\mathcal{M}(\beta, \mathcal{B})$ is a symmetric matrix.
4. For some basis $\mathcal{B}$, $\mathcal{M}(\beta, \mathcal{B})$ is a diagonal matrix.

Proof. Let $\beta \in V^{(2)}$.

(1 ⇒ 2) Suppose β is symmetric and let $\{\boldsymbol{e_1}, \ldots, \boldsymbol{e_n}\}$ be any basis for V. Then since $\beta(\boldsymbol{e_i}, \boldsymbol{e_j}) = \beta(\boldsymbol{e_j}, \boldsymbol{e_i})$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, n$, $\mathcal{M}(\beta)$ must be symmetric for every basis.

(2 ⇒ 3) This is immediate.

(3 ⇒ 1) Suppose $\mathcal{M}(\beta)$ is symmetric with respect to some basis $\{\boldsymbol{e_1}, \ldots, \boldsymbol{e_n}\}$. Then for all $\boldsymbol{v}, \boldsymbol{w} \in V$,

$$\begin{aligned}
\beta(\boldsymbol{v}, \boldsymbol{w}) &= \beta\left(\sum_{i=1}^{n} v_i \boldsymbol{e_i}, \sum_{j=1}^{n} w_j \boldsymbol{e_j}\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} v_i w_j \beta(\boldsymbol{e_i}, \boldsymbol{e_j}) && \text{(bilinearity)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} v_i w_j \beta(\boldsymbol{e_j}, \boldsymbol{e_i}) && \text{(symmetry of } \mathcal{M}(\beta)\text{)} \\
&= \beta\left(\sum_{i=j}^{n} w_j \boldsymbol{e_j}, \sum_{i=1}^{n} v_i \boldsymbol{e_j}\right) \\
&= \beta(\boldsymbol{w}, \boldsymbol{v})
\end{aligned}$$

and therefore β is symmetric.

(4 ⇒ 3) Diagonal matrices are always symmetric, so this is immediate.

(1 ⇒ 4) Suppose β is symmetric. Then for any vectors $\boldsymbol{v}, \boldsymbol{w}$ we have

$$\begin{aligned}
\beta(\boldsymbol{v}+\boldsymbol{w}, \boldsymbol{v}+\boldsymbol{w}) &= \beta(\boldsymbol{v}, \boldsymbol{v}) + \beta(\boldsymbol{w}, \boldsymbol{v}) + \beta(\boldsymbol{v}, \boldsymbol{w}) + \beta(\boldsymbol{w}, \boldsymbol{w}) \\
&= \beta(\boldsymbol{v}, \boldsymbol{v}) + \beta(\boldsymbol{v}, \boldsymbol{w}) + \beta(\boldsymbol{v}, \boldsymbol{w}) + \beta(\boldsymbol{w}, \boldsymbol{w}) \\
&= \beta(\boldsymbol{v}, \boldsymbol{v}) + 2\beta(\boldsymbol{v}, \boldsymbol{w}) + \beta(\boldsymbol{w}, \boldsymbol{w})
\end{aligned}$$

and this rearranges to

$$2\beta(\boldsymbol{v}, \boldsymbol{w}) = \beta(\boldsymbol{v}+\boldsymbol{w}, \boldsymbol{v}+\boldsymbol{w}) - \beta(\boldsymbol{v}, \boldsymbol{v}) - \beta(\boldsymbol{w}, \boldsymbol{w}) \tag{9.1}$$

If β is the zero map, then β is the zero matrix (which is a diagonal matrix). As well, if V is 1-dimensional, then β is a 1×1 matrix and is diagonal. So, suppose β is not the zero map and $\dim(V) = n > 1$. Then Equation 9.1 implies that there is some vector $\boldsymbol{u_1}$ for which $\beta(\boldsymbol{u_1}, \boldsymbol{u_1}) = k_{1,1} \neq 0$. Define

$$U_2 = \{\boldsymbol{v} \in V \ : \ \beta(\boldsymbol{v}, \boldsymbol{u_1}) = 0\}$$

Letting $T_2 \in \mathcal{L}(V, \mathbb{K})$ be given by $T_2(\boldsymbol{x}) = \beta(\boldsymbol{x}, \boldsymbol{u_1})$, then we have that $U_2 = \operatorname{Null}(T_2)$, and since T is not the zero map, it must have rank 1, whence $\dim(U_2) = n - 1$.

See now that we have

$$V = \operatorname{Span}(\boldsymbol{u_1}) \oplus U_2$$

(the directness of the sum requires a small argument that we'll reserve for later).

At this point I'd like to look at where we're at. Taking $\{\boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$ to be a basis for U, we have

$$\mathcal{M}(\beta) = \begin{pmatrix} \beta(\boldsymbol{u_1}, \boldsymbol{u_1}) & \beta(\boldsymbol{u_1}, \boldsymbol{v_2}) & \cdots & \beta(\boldsymbol{u_1}, \boldsymbol{v_n}) \\ \beta(\boldsymbol{v_2}, \boldsymbol{u_1}) & \beta(\boldsymbol{v_2}, \boldsymbol{v_2}) & \cdots & \beta(\boldsymbol{v_2}, \boldsymbol{v_n}) \\ \vdots & \vdots & & \vdots \\ \beta(\boldsymbol{v_n}, \boldsymbol{u_1}) & \beta(\boldsymbol{v_n}, \boldsymbol{v_2}) & \cdots & \beta(\boldsymbol{v_n}, \boldsymbol{v_n}) \end{pmatrix} = \left(\begin{array}{c|ccc} k_{1,1} & 0 & \cdots & 0 \\ \hline 0 & \beta(\boldsymbol{v_2}, \boldsymbol{v_2}) & \cdots & \beta(\boldsymbol{v_2}, \boldsymbol{v_n}) \\ \vdots & \vdots & & \vdots \\ 0 & \beta(\boldsymbol{v_n}, \boldsymbol{v_2}) & \cdots & \beta(\boldsymbol{v_n}, \boldsymbol{v_n}) \end{array}\right)$$

The goal will be to iteratively chip away at this bottom-right block.

Now, for each $i \geq 2$, choose $\boldsymbol{u_i} \in U_i$ so that $\beta(\boldsymbol{u_i}, \boldsymbol{u_i}) = k_{i,i} \neq 0$, and then define

$$U_i = \{\boldsymbol{v} \in V \ : \ \beta(\boldsymbol{v}, \boldsymbol{u_{i-1}}) = 0\}.$$

By construction, just as above

$$\dim(U_i) = \dim(U_{i-1}) - 1.$$

This process terminates after $m \leq n + 1$ steps, at which point

$$\beta(\boldsymbol{v}, \boldsymbol{w}) = 0 \text{ for all } \boldsymbol{v}, \boldsymbol{w} \in U_m$$

(and this can be at most $n + 1$, since $\dim(V) = n$ and thus $U_{n+1} = \{\boldsymbol{0}\}$.)

At this point, we have

$$V = \operatorname{Span}(\boldsymbol{u_1}) \oplus \operatorname{Span}(\boldsymbol{u_2}) \oplus \cdots \oplus \operatorname{Span}(\boldsymbol{u_{m-1}}) \oplus U_m$$

(where we simply omit U_m above in the case that $m = n + 1$). And now, taking a basis $\{\boldsymbol{v_m}, \dots, \boldsymbol{v_n}\}$ for U_m, we have

$$\mathcal{M}(\beta, \ \{\boldsymbol{u_1}, \dots, \boldsymbol{u_{m-1}}, \boldsymbol{v_m}, \dots, \boldsymbol{v_n}\}) = \left(\begin{array}{cccc|ccc} k_{1,1} & & & & & & \\ & k_{2,2} & & & & & \\ & & \ddots & & & & \\ & & & k_{m-1,m-1} & & & \\ \hline & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{array}\right)$$

Now, to verify that this set is a basis (at which point the direct sum claim) will follow, let

$$\boldsymbol{w} = c_1 \boldsymbol{u_1} + \cdots + c_{m-1} \boldsymbol{u_{m-1}} + c_m \boldsymbol{v}_m + \cdots + c_n \boldsymbol{v_n}$$

By construction

$$\begin{cases} \beta(\boldsymbol{w}, \boldsymbol{u_1}) = c_1\beta(\boldsymbol{u_1}, \boldsymbol{u_1}) \\ \vdots \\ \beta(\boldsymbol{w}, \boldsymbol{u_{m-1}}) = c_{m-1}\beta(\boldsymbol{u_{m-1}}, \boldsymbol{u_{m-1}}) \end{cases}$$

So if $\boldsymbol{w} = \mathbf{0}$, then the above implies $c_1 = \cdots = c_{m-1} = 0$, and since the $\boldsymbol{v_i}$ vectors were already linearly independent, so $c_m = \cdots = c_n = 0$.

☐

Definition: alternating form

A bilinear form $\beta \in V^{(2)}$ is called **alternating** if, for every $\boldsymbol{v} \in V$,

$$\beta(\boldsymbol{v}, \boldsymbol{v}) = 0.$$

The set of alternating bilinear forms is denoted $V^{(2)}_{alt}$.

Remark. Alternating forms generalize the idea of a determinant of a 2 × 2 matrix.

Example 9A.12: canonical example

Show that the following map β is an alternating bilinear form on $\mathbb{R}^2$:

$$\beta((a, b),\ (c, d)) = ad - bc.$$

Theorem 9A.13

A bilinear form α on V is alternating if and only if, for all $\boldsymbol{u}$ and $\boldsymbol{v}$,

$$\alpha(\boldsymbol{u}, \boldsymbol{v}) = -\alpha(\boldsymbol{v}, \boldsymbol{u}).$$

Proof. INCOMPLETE ☐

Corollary 9A.14

α is an alternating bilinear form if and only if, for every basis $\mathcal{B}$,

$$\mathcal{M}(\alpha, \mathcal{B})^t = -\mathcal{M}(\alpha, \mathcal{B})$$

Definition: skew-symmetric matrix

A matrix A is called **skew-symmetric** if it satisfies $A^T = -A$.

Theorem 9A.15

The sets $V^{(2)}_{sym}$ and $V^{(2)}_{alt}$ are subspaces of $V^{(2)}$. Moreover

$$V^{(2)} = V^{(2)}_{sym} \oplus V^{(2)}_{alt}.$$

Proof. INCOMPLETE

□

Corollary 9A.16

Let $U_{sym}, U_{skew} \subset \mathbb{K}^{n,n}$ be the subspaces of symmetric and skew-symmetric matrices, respectively. Then

$$\mathbb{K}^{n,n} = U_{sym} \oplus U_{skew}.$$

9A.III Quadratic Forms

Definition: quadratic form

For any bilinear form β, the function $q_\beta : V \to \mathbb{K}$ given by $q_\beta(\boldsymbol{v}) = \beta(\boldsymbol{v}, \boldsymbol{v})$ is called a **quadratic form** on V.

Remark. Quadratic forms generalize the notion of a norm/distance (from the origin).

Example 9A.17

Let $q : \mathbb{R}^2 \to \mathbb{R}$ be the function

$$q(x, y, z) = x^2 + y^2 + z^2.$$

Show that q is a quadratic form by finding the associated bilinear form.

Take β to be the usual dot product:

$$\beta((x_1, y_1, z_1),\ (x_2, y_2, z_2)) = x_1x_2 + y_1y_2 + z_1z_2.$$

Then $\beta((x, y, z),\ (x, y, z)) = x^2 + y^2 + z^2 = q(x, y, z)$.

It turns out that there are no interesting quadratic forms on $\mathbb{K}^n$.

Theorem 9A.18

A function $q : \mathbb{K}^n \to \mathbb{K}$ is a quadratic form if and only if there are scalars $A_{i,j}$ ($i = 1, \dots, n$ and $j = 1, \dots, n$) for which

$$q(x_1, \dots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j} x_i x_j.$$

(Or, abusing notation slightly) ...if and only if there is a matrix A for which

$$q(\boldsymbol{v}) = \boldsymbol{v}^t A \boldsymbol{v}.$$

Proof. INCOMPLETE ☐

Theorem 9A.19: Characterization of quadratic forms

Let $q : V \to \mathbb{K}$ be a function. The following are equivalent

1. q is a quadratic form.
2. There is a unique symmetric bilinear for β on V so that $q(\boldsymbol{v}) = \beta(\boldsymbol{v}, \boldsymbol{v})$.
3. $q(\lambda \boldsymbol{v}) = \lambda^2 q(\boldsymbol{v})$ for all scalars λ and all $\boldsymbol{v} \in V$, and the function

$$\beta(\boldsymbol{u}, \boldsymbol{v}) = q(\boldsymbol{u} + \boldsymbol{v}) - q(\boldsymbol{u}) - q(\boldsymbol{v})$$

is a bilinear form on V.

4. $q(2\boldsymbol{v}) = 2^2 q(\boldsymbol{v})$ for all $\boldsymbol{v} \in V$, and the function

$$\beta(\boldsymbol{u}, \boldsymbol{v}) = q(\boldsymbol{u} + \boldsymbol{v}) - q(\boldsymbol{u}) - q(\boldsymbol{v})$$

is a bilinear form on V.

Proof. ☐

Theorem 9A.20

For any quadratic form $q : V \to \mathbb{K}$, there is a basis $\{\boldsymbol{b_1}, \ldots, \boldsymbol{b_n}\}$ of V and scalars $\lambda_1, \ldots, \lambda_n$ so that, for all $\boldsymbol{v} = v_1 \boldsymbol{b_1} + \cdots + v_n \boldsymbol{v}$,

$$q(\boldsymbol{v}) = q(v_1 \boldsymbol{b_1} + \cdots + v_n \boldsymbol{v}) = \lambda_1 v_1^2 + \cdots + \lambda_n v_n^2.$$

Proof. ☐

9B Alternating Multilinear Forms

One can extend the idea of a bilinear form to a *multilinear form* in a straightforward way:

Notation: n-fold Cartesian Product

Let V be a $\mathbb{K}$-vector space and $m > 0$. Then

$$V^m := \underbrace{V \times \cdots \times V}_{m}$$

is the n-fold Cartesian product of V.

Fact. V^m is a $\mathbb{K}$-vector space with componentwise operations.

Definition: multilinear form

A map $\beta : V^m \to \mathbb{K}$ is called a **multilinear form** or an **m-linear form** if it is linear in each component. So for all scalars k and all vectors $\boldsymbol{u}, \boldsymbol{v_1}, \dots, \boldsymbol{v_m}$,

$$\begin{gathered}\beta(\boldsymbol{v_1} + k\boldsymbol{u}, \boldsymbol{v_2}, \dots, \boldsymbol{v_m}) = \beta(\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_m}) + k\beta(\boldsymbol{u}, \boldsymbol{v_2}, \dots, \boldsymbol{v_m}) \\ \vdots \\ \beta(\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_m} + k\boldsymbol{u}) = \beta(\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_m}) + k\beta(\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{u})\end{gathered}$$

The set of m-linear forms is denoted $V^{(m)}$.

Example 9B.1

Every bilinear form is a “2-linear form.”

Example 9B.2

Show that the map $\beta : \mathbb{R}^4 \to \mathbb{R}$ given by

$$\beta(\boldsymbol{v_1}, \dots, \boldsymbol{v_4}) = (\boldsymbol{v_1} \cdot \boldsymbol{v_2})(\boldsymbol{v_3} \cdot \boldsymbol{v_4})$$

is a 4-linear form.

This is straightforward to show. Multilinearity is essentially inherited from the bilinearity of the dot product.

Definition: alternating multilinear form

An m-linear form β is called **alternating** if

$$\beta(\boldsymbol{v_1}, \dots, \boldsymbol{v_m}) = 0$$

whenever $\boldsymbol{v_i} = \boldsymbol{v_j}$ for some distinct i, j in $1, 2, \dots, m$.

The set of alternating m-linear forms is denoted $V^{(m)}_{alt}$.

Example 9B.3

Let V be a $\mathbb{K}$-vector space and β a 3-linear form on V. Show that the map $\alpha : V^3 \to \mathbb{K}$ given by

$$\begin{aligned}\alpha(\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}) =&\beta(\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}) + \beta(\boldsymbol{v_2}, \boldsymbol{v_3}, \boldsymbol{v_1}) + \beta(\boldsymbol{v_3}, \boldsymbol{v_1}, \boldsymbol{v_2}) \\ &- \beta(\boldsymbol{v_2}, \boldsymbol{v_1}, \boldsymbol{v_3}) - \beta(\boldsymbol{v_3}, \boldsymbol{v_2}, \boldsymbol{v_1}) - \beta(\boldsymbol{v_1}, \boldsymbol{v_3}, \boldsymbol{v_2})\end{aligned}$$

is an alternating 3-linear form on V.

That α is 3-linear follows immediately from the fact that it's defined entirely in terms of the βs, which are themselves 3-linear (although the reader is encouraged to check this).

Checking that α is alternating requires checking that it holds when the same vector is placed in any two of the input arguments. This means checking

$$\alpha(\boldsymbol{v}, \boldsymbol{v}, \boldsymbol{w}) = \alpha(\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{v}) = \alpha(\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{v}) = 0.$$

We show only the first of these cases as the others are straightforward and nearly identical.

$$\begin{aligned}\alpha(\boldsymbol{v}, \boldsymbol{v}, \boldsymbol{w}) =&\beta(\boldsymbol{v}, \boldsymbol{v}, \boldsymbol{w}) + \beta(\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{v}) + \beta(\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{v}) \\ &-\beta(\boldsymbol{v}, \boldsymbol{v}, \boldsymbol{w})-\beta(\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{v})-\beta(\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{v}) \\ =&0.\end{aligned}$$

Proposition 9B.4

Suppose β is an alternating m-linear form. Then

$$\beta(\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}) = 0$$

whenever $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}\}$ is linearly dependent.

Proof. Sketch Write down the offending vector as a linear combination of the others. Observe that, through multilinearity, this can be rewritten as a sum of the β's where each term has a repeated $\boldsymbol{v_m}$ in the list. ☐

Corollary 9B.5

If $m > \dim V$, then $V_{alt}^{(m)} = \{0\}$.

9B.I Alternating Forms and Permutations

Theorem 9B.6: Swapping inputs changes sign

If β is an alternating m-linear form on V, then

$$\beta(\boldsymbol{v_1}, \ldots, \boldsymbol{v_i}, \ldots, \boldsymbol{v_j}, \ldots, \boldsymbol{v_m}) = -\beta(\boldsymbol{v_1}, \ldots, \boldsymbol{v_j}, \ldots, \boldsymbol{v_i}, \ldots, \boldsymbol{v_m})$$

Proof. INCOMPLETE ☐

Definition: permutation, transposition

A **permutation**, σ, of an ordered m-tuple $(1, 2, \ldots, m)$ is formally a bijection

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

the amounts to simply reordering of the entries. The relabeled m-tuple will be written $\big(\sigma(1), \sigma(2), \ldots, \sigma(m)\big)$.

A **transposition** is a reordering obtained by swapping exactly two entries.

Below we state some facts that one would typically see in a first semester course in abstract algebra (relevant Google search term: “symmetric group”).

Theorem 9B.7: Group Theory Facts

1. There are $m!$ distinct permutations of any m-tuple $(1, 2, \ldots, m)$.

2. Every permutation can be realized as a (non-unique) sequence of transpositions.
3. Given any two sequences of transpositions corresponding to a single permutation, the two sequences will either both have an even number of transpositions, or both have an odd number of transpositions.

In this way, the following is well-defined.

Definition: sign of a permutation

The **sign** of a permutation σ is

$$\operatorname{sgn}(\sigma) = (-1)^N$$

where N is the length of the sequence of transpositions which produce σ.

Example 9B.8

Let σ be the permutation for which

$$\big(\sigma(1), \sigma(2), \sigma(3)\big) = (3, 2, 1)$$

Write σ in two different ways as sequences of transpositions. Then find $\operatorname{sgn}(\sigma)$.

σ is already a transposition, so $\operatorname{sgn}(\sigma) = (-1)^1 = -1$. As well

$$\sigma : (1,2,3) \underset{\tau_1}{\mapsto} (2,1,3) \underset{\tau_2}{\mapsto} (2,3,1) \underset{\tau_3}{\mapsto} (3,2,1).$$

so we also have $\operatorname{sgn}(\sigma) = (-1)^3 = -1$.

Theorem 9B.6 can thus be upgraded to

Theorem 9B.9: Permutations and alternating multilinear forms

If β is an altnerating m-linear form on V and σ is a permutation, then for any set of vectors $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}\}$,

$$\beta(\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}) = \operatorname{sgn}(\sigma)\, \beta\big(\boldsymbol{v_{\sigma(1)}}, \, \ldots, \, \boldsymbol{v_{\sigma(m)}}\big).$$

Proof. INCOMPLETE

□

Theorem 9B.10: Formula for alternating $(\dim V)$-linear forms on V

Let V be an n-dimensional $\mathbb{K}$-vector space with basis $\{\boldsymbol{e_1}, \ldots, \boldsymbol{e_n}\}$, and let $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}\}$ be a set of vectors in V. Write

$$\boldsymbol{v_i} = \sum_{j=1}^{n} A_{j,i} \boldsymbol{e_j} \qquad \text{where} \quad A_{j,i} \in \mathbb{K}$$

Then, for every $\beta \in V_{alt}^{(n)}$,

$$\beta(\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}) = \beta(\boldsymbol{e_1}, \ldots, \boldsymbol{e_n}) \left(\sum_{\substack{\text{perm.} \\ \sigma}} \operatorname{sgn}(\sigma) A_{\sigma(1),1} A_{\sigma(2),2} \cdots A_{\sigma(n),n} \right)$$

Proof. INCOMPLETE

□

Theorem 9B.11

Let V be an n-dimensional $\mathbb{K}$-vector space. The vector space $V_{alt}^{(n)}$ is 1-dimensional.

This result tells us that every alternating n-linear form on V will be a scalar multiple of the "determinant"... whatever *that* is... we should probably define it in the next section.

Proof. INCOMPLETE

$\square$

9C Determinants

Definition

Suppose that m is a polisitve integer and $T \in \mathcal{L}(V)$. For $\alpha \in V_{alt}^{(m)}$, define $\alpha_T \in V^{(m)}$ by

$$\alpha_T(\boldsymbol{v_1}, \dots, \boldsymbol{v_m}) = \alpha(T(\boldsymbol{v_1}), \dots, T(\boldsymbol{v_m}))$$

Proposition 9C.1

For any $\alpha \in V_{alt}^{(m)}$ and any operator $T \in \mathcal{L}(V)$, α_T is alternating. Moreover, for each T, the function $\Phi_T : V_{alt}^{(m)} \to V_{alt}^{(m)}$ given by

$$\Phi_T(\alpha) = \alpha_T$$

is a linear transformation.

Proof. Let $T \in \mathcal{L}(V)$ and let $\alpha \in V_{alt}^{(m)}$. Let $(\boldsymbol{v_1}, \dots, \boldsymbol{v_m})$ be an m-tuple of vectors where $\boldsymbol{v_i} = \boldsymbol{v_j}$ for some $i \neq j$. Then $T(\boldsymbol{v_i}) = T(\boldsymbol{v_j})$, hence

$$\alpha_T(\boldsymbol{v_1}, \dots, \boldsymbol{v_m}) = \alpha(T(\boldsymbol{v_1}, \dots, T(\boldsymbol{v_m})) = 0.$$

The linearity of Φ_T follows from the fact that linear combinations of alternating m-linear forms are still alternating m-linear forms:

$$\begin{aligned}\Phi_T(\alpha + k\beta)(\boldsymbol{v_1}, \dots, \boldsymbol{v_m}) &= (\alpha + k\beta)(T(\boldsymbol{v_1}), \dots, T(\boldsymbol{v_m})) \\ &= \alpha(T(\boldsymbol{v_1}), \dots, T(\boldsymbol{v_m})) + k\beta(T(\boldsymbol{v_1}), \dots, T(\boldsymbol{v_m})) \\ &= \alpha_T(\boldsymbol{v_1}, \dots, \boldsymbol{v_m}) + k\beta(\boldsymbol{v_1}, \dots, \boldsymbol{v_m})\end{aligned}$$

□

Quick discussion of uniqueness...If $\dim(V) = n$, then $V_{alt}^{(n)}$ is one-dimensional (Theorem 9B.11), so Φ_T acts by multiplying by a scalar, call it k_T. This scalar is fixed (hence unique and not dependent upon α), and so we give a name to it.

Definition: determinant of an operator

Let T be an operator on V – an n-dimensional $\mathbb{K}$-vector space. The **determinant of T** is the unique scalar $\det T$ so that, for all $\alpha \in V_{alt}^{(n)}$,

$$\alpha_T = (\det T)\alpha.$$

Example 9C.2

Let $\dim(V) = n$ and $T = \mathrm{Id} \in \mathcal{L}(V)$. Find $\det(T)$.

For all tuples $(\boldsymbol{v_1}, \ldots, \boldsymbol{v_n})$ and $\alpha \in V_{alt}^{(n)}$,

$$\alpha_{\mathrm{Id}}(\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}) = \alpha(\mathrm{Id}\, \boldsymbol{v_1}, \ldots, \mathrm{Id}\, \boldsymbol{v_n}) = \alpha(\boldsymbol{v_1}, \ldots, \boldsymbol{v_n})$$

whence $\det \mathrm{Id} = 1$.

It's worth noting that our definition of the determinant is independent of the basis, so one can compute the determinant from a convenient basis.

Example 9C.3

Let $\dim(V) = n$ and $T \in \mathcal{L}(V)$. If $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}\}$ is a basis of eigenvectors with eigenvalues $\lambda_1, \ldots, \lambda_n$. Find $\det(T)$.

$$\alpha_T(\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}) = \alpha(T\boldsymbol{v_1}, \ldots, T\boldsymbol{v_n}) = \alpha(\lambda_1\boldsymbol{v_1}, \ldots, \lambda_n\boldsymbol{v_n}) = \lambda_1 \cdots \lambda_n \alpha(\boldsymbol{v_1}, \ldots, \boldsymbol{v_n})$$

where the last equality follows from n-linearity of α. Therefore $\det(T) = \lambda_1 \cdots \lambda_n$.

Example 9C.4

Let $T \in \mathcal{L}(\mathbb{R}^2)$ be the linear operator given by

$$T(x, y) = (ax + by, cx + dy)$$

Find $\det(T)$.

Letting $\boldsymbol{e_1} = (1, 0)$ and $\boldsymbol{e_2}$, we have that

$$T : \begin{cases} \boldsymbol{e_1} \mapsto a\boldsymbol{e_1} + c\boldsymbol{e_2} = A_{1,1}\boldsymbol{e_1} + A_{2,1}\boldsymbol{e_2} \\ \boldsymbol{e_2} \mapsto b\boldsymbol{e_1} + d\boldsymbol{e_2} = A_{1,2}\boldsymbol{e_1} + A_{2,2}\boldsymbol{e_2} \end{cases}$$

From Theorem 9B.10, we have that

$$\begin{aligned} \alpha_T(\boldsymbol{e_1}, T\boldsymbol{e_2}) &= \alpha(A_{1,1}\boldsymbol{e_1} + A_{2,1}\boldsymbol{e_2},\ A_{1,2}\boldsymbol{e_1} + A_{2,2}T\boldsymbol{e_2}) \\ &= \alpha(\boldsymbol{e_1}, \boldsymbol{e_2}) \left(\sum_{\substack{\text{perm.} \\ \sigma}} \mathrm{sgn}(\sigma) A_{\sigma(1),1} A_{\sigma(2),2} \right) \end{aligned}$$

So

$$\det(T) = \sum_{\substack{\text{perm.} \\ \sigma}} \mathrm{sgn}(\sigma) A_{\sigma(1),1} A_{\sigma(2),2}$$

Observe that there are only two permutations of $(1, 2)$:

$$\iota : (1, 2) \mapsto (1, 2) \qquad \text{and} \qquad \tau : (1, 2) \mapsto (2, 1)$$

with $\mathrm{sgn}(\iota) = 1$ and $\mathrm{sgn}(\tau)$. As such

$$\det(T) = \mathrm{sgn}(\iota) A_{\iota(1),1} A_{\iota(2),2} + \mathrm{sgn}(\tau) A_{\tau(1),1} A_{\tau(2),2} = A_{1,1}A_{2,2} - A_{2,1}A_{2,2} = ad - bc.$$

Definition: "nicest" (with respect to a basis) alternating n-linear form

Let $\mathcal{B} = \{\boldsymbol{b_1}, \dots, \boldsymbol{b_n}\}$ be a basis for V and $\{\varphi_1, \dots, \varphi_n\}$ be the dual basis, i.e.

$$\varphi_i \in \mathcal{L}(V, \mathbb{K}) \qquad \text{where} \qquad \varphi_i(\boldsymbol{b_j}) = \begin{cases} 1 & \text{when } i = j, \\ 0 & \text{when } i \neq j. \end{cases}$$

Call the alternating n-linear form constructed in the proof of Theorem 9B.11 the **"nicest" alternating n-linear form** with respect to the $\mathcal{B}$-basis:

$$\begin{aligned} &\alpha : V^m \to \mathbb{K} \\ &\alpha(\boldsymbol{v_1}, \dots, \boldsymbol{v_n}) = \sum_{\substack{\text{perm.} \\ \sigma}} \varphi_{\sigma(1)}(\boldsymbol{v_1}) \cdots \varphi_{\sigma(n)}(\boldsymbol{v_n}). \end{aligned}$$

Remark. The "nicest" form α with respect to the basis $\mathcal{B} = \{\boldsymbol{b_1}, \dots, \boldsymbol{b_n}\}$ satisfies

$$\alpha(\boldsymbol{b_1}, \dots, \boldsymbol{b_n}) = 1.$$

Definition: determinant of a matrix

Let $T \in \mathcal{L}(V)$ be an operator and let $A = \mathcal{M}(T, \mathcal{B})$ be the $n \times n$ matrix of T with respect to the $\mathcal{B}$ basis. Then the **determinant of $\boldsymbol{A}$**, $\det A$, is given by $\det(A) := \det(T)$.

Example 9C.5

Let $T \in \mathcal{L}(\mathbb{R}^2)$ be the linear operator given by

$$T(x, y) = (ax + by, cx + dy)$$

In the standard basis, $A = \mathcal{M}(T) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and, from our work, $\det(A) = ad - bc$.

Theorem 9C.6: Computing $\det A$

Let $A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,n} \end{pmatrix}$. Then

$$\det A = \sum_{\substack{\text{perm.} \\ \sigma}} \operatorname{sgn}(\sigma) A_{\sigma(1),1} \cdots A_{\sigma(n),n}.$$

Proof. Given our matrix A, we can find an operator $T \in \mathcal{L}(V)$, a basis $\mathcal{B} = \{\boldsymbol{b_1}, \dots, \boldsymbol{b_n}\}$ for which

$$A = \mathcal{M}(T, \mathcal{B}).$$

Let α be the "nicest" form with respect to the $\mathcal{B}$-basis. It follows that

$$\begin{aligned}\det(A) &= \det(T)\\ &= \det(T)\alpha(\boldsymbol{b_1},\dots,\boldsymbol{b_n})\\ &= \alpha(T\boldsymbol{b_1},\dots,T\boldsymbol{b_n})\\ &= \alpha(\boldsymbol{b_1},\dots,\boldsymbol{b_n})\left(\sum_{\substack{\text{perm.}\\ \sigma}} \operatorname{sgn}(\sigma)A_{\sigma(1),1}\cdots A_{\sigma(n),n}.\right) && \text{(by Theorem 9B.10)}\\ &= \sum_{\substack{\text{perm.}\\ \sigma}} \operatorname{sgn}(\sigma)A_{\sigma(1),1}\cdots A_{\sigma(n),n}.\end{aligned}$$

□

Letting $\boldsymbol{a_i}$ be the i^{th} column of an $n\times n$ matrix A and recalling that $\boldsymbol{a_i}$ is just the matrix representation of $T\boldsymbol{b_i}$, the computations in the above previous proof also yield the following:

Corollary 9C.7

The map $\alpha : (\mathbb{K}^n)^n \to \mathbb{K}$ given by

$$\alpha(\boldsymbol{a_1},\dots,\boldsymbol{a_n}) = \det\begin{pmatrix} | & & | \\ \boldsymbol{a_1} & \cdots & \boldsymbol{a_n} \\ | & & | \end{pmatrix}$$

is an alternating n-linear form.

Exercise 9C.8: Determinant of an upper triangular matrix

Suppose $A = \begin{pmatrix} A_{1,1} & * & \cdots & * \\ 0 & A_{2,2} & \dots & * \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & A_{n,n}\end{pmatrix}$. Then

$$\det(A) = A_{1,1}A_{2,2}\cdots A_{n,n}.$$

9C.I Properties of Determinants

Proposition 9C.9: Determinant is multiplicative

1. Suppose $S,T\in\mathcal{L}(V)$. Then $\det(ST)=\det(S)\det(T)$
2. Suppose $A,B\in\mathbb{K}^{n\times n}$. Then $\det(AB)=\det(A)\det(B)$.

Proof. Let $\{\boldsymbol{e_1}, \dots, \boldsymbol{e_n}\}$ and let $\alpha \in V_{alt}^{(n)}$. Then

$$\begin{aligned} \det(ST)\alpha_{ST}(\boldsymbol{e_1}, \dots, \boldsymbol{e_n}) &= \alpha_{ST}(\boldsymbol{e_1}, \dots, \boldsymbol{e_n}) \\ &= \alpha(ST\boldsymbol{e_1}, \dots, ST\boldsymbol{e_2}) \\ &= \det(S)\alpha(T\boldsymbol{e_1}, \dots, T\boldsymbol{e_2}) \\ &= \det(S)\det(T)\alpha(\boldsymbol{e_1}, \dots, \boldsymbol{e_n}) \end{aligned}$$

And therefore $\det(ST) = \det(S)\det(T)$. That the same result holds for matrices is immediate from the definition of the determinant of a matrix. ☐

Proposition 9C.10

An operator $T \in \mathcal{L}(V)$ is invertible iff $\det(T) \neq 0$. Moreover, when T is invertible, $\det(T^{-1}) = \dfrac{1}{\det(T)}$.

Sketch. Observe that

$$1 = \det(\mathrm{Id}) = \det(TT^{-1}) = \det(T)\det(T^{-1}).$$

The proof of the converse requires the observation that $\alpha(\boldsymbol{v_1}, \dots, \boldsymbol{v_n}) = 0$ if and only if the $\boldsymbol{v_i}$'s form a linearly independent set ☐

Proposition 9C.11

1. Suppose A is a square matrix. Then $\det(A^t) = \det(A)$.
2. Suppose $T \in \mathcal{L}(V)$ and let $T' \in \mathcal{L}(V') = \mathcal{L}\big(\mathcal{L}(V, \mathbb{K})\big)$ be the dual map. Then $\det(T') = \det(T)$.

Proof. Recall from a homework assignment that the matrix of T in some basis is precisely the transpose of the matrix for T' in the dual basis. So item 2 follows from item 1.

To prove item 1, we make the following observations about permutations. Permutations of the set $\{1, \dots, n\}$ are formally bijections

$$\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$$

hence

- σ is invertible.
- $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$
- summing over σ's is equal to summing over the σ^{-1}s
- $\sigma(j) = i$ if and only if $\sigma^{-1}(i) = j$

This last observation, when combined with the transpose, gives

$$A^t_{\sigma(i),i} = A^t_{j,i} = A_{i,j} = A_{\sigma^{-1}(j),j}$$

and thus

$$\det(A^t) = \sum_{\substack{\text{perm.}\\ \sigma}} \operatorname{sgn}(\sigma) A^t_{\sigma(1),1} \cdots A^t_{\sigma(n),n} = \sum_{\substack{\text{perm.}\\ \sigma^{-1}}} \operatorname{sgn}(\sigma^{-1}) A_{\sigma^{-1}(1),1} \cdots A^t_{\sigma^{-1}(n),n} = \det(A).$$

☐

Theorem 9C.12: Determinants and row operations

Let A be an $n \times n$ matrix and let B be the $n \times n$ matrix formed by performing an elementary row operation on A.

1. If that elementary row operation is swapping Row_i and Row_j, then $\det(B) = -\det(A)$.
2. If that elementary row operation is scaling Row_i by $k \neq 0$, then $\det(B) = k\det(A)$.
3. If that elementary row operation is adding $k \cdot Row_i$ to Row_j, then $\det(B) = \det(A)$.

Proof. Each of these follow from the fact that $\det(A^t) = \det(A)$ and the multilinearity of the determinant (on the columns) of A. The first two are immediate as they are analogous to swapping/scaling columns of A, but the last one is possibly non-obvious. WLOG, we show only that this holds when adding a multiple of column 1 to column n.

$$\begin{aligned}\det\begin{pmatrix} | & & | \\ \boldsymbol{a_1} & \cdots & (k\boldsymbol{a_1} + \boldsymbol{a_n}) \\ | & & | \end{pmatrix} &= k\det\begin{pmatrix} | & & | \\ \boldsymbol{a_1} & \cdots & \boldsymbol{a_1} \\ | & & | \end{pmatrix} + \det\begin{pmatrix} | & & | \\ \boldsymbol{a_1} & \cdots & \boldsymbol{a_n} \\ | & & | \end{pmatrix} \\ &= 0 + \det\begin{pmatrix} | & & | \\ \boldsymbol{a_1} & \cdots & \boldsymbol{a_n} \\ | & & | \end{pmatrix}.\end{aligned}$$

☐

Exercise 9C.13

Verify the above claim by computing the following determinants explicitly in the 2×2 case:

1. $\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$
2. $\det\begin{pmatrix} c & d \\ a & b \end{pmatrix} =$
3. $\det\begin{pmatrix} ka & kb \\ c & d \end{pmatrix} =$
4. $\det\begin{pmatrix} a & b \\ (c + ak) & (b + dk) \end{pmatrix} =$

9C.II Determinants and Eigenvalues

Proposition 9C.14

λ is an eigenvalue for $T \in \mathcal{L}(V)$ if and only if $\det(T - \lambda\,\mathrm{Id}) = 0$.

Proof. Eigenvalues λ are specifically those values for which $\mathrm{Null}(T - \lambda I)$ is nontrivial, hence $T - \lambda I$ is not invertible. Now apply Proposition 9C.10. $\square$

Definition: characteristic polynomial

Suppose $T \in \mathcal{L}(V)$. The polynomial

$$p(x) = \det(x \,\mathrm{Id} - T)$$

is the **characteristic polynomial of $\boldsymbol{T}$**.

Remark. The above is the convention that your book takes (and it ensures that the monic polynomial has a positive leading term). If one takes the convention that the characteristic polynomial of T is $\det(T - xI)$, then this only differs by possibly a sign. That said, $p(x)$ and $-p(x)$ have the same roots, so it's not a big deal.

Theorem 9C.15: Cayley–Hamilton

Let $T \in \mathcal{L}(V)$ with characteristic polynomial $p(x)$. Then $p(T) = 0$.

Equivalently, $p(x)$ is a multiple of the minimal polynomial $q(x)$ for T.

This fact is actually true quite generally (think modules over commutative rings). In the context of this course, however, we're focused on only subfields of $\mathbb{C}$, where it is easier to show.

We first prove this for $T \in \mathcal{L}(V)$ where V is a finite-dimensional $\mathbb{C}$-vector space so that the minimal polynomial factors, and we'll generalize the result. The strategy of proof will be to argue that we can choose a basis for T in which the matrix $\mathcal{T}$ in this basis is upper-triangular with eigenvalues appearing on the diagonal, and then apply Exercise 9C.8.

Lemma 9C.16

Let $p(x) = (x - \lambda_1)^{\ell_1} \cdots (x - \lambda_m)^{\ell_m}$ be the minimal for the operator $T \in \mathcal{L}(V)$. Then

$$V = \mathrm{Null}(T - \lambda_1 I)^{\ell_1} \oplus \cdots \oplus \mathrm{Null}(T - \lambda_m I)^{\ell_m}.$$

Proof. The sum on the right is always a subspace of V, so we aim to show that, for every $\boldsymbol{v} \in V$, there are vectors $\boldsymbol{u_j} \in \mathrm{Null}(T - \lambda_j I)^{\ell_j}$ for which

$$\boldsymbol{v} = \boldsymbol{u_1} + \cdots + \boldsymbol{u_m}.$$

$\boldsymbol{V}$ is a sum of null spaces . We employ the same technique as in the proof of Theorem 5D.11.

Let $p(x) = (x - \lambda_1)^{\ell_1} \cdots (x - \lambda_m)^{\ell_m}$ be the minimal for T. Looking at the partial fraction decomposition of $\frac{1}{p(x)}$, there are polynomials $k_1(x), \ldots, k_m(x)$ for which

$$\frac{1}{p(x)} = \frac{1}{(x - \lambda_1)^{\ell_1} \cdots (x - \lambda_m)^{\ell_m}} = \sum_{j=1}^{m} \frac{k_j(x)}{(x - \lambda_j)^{\ell_j}}.$$

For each j, define the polynomial

$$q_j(x) = \frac{k_j(x)\, p(x)}{(x - \lambda_j)^{\ell_j}}.$$

This can be rearranged to $(x-\lambda_j)^{\ell_j} q_j(x) = k_j(x)p(x)$, and now we have that

$$(T-\lambda_j I)^{\ell_j} q_j(T)\boldsymbol{v} = k_j(T)p(T)\boldsymbol{v} = \mathbf{0}$$

(because $p(T)=0$). Taking $\boldsymbol{u_j} = q_j(T)\boldsymbol{v}$, we see that $\boldsymbol{u_j} \in \operatorname{Null}(T-\lambda_j I)^{\ell_j}$, as desired.

The sum of null spaces is a direct sum. For each j, choose $\boldsymbol{u_j} \in \operatorname{Null}(T-\lambda_j I)^{\ell_j}$ so that

$$\boldsymbol{u_1} + \cdots + \boldsymbol{u_m} = \mathbf{0}. \tag{9.2}$$

Observe that, using $q_j(x)$ as above, we have

$$q_j(x) = k_j(x)(x-\lambda_1)^{\ell_1}\cdots(x-\lambda_{j-1})^{\ell_{j-1}}(x-\lambda_{j+1})^{\ell_{j+1}}\cdots(x-\lambda_m)^{\ell_m}$$

so $q_j(T)\boldsymbol{u_k} = \mathbf{0}$ when $j \neq k$. Thus, applying $q_j(T)$ to both sides of Equation 9.2, one gets

$$\begin{aligned} q_j(T)(\boldsymbol{u_1} + \cdots + \boldsymbol{u_m}) &= q_j(T)\mathbf{0} \\ \mathbf{0} + \cdots + \mathbf{0} + \boldsymbol{u_j} + \mathbf{0} + \cdots \mathbf{0} &= \mathbf{0} \\ \implies \boldsymbol{u_j} &= \mathbf{0} \end{aligned}$$

and therefore the sum is a direct sum.

□

Before stating the next result, we first look at an example for motivation.

Example 9C.17

Let $T \in \mathcal{L}(\mathbb{C}^4)$ be the operator whose matrix (in the standard basis) is given by

$$\mathcal{M}(T) = \begin{pmatrix} 17 & 20 & 16 & 4 \\ -6 & -5 & -8 & -2 \\ -6 & -10 & -3 & -2 \\ 20 & 33 & 25 & 11 \end{pmatrix}.$$

Find a basis $\mathcal{G}$ for $\mathbb{C}^4$ so that $\mathcal{M}(T,\mathcal{G})$ is upper-triangular the eigenvalues appearing along the diagonal.

It is an exercise to check that T has a minimal polynomial of $(x-5)^3$, so only the eigenvalue $\lambda = 5$. First we find a basis for $\operatorname{Null}(T-5I)$:

$$\operatorname{Null}(T-5I) = \operatorname{Span}\big((3,-2,0,1),\ (7,-5,1,0)\big).$$

and so we take

$$\boldsymbol{g_1} = (3,-2,0,1) \qquad \text{and} \qquad \boldsymbol{g_2} = (7,-5,1,0).$$

Now we find a basis for $\operatorname{Null}(T-5I)^2$. We note, however, that $\operatorname{Null}(T-5I)^n$ is a subspace of $\operatorname{Null}(T-5I)^{n+1}$ for all n, so we want to actually *extend* the basis for $\operatorname{Null}(T-5I)$ to a basis for $\operatorname{Null}(T-5I)^2$.

$$\operatorname{Null}(T-5\,\mathrm{Id})^2 = \operatorname{Span}\big((0,0,0,1),(-1,0,2,0),(-3,2,0,0)\big).$$

and so we take

$$\boldsymbol{g_3} = (0, 0, 0, 1).$$

Lastly we extend the basis for $\operatorname{Null}(T - 5\operatorname{Id})^2$ into a basis for $\operatorname{Null}(T - 5\operatorname{Id})^3 = \mathbb{C}^4$, taking $\boldsymbol{g_4} = (1, 0, 0, 0)$.

And now in the $\mathcal{G} = \{\boldsymbol{g_1}, \boldsymbol{g_2}, \boldsymbol{g_3}, \boldsymbol{g_4}\}$ basis, it can be checked that

$$\mathcal{M}(T, \mathcal{G}) = \begin{pmatrix} 5 & 0 & 6 & 24 \\ 0 & 5 & -2 & -8 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 5 \end{pmatrix}.$$

So why is the matrix in the last example upper triangular? Here's the observation: For any vector $\boldsymbol{v} \in \operatorname{Null}(T - 5I)^2$, the vector $(T - 5I)\boldsymbol{v} \in \operatorname{Null}(T - 5I)$. So, there are scalars k_1, k_2 for which

$$(T - 5I)\boldsymbol{g_3} = k_1\boldsymbol{g_1} + k_2\boldsymbol{g_2} \qquad \implies \qquad T\boldsymbol{g_3} = 5\boldsymbol{g_3} + k_1\boldsymbol{g_1} + k_2\boldsymbol{g_3}.$$

Thus the matrix representation $\mathcal{M}(\boldsymbol{g_3}, \mathcal{G}) = \begin{pmatrix} k_1 \\ k_2 \\ 5 \\ 0 \end{pmatrix}$.

Lemma 9C.18: A generalized eigenbasis

Let $T \in \mathcal{L}(V)$. There is a basis $\mathcal{G}$ for T (a "generalized eigenbasis") in which $\mathcal{M}(T, \mathcal{G})$ is upper triangular and the diagonal entries are the eigenvalues.

Proof of Lemma 9C.18. Let $p(x) = (x - \lambda_1)^{\ell_1} \cdots (x - \lambda_m)^{\ell_m}$ be the minimal polynomial for T. From Lemma 9C.16, we have that

$$V = \operatorname{Null}(T - \lambda_1 I)^{\ell_1} \oplus \cdots \oplus \operatorname{Null}(T - \lambda_m I)^{\ell_m}$$

and so it suffices to consider each null space independently.

Find a basis for $\operatorname{Null}(T - \lambda_j I)$. For each $i = 1, \ldots, \ell_j$, extend the basis for $\operatorname{Null}(T - \lambda_j I)^i$ to a basis for $\operatorname{Null}(T - \lambda_j I)^{i+1}$. Repeat this process for each $j = 1, \ldots, m$ to achieve a basis $\boldsymbol{G} = \{\boldsymbol{g_1}, \ldots, \boldsymbol{g_n}\}$.

To see that this matrix is upper-triangular, take a basis vector $\boldsymbol{g}_d$ of $\operatorname{Null}(T - \lambda_j I)^{i+1}$ which is not contained in $\operatorname{Null}(T - \lambda_j I)^i$, we see that

$$\begin{aligned} &(T - \lambda I)\boldsymbol{g_d} \in \operatorname{Null}(T - \lambda_j)^i \\ \implies &(T - \lambda_j I)\boldsymbol{g_d} = [\text{linear combination of } \operatorname{Null}(T - \lambda_j)^i \text{ basis}] \\ \implies &T\boldsymbol{g_d} = [\text{linear combination of } \operatorname{Null}(T - \lambda_j)^i \text{ basis}] + \lambda_j \boldsymbol{g_d} \end{aligned}$$

and so column d of the matrix will be

$$\mathcal{M}(T\boldsymbol{g_d}, \mathcal{G}) = \begin{pmatrix} * \\ \vdots \\ * \\ \lambda_j \\ 0 \\ \cdots \\ 0 \end{pmatrix}$$

□

In the last proof, the number of times that the eigenvalue λ_j appears on the diagonal is always greater than or equal to ℓ_j (since you get at least one new basis vector for each $\text{Null}(T - \lambda_j I)^i$, as $i = 1, \ldots, \ell_j$).

Proof of the Cayley–Hamilton theorem. Let $T \in \mathcal{L}(V)$ be an operator on an n-dimensional vector space.

Case 1. Suppose V is a $\mathbb{C}$-vector space. Find a generalized eigenbasis $\mathcal{G}$ for V, a la Lemma 9C.18, so that the matrix $\mathcal{M}(T, \mathcal{G})$ is upper-triangular with eigenvalues on the diagonal. For simplicity, let d_j be the number of times that λ_j appears on the diagonal. As $\mathcal{M}(T)$ is upper triangular, the characteristic polynomial $p(x)$ for T is the product of the diagonal entries (see Exercise 9C.8), hence

$$p(x) = (x - \lambda_1)^{d_1} \cdots (x - \lambda_m)^{d_m}.$$

Let $q(x) = (x - \lambda_1)^{\ell_1} \cdots (x - \lambda_m)^{\ell_m}$ be the minimal polynomial for T. Our above observation yields

$$p(x) = (x - \lambda_1)^{d_1 - \ell_1} \cdots (x - \lambda_m)^{d_m - \ell_m} q(x)$$

whence

$$\begin{aligned} p(T) &= (T - \lambda_1 I)^{d_1 - \ell_1} \cdots (T - \lambda_m I)^{d_m - \ell_m} q(T) \\ &= (T - \lambda_1 I)^{d_1 - \ell_1} \cdots (T - \lambda_m I)^{d_m - \ell_m} 0 \\ &= 0 \end{aligned}$$

Case 2. Suppose V is a $\mathbb{K}$-vector space with $\mathbb{K} \neq \mathbb{C}$ and $\dim(V) = n$. Let $A = \mathcal{M}(T)$ be the matrix for T in the standard basis. Since A has entries in $\mathbb{K}$ and a standing assumption in this class is that $\mathbb{K} \subseteq \mathbb{C}$, we can let $S \in \mathcal{L}(\mathbb{C}^n)$ be the operator whose matrix in the standard basis is also A. Then we have that, for all x,

$$p(x) = \det(xI - T) = \det(xI - A) = \det(xI - S).$$

Appealing to case 1, we must have that $0 = p(S) = p(A) = p(T)$.

□

9C.III Bits and Bobs

Definition

Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional vector space, and let λ be an eigenvalue of T. The **algebraic multiplicity** of λ is the number of times that the $(\lambda - x)$ appears in the characteristic polynomial.

The **geometric multiplicity** of λ is the dimension of the eigenspace $E(\lambda, T)$.

For lack of any good notation, one may write $\mathrm{AlgMult}(\lambda)$ and $\mathrm{GeoMult}(\lambda)$ to represent the algebraic and geometric multiplicities, respectively.

Lemma 9C.19

For any operator $T \in \mathcal{L}(V)$, $1 \leq \mathrm{GeoMult}(\lambda) \leq \mathrm{AlgMult}(\lambda)$.

Sketch. That $\mathrm{GeoMult}(\lambda) \geq 1$ is immediate: λ is an eigenvalue for T if and only if $\mathrm{Null}(T - \lambda I)$ is nontrivial (i.e. has dimension greater than 0).
To see that $\mathrm{GeoMult}(\lambda) \leq \mathrm{AlgMult}(\lambda)$, let $p(x) = \cdots (x - \lambda)^{\ell} \cdots$ be the minimal polynomial for T. We observe that

$$\mathrm{GeoMult}(\lambda) = \dim E(\lambda, T) = \dim \mathrm{Null}(T - \lambda I) \leq \dim \mathrm{Null}(T - \lambda I)^{\ell} = \mathrm{AlgMult}(\lambda).$$

☐

Theorem 9C.20

An operator $T \in \mathcal{L}(V)$ is diagonalizable if and only if, for every eigenvalue λ, $\mathrm{AlgMult}(\lambda) = \mathrm{GeoMult}(\lambda)$.

Proof. This follows immediately from the work in the above lemma, and letting ℓ be as in the lemma, the fact that T is only diagonalizable if $\ell = 1$ (thm:diag-and-min-poly). ☐

Theorem 9C.21: Laplace's Cofactor Expansion

Let A be an $n \times n$ matrix and let $M_{i,j}$ denote the $(n-1) \times (n-1)$ submatrix obtained by deleting Row i and Column j from A. The determinant of an $n \times n$ matrix A can be computed *along the i^{th} row* as the sum

$$\det A = \sum_{\ell=1}^{n} (-1)^{i+\ell} A_{i,\ell} \det(M_{i,\ell})$$

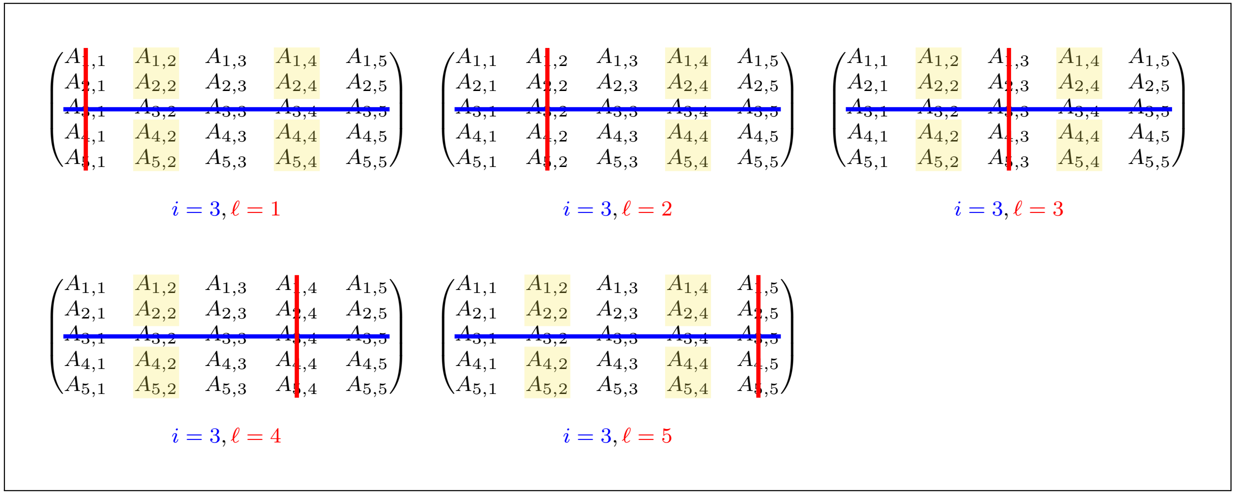
or *along the j^{th} column* as the sum

$$\det A = \sum_{\ell=1}^{n} (-1)^{\ell+j} A_{\ell,j} \det(M_{\ell,j}).$$

The quantity $(-1)^{i+j} \det(M_{i,j})$ is sometimes called the (i,j)**-cofactor** and the above sums are called **cofactor expansions**.

Proof. Since the determinant is the *unique* alternating n-linear form on the columns of A which is 1 when $A = I$, it suffices to check that the expression above has these properties.

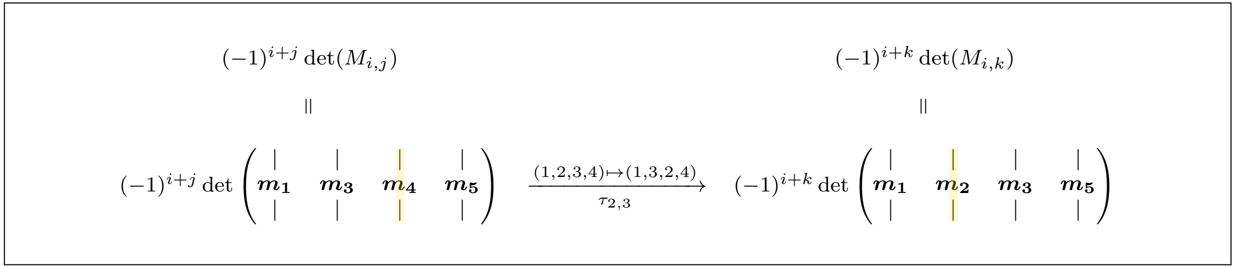
Alternating. We argue only in the case of expanding along a particular row (that $\det(A) = \det(A^t)$ implies the column expansion result holds.) We choose to expand along Row i. Suppose that A has Column j and Column k are equal (with $j \neq k$). Upon removing Row i and Column ℓ from the matrix, when $\ell \neq j$ and $\ell \neq k$ the resulting matrix $M_{i,\ell}$ still has repeated columns.



As such, we have

$$\sum_{\ell=1}^{n} A_{i,\ell}(-1)^{i+\ell}\det(M_{i,\ell}) = A_{i,j}(-1)^{i+j}\det(M_{i,j}) + A_{i,k}(-1)^{i+k}\det(M_{i,k}). \tag{9.3}$$

Now, $M_{i,j}$ and $M_{i,k}$ have the same columns, but permuted.



Let σ be the permutation sending $M_{i,j}$ to $M_{i,k}$. Then, acknowledging that $A_{i,j} = A_{i,k}$, Equation 9.3 becomes

$$\begin{aligned} & A_{i,j}(-1)^{i+j}\det(M_{i,j}) + A_{i,k}(-1)^{i+k}\det(M_{i,k}) && (9.4) \\ &= A_{i,j}(-1)^{i+j}\det(M_{i,j}) + A_{i,k}(-1)^{i+k}\operatorname{sgn}(\sigma)\det(M_{i,j}) \\ &= A_{i,j}(-1)^{i+j}\det(M_{i,j})\Big[1 + (-1)^{k-j}\operatorname{sgn}(\sigma)\Big] && (9.5) \\ & && (9.6) \end{aligned}$$

WLOG, assuming $k > j$. Observe that column $\boldsymbol{a_j}$ in A becomes column $\boldsymbol{m_j}$ in $M_{i,k}$, and column $\boldsymbol{a_k}$ in A becomes column $\boldsymbol{m_{k-1}}$ in $M_{i,j}$. As such, it must take $(k - 1 - j)$ column transpositions in order to change $M_{i,j}$ into $M_{i,k}$:

$$\sigma = \tau_{k-1,k-2} \circ \cdots \circ \tau_{j+1,j}.$$

Therefore, Equation 9.5 becomes

$$\begin{aligned}
&A_{i,j}(-1)^{i+j}\det(M_{i,j})\left[1+(-1)^{k-j}\operatorname{sgn}(\sigma)\right]\\
&= A_{i,j}(-1)^{i+j}\det(M_{i,j})\left[1+(-1)^{k-j}(-1)^{k-j-1}\right]\\
&= A_{i,j}(-1)^{i+j}\det(M_{i,j})\left[1+(-1)^{2k-2j-1}\right]\\
&= A_{i,j}(-1)^{i+j}\det(M_{i,j})\left[1+(-1)\right]=0
\end{aligned}$$

and thus the expression is alternating.

Multilinear. The multilinearity follows from the fact that it is defined as a linear combination of determinants, which are themselves multilinear.

One. Taking $A = I$, the $n \times n$ identity matrix, the only nonzero A_{ij} in the alternating sum occurs when $i = j$, and the resulting submatrix $M_{i,i}$ is the identity matrix, and $\det(M_{i,i}) = 1$.

□

Since cofactor expansion suggests that we get the freedom to choose which row or column we expand along, the reasonable thing to do is to find the row or column with the most rows possible and use that.

Example 9C.22

Use cofactor expansion to compute $\det A$, where $A = \begin{pmatrix} 3 & 0 & \pi^2 & -1 & 0\\ \zeta(2) & 4 & \sqrt{e} & 0.7 & 1\\ 0 & 0 & 1 & 0 & 0\\ 5 & 1 & \cos(1) & 5 & 0\\ -5 & 0 & \frac{1}{100} & 2 & 0 \end{pmatrix}$.

First expanding along row 3:

$$\det(A) = \det\begin{pmatrix} 3 & 0 & \pi^2 & -1 & 0\\ \zeta(2) & 4 & \sqrt{e} & 0.7 & 1\\ 0 & 0 & \text{\textcircled{1}} & 0 & 0\\ 5 & 1 & \cos(1) & 5 & 0\\ -5 & 0 & \frac{1}{100} & 2 & 0 \end{pmatrix} = (1)\det\begin{pmatrix} 3 & 0 & -1 & 0\\ \zeta(2) & 4 & 0.7 & 1\\ 5 & 1 & 5 & 0\\ -5 & 0 & 2 & 0 \end{pmatrix}$$

Now expanding along column 4:

$$\det(A) = (1)\det\begin{pmatrix} 3 & 0 & -1 & 0\\ \zeta(2) & 4 & 0.7 & \text{\textcircled{1}}\\ 5 & 1 & 5 & 0\\ -5 & 0 & 2 & 0 \end{pmatrix} = (1)(1)\det\begin{pmatrix} 3 & 0 & -1\\ 5 & 1 & 5\\ -5 & 0 & 2 \end{pmatrix}$$

Now expanding along column 2:

$$\det(A) = (1)(1)\det\begin{pmatrix} 3 & 0 & -1 \\ 5 & \textcircled{1} & 5 \\ -5 & 0 & 2 \end{pmatrix} = (1)(1)(1)\det\begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix} = (1)(1)(1)(1) = 1$$

9D Tensor Products

This section was just for funsies at the end of the course and is mostly supposed to be a survey/brief introduction to tensor products. It's going to be lacking in proofs in favor of a few explicit computations. While your instructor agrees that the approach taken by the textbook author is sufficiently aligned with his higher-level treatment of other material, in this instance, your instructor thinks that a few explicit, low-dimensional computations are a necessary supplement.

Fact. Let V, W be $\mathbb{K}$-vector spaces. Then the Cartesian product $V \times W$ is a vector space with component-wise operations. Letting $+_V$ and $+_W$ be the vector operations on V and W (respectively) and k a scalar:

$$(\boldsymbol{v_1}, \boldsymbol{w_1}) + (\boldsymbol{v_2}, \boldsymbol{w_2}) = (\boldsymbol{v_1} +_V \boldsymbol{v_2}, \boldsymbol{w_1} +_W \boldsymbol{w_2})$$

$$k(\boldsymbol{v}, \boldsymbol{w}) = (k\boldsymbol{v}, k\boldsymbol{w})$$

More generally, if $V_1, \ldots, V_m$ are $\mathbb{K}$-vector spaces, the m-fold Cartesian product $V_1 \times \cdots \times V_m$ is a $\mathbb{K}$-vector space.

Tensor products are extremely useful for a myriad of reasons, many (most?) of which are simply outside of the scope of this course. In the context of this class, maybe two compelling reasons are the following:

- Bilinear forms $\beta : V \times V \to \mathbb{K}$ are not actually linear maps, which means that we can't utilize the full power of linear algebra to work with them. Observe that,

  $$\beta(k(\boldsymbol{v_1}, \boldsymbol{v_2})) = \beta(k\boldsymbol{v_1}, k\boldsymbol{v_2}) = k^2\beta(\boldsymbol{v_1}, \boldsymbol{v_2})$$

  and other than for some very specific k or v_1, v_2-values, this is generally not equal to $k\beta(\boldsymbol{v_1}, \boldsymbol{v_2})$. With a tensor product, we are able to turn bilinear (or even multilinear) maps into honest linear maps.
- Vector spaces do not naturally come equipped with a product operation ($\mathbb{R}^3$ and the cross product is a special example, as is $\mathcal{P}(\mathbb{K})$ with the usual polynomial multiplication). The tensor product allows us to define something akin to a product operation.

Definition

Let $V_1, \ldots, V_m$ be $\mathbb{K}$-vector spaces. An ***m*-linear functional** is a map

$$\beta : V_1 \times \cdots \times V_m \to \mathbb{K}$$

where

$$x \mapsto \beta(v_1, \ldots, v_{i-1}, x, v_{i+1}, \ldots, x_m)$$

is a linear map for each i.

Recall that, for each vector space V, the dual space V' is the vector space of linear maps $\mathcal{L}(V, \mathbb{K})$.

Definition

Let $V_1, \ldots, V_m$ be $\mathbb{K}$-vector spaces. The **tensor product** of these vector spaces is

$$v_1 \otimes \cdots \otimes V_m := \{\text{multilinear functionals } \beta : V_1' \times \cdots \times V_m' \to \mathbb{K}\}.$$

Given vectors $\boldsymbol{v_i} \in V_i$, the **tensor product** of $v_1, \ldots, v_n$ is the element of $V_1 \otimes \cdots \otimes V_m$, denoted $\boldsymbol{v_1} \otimes \cdots \otimes \boldsymbol{v_m}$, so that

$$\boldsymbol{v_1} \otimes \cdots \otimes \boldsymbol{v_m}(\varphi_1, \ldots, \varphi_m) = \varphi_1(\boldsymbol{v_1})\varphi_2(\boldsymbol{v_2}) \cdots \varphi_m(\boldsymbol{v_m}).$$

for all linear functionals $\varphi_i \in V_i'$.

This seems like kind of an odd choice of definition to fix our non-linear multilinearity problem, but I think the spirit of the choice is that it "hides" the multilinearity part. For all intents and purposes, you can (should?) consider this to just be a technicality and not get bogged down by it. A "natural" choice of basis is coming shortly, and once you have that you're find to think of this as a more normal vector space.

Proposition 9D.1

Let $\boldsymbol{v_1} \otimes \cdots \otimes \boldsymbol{v_m} \in V_1 \otimes \cdots \otimes V_m$, let $\lambda \in \mathbb{K}$, and for each i let $\boldsymbol{w_i} \in V_i$. Then

$$\begin{aligned} &\boldsymbol{v_1} \otimes \cdots \otimes (\boldsymbol{v_i} + \boldsymbol{w_i}) \otimes \cdots \otimes \boldsymbol{v_m} \\ &= (\boldsymbol{v_1} \otimes \cdots \otimes \boldsymbol{v_i} \otimes \cdots \otimes \boldsymbol{v_m}) + (\boldsymbol{v_1} \otimes \cdots \otimes \boldsymbol{w_i} \otimes \cdots \otimes \boldsymbol{v_m}) \end{aligned}$$

and

$$\boldsymbol{v_1} \otimes \cdots \otimes (\lambda \boldsymbol{v_i}) \otimes \cdots \otimes \boldsymbol{v_m} = \lambda(\boldsymbol{v_1} \otimes \cdots \otimes \boldsymbol{v_i} \otimes \cdots \otimes \boldsymbol{v_m})$$

This is a straightforward proof.

Proposition 9D.2

Let $V_1, \ldots, V_m$ be $\mathbb{K}$-vector spaces. Then the tensor product space $V_1 \otimes \cdots \otimes V_m$ is a $\mathbb{K}$-vector space.

This is a straightforward proof.

Since the tensor product space is a vector space, it's natural to ask about its basis.

Theorem 9D.3

Let $V_1, \ldots, V_m$ be $\mathbb{K}$-vector spaces with bases $\mathcal{B}_1, \ldots, \mathcal{B}_m$, respectively. Then

$$\{\boldsymbol{b_1} \otimes \cdots \otimes \boldsymbol{b_m} \ : \ \boldsymbol{b_i} \in \mathcal{B}_i\}$$

is a basis for $V_1 \otimes \cdots \otimes V_m$.

Proof. Left as an exercise; but straightforward. ☐

Corollary 9D.4

$$\dim(V_1 \otimes \cdots \otimes V_m) = \prod_{j=1}^{m} \dim V_j$$

9D.I Universal Property of Tensor Products

One of the reasons that tensor products are so useful is the fact that they turn multilinear maps into linear maps (and in a unique way). This is sometimes called a *universal property*.

Definition: multilinear map

Let $V_1, \ldots, V_m, W$ be $\mathbb{K}$-vector spaces. A map

$$F : V_1 \times \cdots \times V_m \to W$$

is a **multilinear map** if it is linear in each component, that is, for each $i = 1, \ldots, m$, and each scalar $\lambda \in \mathbb{K}$ and all vectors $\boldsymbol{x_i}, \boldsymbol{y_1} \in V_i$,

$$\begin{aligned} &F(\boldsymbol{v_1}, \ldots, \lambda \boldsymbol{x_i} + \boldsymbol{y_i}, \ldots, \boldsymbol{v_m}) \\ &= \lambda F(\boldsymbol{v_1}, \ldots, \boldsymbol{x_i}, \ldots, \boldsymbol{v_m}) + F(\boldsymbol{v_1}, \ldots, \boldsymbol{y_i}, \ldots, \boldsymbol{v_m}) \end{aligned}$$

Theorem 9D.5: Universal Property of Tensor Products

For every multilinear map

$$\Gamma : V_1 \times \cdots \times V_m \to W$$

there is a *unique* linear map

$$\begin{aligned} \tilde{\Gamma} :&V_1 \otimes \cdots \otimes V_m \to W \\ \tilde{\Gamma}(\boldsymbol{v_1} \otimes \cdots \otimes \boldsymbol{v_m}) = \Gamma(\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}) \end{aligned}$$

Proof. Left as an exercise. ☐

Example 9D.6: dot product on $\mathbb{R}^2$

Consider the usual dot product on $\mathbb{R}^2$, which we write as the function

$$\begin{aligned} \Gamma :&\mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R} \\ \Gamma(\boldsymbol{v}, \boldsymbol{w}) &= \boldsymbol{v} \cdot \boldsymbol{w} = v_1 w_1 + v_2 w_2 \end{aligned}$$

Find $\mathcal{M}(\tilde{\Gamma})$ in some basis and verify that $\mathcal{M}(\tilde{\Gamma})\mathcal{M}(\boldsymbol{v} \otimes \boldsymbol{w}) = \mathcal{M}(\tilde{\Gamma}(\boldsymbol{v} \otimes \boldsymbol{w}))$.

Let $\{\boldsymbol{e_1}, \boldsymbol{e_2}\}$ be the standard basis for $\mathbb{R}^2$, and let

$$\{\boldsymbol{e_1} \otimes \boldsymbol{e_1},\ \boldsymbol{e_1} \otimes \boldsymbol{e_2},\ \boldsymbol{e_2} \otimes \boldsymbol{e_1},\ \boldsymbol{e_2} \otimes \boldsymbol{e_2}\}$$

be the standard basis for $\mathbb{R}^2 \otimes \mathbb{R}^2$. Then we have that

$$\begin{aligned} \tilde{\Gamma}(\boldsymbol{e_1} \otimes \boldsymbol{e_1}) &= \boldsymbol{e_1} \cdot \boldsymbol{e_2} = 1 & \qquad \tilde{\Gamma}(\boldsymbol{e_1} \otimes \boldsymbol{e_2}) &= \boldsymbol{e_1} \cdot \boldsymbol{e_2} = 0 \\ \tilde{\Gamma}(\boldsymbol{e_2} \otimes \boldsymbol{e_1}) &= \boldsymbol{e_2} \cdot \boldsymbol{e_1} = 0 & \qquad \tilde{\Gamma}(\boldsymbol{e_2} \otimes \boldsymbol{e_2}) &= \boldsymbol{e_2} \cdot \boldsymbol{e_2} = 1 \end{aligned}$$

In the standard (tensor product) basis, we thus have

$$\mathcal{M}(\tilde{\Gamma}) = \begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix}$$

Now, let $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^2$. We see that

$$\begin{aligned} \boldsymbol{v} \otimes \boldsymbol{w} &= (v_1 \boldsymbol{e_1} + v_2 \boldsymbol{e_2}) \otimes (w_1 \boldsymbol{e_1} + w_2 \boldsymbol{e_2}) \\ &= v_1 w_1 \boldsymbol{e_1} \otimes \boldsymbol{e_2} \, + \, v_1 w_2 \boldsymbol{e_1} \otimes \boldsymbol{e_2} \, + \, v_2 w_1 \boldsymbol{e_2} \otimes \boldsymbol{e_1} \, + \, v_2 w_2 \boldsymbol{e_2} \otimes \boldsymbol{e_2} \end{aligned}$$

and thus, in the standard (tensor product) basis, we see that

$$\mathcal{M}(\tilde{\Gamma})\mathcal{M}(\boldsymbol{v} \otimes \boldsymbol{w}) = \begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 w_1 \\ v_1 w_2 \\ v_2 w_1 \\ v_2 w_2 \end{pmatrix} = \begin{pmatrix} v_1 w_1 + v_2 w_2 \end{pmatrix} = \mathcal{M}(\boldsymbol{v} \cdot \boldsymbol{w}).$$

Example 9D.7: cross product on $\mathbb{R}^3$

Consider the usual cross product on $\mathbb{R}^3$, which we write as the function

$$\begin{aligned} \Gamma :&\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3 \\ \Gamma(\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{v} \times \boldsymbol{w} &= (v_2 w_3 - v_3 w_2, v_3 w_1 - v_1 w_3, v_1 w_2 - v_2 w_1) \end{aligned}$$

Find $\mathcal{M}(\tilde{\Gamma})$ in some basis and verify that $\mathcal{M}(\tilde{\Gamma})\mathcal{M}(\boldsymbol{v} \otimes \boldsymbol{w}) = \mathcal{M}(\tilde{\Gamma}(\boldsymbol{v} \otimes \boldsymbol{w}))$.

Let $\{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_e}\}$ be the standard basis for $\mathbb{R}^3$ and let

$$\{\boldsymbol{e_1} \otimes \boldsymbol{e_1}, \, \boldsymbol{e_1} \otimes \boldsymbol{e_2}, \, \boldsymbol{e_1} \otimes \boldsymbol{e_3}, \, \ldots, \boldsymbol{e_3} \otimes \boldsymbol{e_3}\}$$

be the standard basis for $\mathbb{R}^3 \otimes \mathbb{R}^3$. See that, in the standard (tensor product) basis,

$$\mathcal{M}(\tilde{\Gamma}) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and in the standard (tensor product) basis,

$$\mathcal{M}(\boldsymbol{v} \otimes \boldsymbol{w}) = \begin{pmatrix} v_1 w_1 & v_1 w_2 & v_1 w_3 & v_2 w_1 & v_2 w_2 & v_2 w_3 & v_3 w_1 & v_3 w_2 & v_3 w_3 \end{pmatrix}^t$$

so we have

$$\mathcal{M}(\tilde{\Gamma})\mathcal{M}(\boldsymbol{v} \otimes \boldsymbol{w}) = \begin{pmatrix} v_2 w_3 - v_3 w_2 \\ v_3 w_1 - v_1 w_3 \\ v_1 w_2 - v_2 w_1 \end{pmatrix}.$$

9D.II Tensors of mixed type

Some of the most useful tensor products involve only a single $\mathbb{K}$-vector space V and its dual space V'. Every author will have their own particular notation for this.

Definition

Let V be a $\mathbb{K}$-vector space and V' the dual space. The space of **(m,n)-tensors** (sometimes called **m-contravariant / n-covariant tensors**) is the tensor prodct space

$$\mathcal{T}^{(m,n)}(V) = \underbrace{V \otimes \cdots \otimes V}_{m} \otimes \underbrace{V' \otimes \cdots \otimes V'}_{n}$$

Theorem 9D.8

The space $\mathcal{T}^{(1,1)}(V)$ is isomorphic to $\mathcal{L}(V)$.

For every vector $\boldsymbol{v} \in V$ and covector $\varphi \in V'$, we can define a linear transformation $T_{v\otimes\varphi}$ for which

$$T_{v\otimes\varphi}(\boldsymbol{w}) = \varphi(\boldsymbol{w})\boldsymbol{v}$$

Let's see explicitly what's happening with this linear map

For simplicity, take $V = \mathbb{R}^2$ and $\mathbb{K} = \mathbb{R}$. Using $\{\boldsymbol{e_1}, \boldsymbol{e_2}\}$ and $\{\varepsilon^1, \varepsilon^2\}$ as bases for V and V', respectively, we write

$$\boldsymbol{v} = x\boldsymbol{e_1} + y\boldsymbol{e_2} \qquad \text{and} \qquad \varphi = z\varepsilon_1 + w\varepsilon_2.$$

Now, we compute

$$\begin{aligned} T_{\boldsymbol{v}\otimes\varphi}(\boldsymbol{e_1}) &= \underbrace{(z\varepsilon_1 + w\varepsilon_2)}_{\varphi}(\boldsymbol{e_1})\ \underbrace{(x\boldsymbol{e_1} + y\boldsymbol{e_2})}_{\boldsymbol{v}} \\ &= \underbrace{(z\varepsilon_1(\boldsymbol{e_1}) + w\varepsilon_2(\boldsymbol{e_2}))}_{\varphi(\boldsymbol{e_1})}\ \underbrace{(x\boldsymbol{e_1} + y\boldsymbol{e_2})}_{\boldsymbol{v}} \\ &= z\,\underbrace{(x\boldsymbol{e_1} + y\boldsymbol{e_2})}_{\boldsymbol{v}} \\ &= zx\boldsymbol{e_1} + yz\boldsymbol{e_2} \end{aligned}$$

an a similar computation yields

$$T_{\boldsymbol{v}\otimes\varphi}(\boldsymbol{e_2}) = xw\boldsymbol{e_1} + yw\boldsymbol{e_2}$$

Hence the matrix for T in the (tensor product) basis is

$$\mathcal{M}(T_{\boldsymbol{v}\otimes\varphi}) = \begin{pmatrix} xz & yz \\ xw & yw \end{pmatrix}.$$

I'll leave it to you to check that

$$\boldsymbol{v} \otimes \varphi = xz\boldsymbol{e_1} \otimes \varepsilon_1 + xw\boldsymbol{e_1} \otimes \varepsilon_2 + yz\boldsymbol{e_2} \otimes \varepsilon_1 + yw\boldsymbol{e_2} \otimes \varepsilon_2.$$

Proof of **??**. Assume V is n-dimensional and let $\mathfrak{F} : \mathcal{T}^{(1,1)}(V) \to \mathcal{L}(V)$ be the map

$$\mathfrak{F}(\boldsymbol{v} \otimes \varphi) = T_{\boldsymbol{v}\otimes\varphi}.$$

Explicitly, if $\boldsymbol{v} \otimes \varphi = \sum A_{i,j}\boldsymbol{e_i} \otimes \varepsilon_j$, then

$$\mathcal{F}\Big(\sum A_{i,j}\boldsymbol{e_i} \otimes \varepsilon_j\Big) = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \cdots & A_{n,n} \end{pmatrix}.$$

The claim is that this is an isomorphism.
Left as an exercise. ☐

Example 9D.9

Let V be a $\mathbb{K}$-vector space **evaluation map**

$$\begin{aligned} e :& \mathcal{T}^{(1,1)}(V) \to \mathbb{K} \\ e(\boldsymbol{v}, \varphi) &= \varphi(\boldsymbol{v}) \end{aligned}$$

Since there is an isomorphism between $\mathcal{T}^{(1,1)}$ and $\mathcal{L}(V)$, e can be regarded as a map $E : \mathcal{L}(V) \to \mathbb{K}$.

Find an explicit description for this map in the case that $\mathbb{K} = \mathbb{R}$ and $V = \mathbb{R}^2$ (which is sufficient; the same observation holds in general).

We take the same basis and vectors from the work above:

> For simplicity, take $V = \mathbb{R}^2$ and $\mathbb{K} = \mathbb{R}$. Using $\{\boldsymbol{e_1}, \boldsymbol{e_2}\}$ and $\{\boldsymbol{\varepsilon^1}, \boldsymbol{\varepsilon^2}\}$ as bases for V and V', respectively, we write
>
> $$\boldsymbol{v} = x\boldsymbol{e_1} + y\boldsymbol{e_2} \qquad \text{and} \qquad \varphi = z\varepsilon_1 + w\varepsilon_2.$$
>
> Then
>
> $$\boldsymbol{v} \otimes \varphi = xz\boldsymbol{e_1} \otimes \varepsilon_1 + xw\boldsymbol{e_1} \otimes \varepsilon_2 + yz\boldsymbol{e_2} \otimes \varepsilon_1 + yw\boldsymbol{e_2} \otimes \varepsilon_2.$$
>
> and
>
> $$\mathcal{M}(T_{\boldsymbol{v}\otimes\varphi}) = \begin{pmatrix} xz & yz \\ xw & yw \end{pmatrix}.$$

Now we compute $e(\boldsymbol{v} \otimes \varphi)$:

$$\begin{aligned} e(\boldsymbol{v} \otimes \varphi) &= \varphi(\boldsymbol{v}) \\ (z\varepsilon_1 + w\varepsilon_2)(x\boldsymbol{e_1} + y\boldsymbol{e_2}) \\ &= (z\varepsilon_1 + w\varepsilon_2)(x\boldsymbol{e_1} + y\boldsymbol{e_2}) \\ &= xz\varepsilon_1(\boldsymbol{e_1}) + yz\varepsilon_1(\boldsymbol{e_2}) + wx\varepsilon_2(\boldsymbol{e_1}) + wy\varepsilon_2(\boldsymbol{e_2}) \\ &= xz + 0 + 0 + wy \end{aligned}$$

The evaluation map is the sum of the diagonal entries of $\mathcal{M}(T_{\boldsymbol{v}\otimes\varphi})$.

(This is called the “trace” of the matrix, and the trace is actually the sum of the the eigenvalues.)

Index