MATH115A LECTURE NOTES

ALLEN GEHRET

Abstract. These lecture notes are largely based off of the course textbook [1] and the notes of Artem Chernikov (`math.ucla.edu/~chernikov/teaching/17S-MATH115A/index.html`), except for the sake of time I exclude/rearrange some material and modify some proofs. I recommend you refer to these notes (and the textbook) for learning the mathematical content of the course, and refer to the textbook for additional examples, pictures, and practice problems.

Note: these lecture notes are subject to revision, so the numbering of Lemmas, Theorems, etc. may change throughout the course and I do not recommend you print out too many pages beyond the section where we are in lecture. Any and all questions, comments, and corrections are enthusiastically welcome!

Contents

1. Vector spaces 2
1.1. Fields 2
1.2. Vector spaces 3
1.3. Basic properties of vector spaces 4
1.4. Subspaces 5
1.5. Linear combinations and span 7
1.6. Linear independence 9
1.7. Bases and dimension 11
2. Linear transformations 15
2.1. Basic properties of linear transformations 15
2.2. Null space and range 15
2.3. The matrix representation of a linear transformation 19
2.4. Algebraic description of the operations in $\mathcal{L}(V, W)$ 21
2.5. Composition of linear transformations and matrix multiplication 22
2.6. Calculating the value of a linear transformation using its matrix representation 25
2.7. Associating a linear transformation to a matrix 26
2.8. Invertibility 27
2.9. Isomorphisms 28
2.10. Change of coordinate matrix 30
3. Determinants 32
3.1. Computing the determinant 32
3.2. Properties of the determinant 33
3.3. The determinant of a linear operator (New!) 34
4. Eigenvalues and eigenvectors 36
4.1. Determining eigenvectors and eigenvalues of a linear operator. 39
5. Inner product spaces 44
5.1. Inner products and norms 44
5.2. Orthogonality 47

Date: February 18, 2019.

5.3. Orthonormal bases and Gram-Schmidt orthogonalization 48
5.4. Orthogonal complement 51
6. Appendix: non-linear algebra math 54
6.1. Sets 54
6.2. Set operations - making new sets from old 55
6.3. Functions 56
6.4. Induction 57
References 59

1. Vector spaces

1.1. **Fields.** You are probably familiar with "linear algebra with real scalars/coefficients" and "linear algebra with complex scalars/coefficients". To make the notion of "which scalars are we using" precise, we give the definition of a *field*:

Definition 1.1. A **field** is a set F is a set equipped with two operations $+$ and $\cdot$ (called **(scalar) addition** and **(scalar) multiplication**, respectively) and two special elements 0 and 1, such that for every $a, b, c \in F$:

(F1) $a + b = b + a$ and $a \cdot b = b \cdot a$ (commutativity of addition and multiplication),
(F2) $(a + b) + c = a + (b + c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ (associativity),
(F3) $0 + a = a$ and $1 \cdot a = a$ (additive and multiplicative identities)
(F4) there exists an element $-a \in F$ such that

$$a + (-a) = 0 \quad \text{(additive inverse)}$$

and if $b \neq 0$, then there exists an element $b^{-1} \in F$ such that

$$b \cdot b^{-1} = 1 \quad \text{(multiplicative inverse)}$$

(F5) $a \cdot (b + c) = a \cdot b + a \cdot c$ (distributivity)

For us, the following are the two most important examples:

Examples 1.2.
(1) The set $\mathbb{R}$ of **real numbers**, with the usual operations $+$ and $\cdot$, and the usual 0 and 1 is a field.
(2) The set $\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$ of **complex numbers**, with the usual $+, \cdot, 0, 1$ is also a field.
(3) The set

$$\mathbb{Q} \;=\; \left\{ \frac{m}{n} : m, n \in \mathbb{Z} \text{ and } n \neq 0 \right\}$$

of **rational numbers**, with the usual $+, \cdot, 0, 1$ is a field.
(4) Let Z_2 be a set with two elements $Z_2 = \{0, 1\}$ and define the operations of addition and multiplication as follows:

$$0 + 0 = 1 + 1 := 0 \quad \text{and} \quad 0 + 1 = 1 + 0 := 1 \quad \text{and} \quad 0 \cdot 0 = 0 \cdot 1 = 1 \cdot 0 := 0 \quad \text{and} \quad 1 \cdot 1 := 1.$$

Then Z_2 together with these two operations is a field (a field of two elements). Think of Z_2 as the field of "binary arithmetic".

In most of the class, we will be agnostic about which field we are working over. The reason is because, we could prove a theorem working the scalars coming from $\mathbb{R}$, and then prove a second nearly identical theorem using scalars from $\mathbb{C}$. However, it is easier and more efficient to prove just one theorem which works for any field we want (multiple theorems for the price of one!). Thus we have:

Convention 1.3. ***For the rest of the course*** *we let F denote a field. It is good to keep in mind the examples $F = \mathbb{R}$ or $F = \mathbb{C}$. However, we will rarely be using any special properties of $\mathbb{R}$ or $\mathbb{C}$ beyond just the field axioms, so in general F is allowed to be any field, not just these two.*

1.2. **Vector spaces.**

Definition 1.4. A **vector space** V **over** F is a set V with two operations:

- (vector addition) for every $x, y \in V$, there is an element $x + y \in V$
- (scalar multiplication) for every $a \in F$ and $x \in V$, there is an element $a \cdot x \in V$

such that the following axioms hold:

(VS1) $x + y = y + x$ for every $x, y \in V$ (commutativity of addition)
(VS2) $(x + y) + z = x + (y + z)$ for every $x, y, z \in V$ (associativity of addition)
(VS3) there is an element $0 \in V$ such that $x + 0 = x$ for every $x \in V$ (vector additive identity)
(VS4) for every $x \in V$ there is $y \in V$ such that $x + y = 0$ (vector additive inverse)
(VS5) $1 \cdot x = x$ for every $x \in V$ (where 1 is additive identity of F)
(VS6) $a \cdot (b \cdot x) = (a \cdot b) \cdot x$ for every $a, b \in F$ and $x \in V$ (associativity of scalar multiplication)
(VS7) $a \cdot (x + y) = a \cdot x + a \cdot y$ for every $a \in F$ and $x, y \in V$ (distributivity)
(VS8) $(a + b) \cdot x = a \cdot x + b \cdot x$ for every $a, b \in F$ and $x \in V$ (distributivity)

Given a vector space V over F, we will refer to elements of V as **vectors** and elements of F as **scalars**.

Example 1.5. Given a field F, and natural number $n \geq 1$, consider the set

$$F^n \;=\; \big\{(x_1, \ldots, x_n) : x_i \in F\big\},$$

the set of all n**-tuples** of elements from F. We equip F^n with the operations of *vector addition*:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) \;:=\; (x_1 + y_1, \ldots, x_n + y_n) \quad \text{for all } (x_1, \ldots, x_n), (y_1, \ldots, y_n) \in F^n$$

and *scalar multiplication*:

$$a \cdot (x_1, \ldots, x_n) \;:=\; (a \cdot x_1, \ldots, a \cdot x_n) \quad \text{for all } a \in F \text{ and } (x_1, \ldots, x_n) \in F^n.$$

It follows that F^n with these two operations is a vector space over the field F (i.e., (VS1)-(VS8) hold). As an example, if $F = \mathbb{R}$ and $n = 2$ or $n = 3$, then this gives the familiar vector spaces $\mathbb{R}^2$ (the Cartesian plane) and $\mathbb{R}^3$ (3D Euclidean space)

Example 1.6. Let F be a field, and let $P(F)$ be the set of all polynomials with coefficients in F. That is, $P(F)$ consists of all expressions of the form

$$p(x) \;=\; a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

for some $n \geq 0$, with $a_0, \ldots, a_n \in F$. If $a_i = 0$ for all $i = 0, \ldots, n$, then $p(x)$ is called the **zero polynomial**. The **degree** $\deg p(x)$ of a nonzero polynomial is the largest i such that $a_i \neq 0$ (the degree of the zero polynomial is defined to be -1). Two polynomials $p(x) = a_n x^n + \cdots + a_1 x + a_0$ and $q(x) = b_m x^m + \cdots + b_1 x + b_0$ are equal if they have the same degree and the same coefficients, i.e., $a_i = b_i$ for every i.

We equip $P(F)$ with the operation of *vector addition*: Given $p(x)$ and $q(x)$ as above, we may assume that $m = n$ (if $m < n$, then we set $b_{m+1} = \cdots = b_n = 0$, and if $n < m$, then we set $a_{n+1} = \cdots = a_m = 0$). Then we define:

$$p(x) + q(x) \;:=\; (p + q)(x) \;=\; (a_n + b_n)x^n + (a_{n-1} + b_{n-1})x^{n-1} + \cdots + (a_1 + b_1)x + (a_0 + b_0)$$

We also equip $P(F)$ with the operation of *scalar multiplication*: given $p(x)$ as above, and $c \in F$, define:

$$c \cdot p(x) \;:=\; (cp)(x) \;=\; ca_n x^n + ca_{n-1} x^n + \cdots + ca_1 x + ca_0.$$

It follows that $P(F)$ with these two operations is a vector space over F (need to check that (VS1)-(VS8) hold).

Example 1.7. Let $M_{2\times 2}(\mathbb{R})$ be the set of all 2×2 matrices with entries in $\mathbb{R}$. We define *vector addition* and *scalar multiplication* in the usual way:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} e & f \\ g & h \end{pmatrix} := \begin{pmatrix} a+e & b+f \\ c+g & d+h \end{pmatrix} \quad \text{and}$$

$$\alpha \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} := \begin{pmatrix} \alpha a & \alpha b \\ \alpha c & \alpha d \end{pmatrix}$$

for every $a, b, c, d, e, f, g, h, \alpha \in \mathbb{R}$. With these operations $M_{2\times 2}(\mathbb{R})$ is a vector space over $\mathbb{R}$. Analogously, the space $M_{m\times n}(\mathbb{R})$ of all $m \times n$ matrices with real coefficients is a vector space over $\mathbb{R}$.

Example 1.8 (The most boring vector space)**.** Let $V = \{0\}$, i.e., V contains a single element 0. The operations of vector addition and scalar multiplication are defined in the only way possible:

$$0 + 0 := 0 \quad \text{and} \quad \alpha \cdot 0 := 0 \quad \text{for every } \alpha \in F.$$

With these operations, V is a vector space over F.

1.3. **Basic properties of vector spaces.** In this subsection we derive some very basic properties of vector spaces from the axioms (VS1)-(VS8).

Cancellation Law 1.9. *Let V be a vector space over F and suppose $x, y, z \in V$. Then*

(1) if $x + z = y + z$, then $x = y$;
(2) if $z + x = z + y$, then $x = y$.

Proof. (1) By (VS4), there is an element $\tilde{z} \in V$ such that $z + \tilde{z} = 0$. Then

$$\begin{aligned} x &= x + 0 && \text{by (VS3)} \\ &= x + (z + \tilde{z}) && \text{by the choice of } \tilde{z} \\ &= (x + z) + \tilde{z} && \text{by (VS2)} \\ &= (y + z) + \tilde{z} && \text{by assumption} \\ &= y + (z + \tilde{z}) && \text{by (VS2) again} \\ &= y + 0 && \text{by choice of } \tilde{z} \\ &= y && \text{by (VS3) again.} \end{aligned}$$

Thus $x = y$.

(2) Assume $z + x = z + y$. Then $x + z = y + z$ by applying (VS1) to both sides. Thus $x = y$ by part (1). □

Corollary 1.10. *The vector 0 described in (VS3) is unique, i.e., if $0, 0' \in V$ are such that*

$$x + 0 = x \quad \textit{and} \quad x + 0' = x \quad \textit{for every } x \in V,$$

then $0 = 0'$.

Proof. Assume $0, 0' \in V$ are as above. Then for any $x \in V$ we have

$$x + 0 \;=\; x \;=\; x + 0'$$

by (VS3) for 0 first and then $0'$. By the Cancellation Law 1.9, we conclude that $0 = 0'$. □

This permits us to define the **zero vector** of a vector space V to be the unique vector 0 which satisfies (VS3).

Example 1.11. In the vector space F^n, the zero vector is $(0, 0, \ldots, 0)$ (the n-tuple of all 0's). In $M_{2\times 2}(\mathbb{R})$ the zero vector is $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. In $P(F)$, the zero vector is the zero polynomial.

Corollary 1.12. *Given $x \in V$, the vector y described in (VS4) is unique, i.e., if $y, y' \in V$ are such that $x + y = 0$ and $x + y' = 0$, then $y = y'$. This vector is called the **additive inverse of** x and is denoted by $-x$.*

Proof. With $y, y' \in V$ as above, if $x+y = 0 = x+y'$, then by the Cancellation Law 1.9, we conclude that $y = y'$. □

We also record here some more useful properties of vector spaces:

Proposition 1.13. *Let V be a vector space over a field F. For every $x \in V$ and $a \in F$ we have*

(1) $0 \cdot x = 0$ (the scalar 0 times any vector x equals the zero vector)
(2) $(-a) \cdot x = -(ax) = a \cdot (-x)$ (the first thing is the additive inverse of the scalar a times the vector x, the second thing is the additive inverse of the vector ax, the third thing is the scalar a times the additive inverse of x)
(3) $a \cdot 0 = 0$ (the scalar a times the zero vector is the zero vector).

Proof. Exercise. □

1.4. **Subspaces.**

Definition 1.14. Let V be a vector space over a field F. A subset W of V ($W \subseteq V$) is a **subspace of** V if W itself is a vector space over F, with respect to the vector addition and scalar multiplication defined on V.

That is, W satisfies (VS1)-(VS8) using $+$ and $\cdot$ defined on V.

Example 1.15. Consider the vector space $V = F^n$. The subset

$$W \;:=\; \big\{(x_1, \ldots, x_{n-1}, 0) : x_1, \ldots, x_{n-1} \in F\big\} \;\subseteq\; F^n$$

is a subspace of F^n (we'll show this soon).

Example 1.16. Consider the vector space $V = P(F)$ of all polynomials with coefficients from F. For each $n \geq 0$, define the subset $P_n(F) \subseteq P(F)$ consisting of all polynomials of degree *less than or equal to* n. We'll show $P_n(F)$ is a subspace of $P(F)$.

Example 1.17. Let V be any vector space. Then V is a subspace of V (the biggest possible subspace), and $\{0\} \subseteq V$ is a subspace of V (the smallest possible subspace) where 0 is the zero vector of V. These two subspaces are always guaranteed to be there.

Example 1.18. Let S be a non-empty set and F a field. Let $\mathcal{F}(S, F)$ denote the set of all functions from S to F. Two functions $f, g \in \mathcal{F}(S, F)$ are equal if $f(x) = g(x)$ for every $x \in S$. We can equip $\mathcal{F}(S, F)$ with the operations of *vector addition* and *scalar multiplication*: given $f, g \in \mathcal{F}(S, F)$ and $c \in F$, define $(f+g)(x) := f(x)+g(x)$ and define $(cf)(x) := c \cdot f(x)$. With these operations $\mathcal{F}(S, F)$ is a vector space over F.

In particular, $\mathcal{F}(\mathbb{R}, \mathbb{R})$ is the vector space of all real-valued functions defined on the real numbers. Let $C(\mathbb{R}) \subseteq \mathcal{F}(\mathbb{R}, \mathbb{R})$ denote the subset of all continuous functions. We'll show $C(\mathbb{R})$ is a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$.

The next lemma shows that to check that $W \subseteq V$ is a subspace of V, we need to check fewer things than all of (VS1)-(VS8):

Subspace Test 1.19. *Let V be a vector space over F, and let $W \subseteq V$ be a sub**set** of V. Then W is a sub**space** of V if and only if*

(a) $0 \in W$ *(i.e., the zero vector of* V *is in* W*),*
(b) if $x, y \in W$*, then* $x + y \in W$ *(i.e.,* W *is closed under vector addition), and*
(c) if $x \in W$ *and* $c \in F$*, then* $cx \in W$ *(i.e.,* W *is closed under scalar multiplication).*

Proof. ($\Rightarrow$) Assume that W is a subspace of V. This means that W is a vector space under the operations of addition and scalar multiplication defined on V. In particular, for every $x, y \in W$ and $c \in F$, $x + y, cx \in W$, so (b) and (c) hold. Furthermore, (VS3) holds in W, so there is $0' \in W$ such that $x + 0' = x$ for all $x \in W$. In particular, $0' + 0' = 0'$. Since $0' \in V$, we also have $0' + 0 = 0'$ (since $0 \in V$ satisfies (VS3) in V). By the Cancellation Law 1.9, we conclude that $0' = 0$, and so $0 \in W$ which implies (a).

($\Leftarrow$) Now suppose (a), (b), (c) hold. We will show that W is a subspace of V. Properties (b) and (c) tell us W has two operations defined on it ($+$ and $\cdot$ coming from V), so we need to check that (VS1)-(VS8) hold in order to conclude that W is a vector space with respect to these two operations (and hence a subspace of V). First, as V is a vector space, (VS1), (VS2), (VS5), (VS6), (VS7), (VS8) hold for all elements V. Since W is a subset of V these axioms also hold for all vectors in W. (VS3) holds by (a). It remains to check (VS4). Let $x \in W$ be arbitrary. We need to find some $y \in W$ such that $x + y = 0$. We know that in V, $(-1)x = -x$ is an additive inverse of x (see Proposition 1.13). By (c), $(-1)x \in W$, so we can take $y = (-1)x$. □

We will now use Subspace Test 1.19 to check the subspaces in the previous examples are indeed subspaces.

Example 1.20. Given $V = F^n$ and $W = \{(x_1, \ldots, x_{n-1}, 0) : x_i \in F\}$, we will check (a), (b), and (c). For (a), we note that $(0, 0, \ldots, 0) \in W$, taking $x_1 = \cdots = x_{n-1} = 0$. For (b) and (c), note that given any $(x_1, \ldots, x_{n-1}, 0), (y_1, \ldots, y_{n-1}, 0) \in W$ and $c \in F$, we have

$$(x_1, \ldots, x_{n-1}, 0) + (y_1, \ldots, y_{n-1}, 0) \;=\; (x_1 + y_1, \ldots, x_{n-1} + y_{n-1}, \underbrace{0 + 0}_{=0}) \in W$$

and

$$a \cdot (x_1, \ldots, x_{n-1}, 0) \;=\; (ax_1, \ldots, ax_{n-1}, \underbrace{a \cdot 0}_{=0}) \in W.$$

Example 1.21. Given $V = P(F)$ and $W = P_n(F)$, we will check (a), (b) and (c). For (a), the zero vector in $P(F)$ is the zero polynomial: $p(x) = a_n x^n + \cdots + a_1 x + a_0$, where $a_n = \cdots = a_0 = 0$. The degree of $p(x)$ is -1, so $p(x) \in P_n(F)$. For (b) if both $p(x)$ and $q(x)$ are in $P_n(F)$, then their degrees are at most n, so the degree of $p(x) + q(x)$ is at most n, so $p(x) + q(x) \in P_n(F)$. For (c), if $p(x) \in P_n(F)$ and $c \in F$, then $cp(x)$ either is the zero polynomial (if $c = 0$), or has the same degree as $p(x)$ (if $c \neq 0$), in either case, $cp(x) \in P_n(F)$.

Example 1.22. Given $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$ and $W = C(\mathbb{R})$. The zero vector of V is the constant zero function: $f(x) = 0$ for all $x \in \mathbb{R}$. By basic calculus, all constant functions are continuous (including the constant zero function), the sum of two continuous functions is continuous, and a scalar multiple of a continuous function is continuous. Thus (a), (b), and (c) hold for W.

The intersection of two subspaces is again a subspace. More generally:

Subspace Intersection 1.23. *Let* V *be a vector space over* F*, and suppose* $\{W_i : i \in I\}$ *is a collection of subspaces of* V*. Then the intersection*

$$W \;:=\; \bigcap_{i \in I} W_i \;=\; \{x \in V : x \in W_i \text{ for every } i \in I\}$$

is a subspace of V*.*

Proof. We will verify (a), (b), and (c) of Subspace Test 1.19 for W above. For (a), since each W_i is a subspace, we know that $0 \in W_i$ for each i, and so $0 \in W$. For (b), let $x, y \in W$. Then for each $i \in I$, $x, y \in W_i$. Since for each $i \in I$, W_i is a subspace, it follows by (b) for W_i that $x + y \in W_i$. Since $x + y \in W_i$ for each i, it follows that $x + y \in W$. Finally for (c), let $x \in W$ and $c \in F$. Then $x \in W_i$ for each $i \in I$, so $cx \in W_i$ for each $i \in I$, so $cx \in W$. □

Example 1.24. Let $V = \mathbb{R}^2$, a vector space over $\mathbb{R}$. Let $W_1 := \{(x_1, 0) : x_1 \in \mathbb{R}\}$ and $W_2 := \{(0, x_2) : x_2 \in \mathbb{R}\}$. Then W_1, W_2 are subspaces of V. The intersection $W_1 \cap W_2 = \{(0, 0)\}$ is also a subspace of $\mathbb{R}^2$ (the zero subspace). However, the union $W_1 \cup W_2$ is *not* a subspace of $\mathbb{R}^2$!

1.5. **Linear combinations and span.**

Definition 1.25. Let V be a vector space over F and $S \subseteq V$ a nonempty subset of V. A vector $v \in V$ is called a **linear combination** of vectors of S if there exists $n \in \mathbb{N}$, vectors $u_1, \ldots, u_n \in S$, and scalars $a_1, \ldots, a_n \in F$ such that

$$v \;=\; a_1u_1 + a_2u_2 + \cdots + a_nu_n.$$

In this case we also say that v is a linear combination of $u_1, \ldots, u_n$ and call $a_1, \ldots, a_n$ the **coefficients** of the linear combination.

Here is a fundamental question in linear algebra:

Question 1.26. *Given $v \in V$ and $u_1, \ldots, u_n \in V$, how does one determine whether v is a linear combination of the vectors $u_1, \ldots, u_n$?*

Answer. This reduces to solving a system of linear equations. □

Example 1.27. Let $V = P_1(\mathbb{R})$, $v = x + 5$, and $u_1 = 2x$, $u_2 = 3x - 1$. We want to know if v is a linear combination of u_1, u_2. In other words, are there $a_1, a_2 \in \mathbb{R}$ such that $v = a_1u_1 + a_2u_2$, i.e., can we find $a_1, a_2 \in \mathbb{R}$ such that

$$x + 5 \;=\; a_1(2x) + a_2(3x - 1) \;=\; (2a_1 + 3a_2)x + (-a_2)$$

By writing what it means for two polynomials to be equal, we arrive at a system of linear equations:

$$1 = 2a_1 + 3a_2$$
$$5 = -a_2$$

which we solve: $a_2 = -5$, and then $a_1 = 8$. Thus v is a linear combination of u_1 and u_2; indeed, $v = -5u_1 + 8u_2$.

The following definition allows us to create a subspace starting from a subset $S \subseteq V$:

Definition 1.28. Let $S \subseteq V$ be a nonempty subset of V. The **span of** S, denoted $\operatorname{span}(S)$ is the set of all linear combinations of vectors in S, i.e.,

$$\operatorname{span}(S) \;:=\; \{a_1u_1 + \cdots + a_nu_n : n \in \mathbb{N}, a_i \in F, u_i \in S\} \;\subseteq\; V.$$

As convention, we also define $\operatorname{span}(\emptyset) = \{0\} \subseteq V$. Note that $S \subseteq \operatorname{span}(S)$ since for every $u \in S$, $s = 1 \cdot s \in \operatorname{span}(S)$.

Example 1.29. Consider the vectors $u_1 = (1, 0, 0)$ and $u_2 = (0, 1, 0)$ in $V = \mathbb{R}^3$, and let $S := \{u_1, u_2\} \subseteq V$. Then the vectors in $\operatorname{span}(S)$ are all vectors of the form $a_1u_1 + a_2u_2$, where a_1, a_2 vary over $\mathbb{R}$. Thus

$$\operatorname{span}(S) \;=\; \{a_1(1, 0, 0) + a_2(0, 1, 0) : a_1, a_2 \in \mathbb{R}\} \;=\; \{(a_1, a_2, 0) : a_1, a_2 \in \mathbb{R}\},$$

which is a subspace of $\mathbb{R}^3$ which contains S.

Lemma 1.30. *Let V be a vector space over F, and suppose $S \subseteq V$. Then*

(1) The span of S is a subspace of V.
(2) Any subspace of V that contains S must contain $\text{span}(S)$.

In particular, $\text{span}(S)$ *is the smallest subspace of V which contains S.*

Proof. Both (1) and (2) are obvious if $S = \emptyset$, because $\text{span}(\emptyset) = \{0\}$ is always a subspace of V, and if $W \subseteq V$ is a subspace of V which contains $\emptyset$ (this last part is redundant as all sub**sets** contain $\emptyset$ automatically), then $0 \in W$ by Subspace Test 1.19(a), so W contains $\{0\} = \text{span}(\emptyset)$.

Now suppose $S \neq \emptyset$. Then there is $z \in S$. Since $0 \cdot z = 0$, we have $0 \in \text{span}(S)$. Next, suppose $x, y \in \text{span}(S)$. Then by definition of $\text{span}(S)$ you can write

$$x = a_1u_1 + \cdots + a_mu_m \quad \text{and} \quad y = b_1v_1 + \cdots + b_nv_n$$

for some $a_1, \ldots, a_m, b_1, \ldots, b_n \in F$ and $u_1, \ldots, u_m, v_1, \ldots, v_m \in S$. Then both

$$x + y = a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n \quad \text{and} \quad c \cdot x = (ca_1)u_1 + \cdots + (ca_m)u_m$$

are also linear combinations of vectors in S, for $c \in F$, and so they belong to $\text{span}(S)$. Since (a), (b) and (c) hold in Subspace Test 1.19, we conclude that $\text{span}(S)$ is a subspace of V.

Next, suppose $W \subseteq V$ is an arbitrary subspace of V which contains S. Let $w \in \text{span}(S)$. Then $w = c_1w_1 + \cdots + c_kw_k$ for some $c_1, \ldots, c_k \in F$ and $w_1, \ldots, w_k \in S$. Since $S \subseteq W$ we have $w_1, \ldots, w_k \in W$. Since W is itself a vector space over F, we have $w = c_1w_1 + \cdots + c_kw_k \in W$. Thus $\text{span}(S) \subseteq W$. □

Definition 1.31. A subset S of a vector space V **generates** (or **spans**) V if $\text{span}(S) = V$. In this case, we will also say the vectors of S generate (or span) V.

In general, finding a small (finite, if possible) generating set for a vector space V is an efficient way of describing V and simplifies working with it.

Example 1.32. For any vector space V, $\text{span}(V) = V$, so V is generated by itself.

Example 1.33. In $V = \mathbb{R}^3$, the vectors $(1,0,0), (0,1,0)$ and $(0,0,1)$ generate all of V. Indeed, given any vector $(a,b,c) \in \mathbb{R}^3$, we can express it as a linear combination of these three:

$$(a,b,c) \;=\; a(1,0,0) + b(0,1,0) + c(0,0,1).$$

Thus $(a,b,c) \in \text{span}\big((1,0,0),(0,1,0),(0,0,1)\big) = \mathbb{R}^3$.

Example 1.34. Let $V = M_{2\times 2}(\mathbb{R})$ be the vector space of all 2×2 matrices with coefficients in $\mathbb{R}$. Then we set

$$M_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad M_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad M_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad M_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

and we claim that the four vectors M_1, M_2, M_3, M_4 generate $M_{2\times 2}(\mathbb{R})$. Indeed, note that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \;=\; aM_1 + bM_2 + cM_3 + dM_4.$$

Thus $\text{span}\big(M_1, M_2, M_3, M_4\big) = M_{2\times 2}(\mathbb{R})$.

Example 1.35. Recall that $P(F)$ is the vector space of all polynomials with coefficients in F. Then the set $\{1, x, x^2, x^3, \ldots\}$ generates all of $P(F)$. Indeed, we have

$$\text{span}\big(\{1, x, x^2, \ldots\}\big) \;=\; \{a_0 + a_1x + \cdots + a_nx^n : n \in \mathbb{N} \cup \{0\}, a_i \in F\}$$

Likewise, we have $P_n(F)$ is generated by $\{1, x, x^2, \ldots, x_n\}$.

1.6. **Linear independence.** Usually there are many different subsets which are able to generate the same vector space. For instance

Example 1.36. The set $\{(1,0,0),(0,1,0),(0,0,1)\}$ can generate the entire vector space $\mathbb{R}^3$. The set

$$\big\{(1,0,0),(0,1,0),(0,0,1),(2,3,-1)\big\}$$

also generates all of $\mathbb{R}^3$. In some sense, the last vector $(2,3,-1)$ is *redundant*, at least when it comes to generating a subspace.

In general, it is natural to look for a smallest possible subset of V that generates all of V. This is because it is not very helpful to have redundant vectors at your disposal for describing the subspace a certain subset generates.

We will first look at the situation when it is possible to remove a redundant vector from a generating set and it remains a generating set.

First, note that if V is a vector space over F and $u_1,\dots,u_n \in V$, then the zero vector is *always* a linear combination of $u_1,\dots,u_n$ via the **trivial representation** (using only the scalar $0 \in F$ as every coefficient):

$$0 \;=\; 0\cdot u_1 + 0\cdot u_2 + \cdots + 0\cdot u_n$$

sometimes (but not always) it may or may not be possible to write

$$0 \;=\; a_1u_1 + \cdots + a_nu_n$$

where $a_1,\dots,a_n \in F$ and some $a_i \neq 0$. Such a linear combination of 0 is called a **non-trivial representation** of 0.

Example 1.37. In $\mathbb{R}^2$,

$$(0,0) \;=\; 2\cdot(1,2) + 5\cdot(2,1) + 3\cdot(-4,-3)$$

is a *non-trivial representation* of the zero vector $(0,0)$.

Definition 1.38. A subset S of a vector space V is called **linearly dependent** if there exists a finite number of distinct vectors $u_1,\dots,u_n \in S$ and scalars $a_1,\dots,a_n \in F$ such that at least one $a_i \neq 0$ and

$$a_1u_1 + a_2u_2 + \cdots + a_nu_n \;=\; 0,$$

or in other words, $u_1,\dots,u_n$ admit a non-trivial representation of 0.

If S is not linearly dependent, then we say that S is **linearly independent**.

We also say that the vectors $v_1,\dots,v_n$ are linearly dependent/independent if the set $\{v_1,\dots,v_n\}$ is linearly dependent/independent.

Example 1.39. Let $V = \mathbb{R}^2$. The set $S_1 = \{(0,1),(1,0)\}$ is linearly independent. Indeed, if $(0,0) = a_1(0,1) + a_2(1,0)$, then (a_1,a_2) must be a solution to the system

$$\begin{aligned} 0 &= a_1\cdot 0 + a_2\cdot 1 \\ 0 &= a_1\cdot 1 + a_2\cdot 0 \end{aligned}$$

although, we see that $a_1 = 0$ and $a_2 = 0$ is the only solution. This means that every representation of the zero vector must be the trivial representation and so S_1 is linearly independent.

The set $S_2 = \{(0,1),(1,0),(17,18)\}$ is linearly dependent. Indeed,

$$18\cdot(0,1) + 17\cdot(1,0) + (-1)\cdot(17,18) \;=\; (0,0),$$

which is a non-trivial representation of the zero vector $(0,0)$.

Example 1.40. Let V be a vector space over F.

(1) Any subset $S \subseteq V$ which contains 0 is automatically linearly dependent. Indeed, $0 = 1 \cdot 0$ is a non-trivial representation of 0 (since the coefficient 1 is a nonzero scalar)
(2) The empty set $\emptyset \subseteq V$ is linearly independent since we cannot form a non-trivial representation of zero with it's elements (since it does not have any elements.)
(3) If $S = \{u\} \subseteq V$ consists of a single non-zero vector u, then S is linearly independent. Indeed, if $\{u\}$ is linearly dependent, then $a \cdot u = 0$ for some $a \in F$ such that $a \neq 0$. Then

$$u = (a^{-1} \cdot a) \cdot u = a^{-1} \cdot (a \cdot u) = a^{-1} \cdot 0 = 0.$$

Example 1.41. In $V = M_{2\times 2}(\mathbb{R})$, the set

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

is linearly independent. Indeed, assume that

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = a_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + a_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + a_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

is an arbitrary representation of the zero vector $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \in M_{2\times 2}(\mathbb{R})$. This means that the scalars $a_1, a_2, a_3, a_4 \in \mathbb{R}$ is a solution to the system of linear equations:

$$\begin{aligned} 0 &= a_1 \cdot 1 + a_2 \cdot 0 + a_3 \cdot 0 + a_4 \cdot 0 \\ 0 &= a_1 \cdot 0 + a_2 \cdot 1 + a_3 \cdot 0 + a_4 \cdot 0 \\ 0 &= a_1 \cdot 0 + a_2 \cdot 0 + a_3 \cdot 1 + a_4 \cdot 0 \\ 0 &= a_1 \cdot 0 + a_2 \cdot 0 + a_3 \cdot 0 + a_4 \cdot 1 \end{aligned}$$

This is only possible if $a_1 = a_2 = a_3 = a_4 = 0$. In other words, the only representation of the zero vector is the trivial representation.

Example 1.42. In $V = P_n(F)$, the set $S = \{1, x, \ldots, x^n\}$ is linearly independent. Indeed, assume that $0 = a_0 \cdot 1 + a_1 \cdot x + \cdots + a_n \cdot x^n$ is a representation of the zero vector (i.e., the zero polynomials) in $P_n(F)$. This means that $a_0 = a_1 = \cdots = a_n = 0$, which implies the representation must be the trivial representation.

Lemma 1.43. *Let V be a vector space over F, and let $S_1 \subseteq S_2 \subseteq V$ be two subsets. Then*
(1) if S_1 is linearly dependent, then S_2 is linearly dependent, and
(2) if S_2 is linearly independent, then S_1 is linearly independent.

Proof. (1) Suppose S_1 is linearly dependent. Then there are distinct $u_1, \ldots, u_n \in S_1$ and scalars $a_1, \ldots, a_n \in F$ such that at least one $a_i \neq 0$, such that $a_1u_1 + \cdots + a_nu_n = 0$. Then since $u_1, \ldots, u_n \in S_2$ as well, we have that S_2 is also linearly dependent.
(2) This is the contrapositive of (1). □

The following relates the notion of span and linear independence:

Proposition 1.44. *Let S be a linearly independent subset of a vector space V, and let $v \in V$ be such that $v \notin S$. Then $S \cup \{v\}$ is linearly dependent if and only if $v \in \text{span}(S)$.*

Proof. ($\Rightarrow$) Suppose $S \cup \{v\}$ is linearly dependent. Then we can write

$$0 = a_1u_1 + \cdots + a_nu_n$$

for distinct $u_1, \ldots, u_n \in S \cup \{v\}$ and some scalars $a_1, \ldots, a_n \in F$, not all zero. Because S itself is linearly independent, it is not possible that each $u_i \in S$, for $i = 1, \ldots, n$, i.e., at least one of the u_i,

say u_1, equals v, and $a_1 \neq 0$. Thus $a_1v + a_2u_2 + \cdots + a_nu_n = 0$. Dividing by a_1 and solving for v yields

$$v \;=\; \frac{1}{a_1}(-a_2u_2 - \cdots - a_nu_n) \;=\; \big(-\frac{a_2}{a_1}\big)u_2 + \cdots + \big(-\frac{a_n}{a_1}\big)u_n.$$

This shows that v is a linear combination of $u_2, \ldots, u_n \in S$, and so $v \in \text{span}(S)$.

($\Leftarrow$) Suppose $v \in \text{span}(S)$. Then we have $v = b_1v_1 + \cdots b_mv_m$ for some vectors distinct $v_1, \ldots, v_m \in S$ and some scalars $b_1, \ldots, b_m \in F$. Hence

$$0 \;=\; b_1v_1 + \cdots + b_mv_m + (-1)v.$$

This is a nontrivial representation of the zero vector by the distinct vectors $v_1, \ldots, v_m, v$. Thus $\{v_1, \ldots, v_m, v\}$ are linearly dependent, hence $S \cup \{v\}$ is also linearly dependent by Lemma 1.43(1). □

1.7. **Bases and dimension.**

Definition 1.45. Let V be a vector space over F. A **basis** for a vector space V is a linearly independent subset of V that generates V. If β is a basis for V, we also say that the vectors of β form a basis for V.

Example 1.46. Recall that $\text{span}(\emptyset) = \{0\}$ and $\emptyset$ is linearly indepenent. Thus the empty set $\emptyset$ is a basis for the zero vector space.

Example 1.47. In $V = F^n$, consider the vectors

$$e_1 := (1, 0, 0, \ldots, 0), \quad e_2 := (0, 1, 0, 0, \ldots, 0), \quad \ldots, \quad e_n := (0, 0, \ldots, 0, 1).$$

Then $\beta = \{e_1, e_2, \ldots, e_n\}$ is a basis for F^n and is called the **standard basis** for F^n.

Example 1.48. In $V = M_{m\times n}(F)$, let E^{ij} denote the matrix whose only nonzero entry is a 1 in the ith row and jth column. Then the set

$$\beta \;:=\; \{E^{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$$

is a basis for $M_{m\times n}(F)$.

Example 1.49. In $V = P_n(F)$, the set $\{1, x, x^2, \ldots, x^n\}$ is a basis. We call this basis the **standard basis** for $P_n(F)$.

Example 1.50. In $V = P(F)$, the set $\{1, x, x^2, \ldots\}$ is a basis. This shows in particular that a basis need not be finite.

The next proposition expresses a very important property of bases. Namely that every vector can be expressed in a unique way as a linear combination of elements from a basis. This property shows that bases are the building blocks of vector spaces in the sense that they allow us to efficiently "parameterize" the entire space.

Proposition 1.51. *Let V be a vector space over F and $\beta = \{u_1, \ldots, u_n\}$ be a subset of V. Then the following are equivalent:*

(1) β is a basis for V.
(2) Every vector $v \in V$ can be uniquely expressed as a linear combination of vectors in β, i.e., can be expressed in the form $v = a_1u_1 + \cdots + a_nu_n$ for unique scalars $a_1, \ldots, a_n \in F$.

Proof. (1)⇒(2) Suppose β is a basis of V, and let $v \in V$ be arbitrary. Since β spans V, we have $v \in \text{span}(\beta)$ and so there are coefficients $a_1, \ldots, a_n \in F$ such that $v = a_1u_1 + \cdots + a_nu_n$. Suppose $b_1, \ldots, b_n \in F$ are another collection of coefficients such that $v = b_1u_1 + \cdots + b_nu_n$. Subtracting the second linear combination from the first linear combination yields

$$v - v \;=\; 0 \;=\; (a_1 - b_1)u_1 + \cdots + (a_n - b_n)u_n,$$

which is a representation of the zero vector using the vectors $u_1, \ldots, u_n$. Since these vectors are also linearly independent, this must be the zero representation and so the coefficients must all be $= 0$. In other words, $a_i - b_i = 0$ for all $i = 1, \ldots, n$, or rather, $a_i = b_i$ for all $i = 1, \ldots, n$. To summarize, there is only one way to express v as a linear combination of $u_1, \ldots, u_n$.

(2)⇒(1) Suppose every $v \in V$ can be uniquely expressed as a linear combination of $u_1, \ldots, u_n$. Then in particular $\text{span}(\beta) = V$ (this is true even without "uniquely"). It remains to check that β is linearly independent. Assume $0 = a_1u_1 + \cdots + a_nu_n$ for certain coefficients $a_1, \ldots, a_n \in F$. We also know that $0 = 0u_1 + \cdots + 0u_n$. Since by assumption there is only one way to express 0 as a linear combination of the vectors from β, it must be the case that $a_1 = a_2 = \cdots = a_n = 0$. We conclude that β is also linearly independent, hence a basis of V. □

The following provides a method for producing a basis for a vector space:

Basis Existence 1.52. *If V is a vector space generated by a* finite *set $S \subseteq V$, then some subset of S is a basis for V. In particular, V has a finite basis.*

Proof. First, if $S = \emptyset$, or $S = \{0\}$, then $V = \text{span}(S) = \{0\}$, and so $\emptyset$ is a subset of S which is a basis for V.

Otherwise, suppose that S contains a vector $v \neq 0$. By a previous example, the set $\{v\}$ is linearly independent. Now, consider all finitely many subsets of S, and pick one that is linearly independent and has largest possible size; say, $\beta = \{u_1, \ldots, u_n\} \subseteq S$ is linearly independent. Since $\{v\}$ is a linearly independent set of size 1, we know that $\beta \neq \emptyset$, i.e., $n \geq 1$. We claim that β is a basis for V. By choice of β, we know that β is linearly independent. By Lemma 1.30(2), we must show that $S \subseteq \text{span}(\beta)$ (because if the subspace $\text{span}(\beta)$ contains S, then it contains $\text{span}(S) = V$, so $\text{span}(\beta) = V$). Let $v \in S$. If $v \in \beta$, then $v \in \text{span}(\beta)$. Otherwise, if $v \notin \beta$, then $\beta \cup \{v\}$ is linearly dependent (because β is a maximal linearly independent subset of S), so $v \in \text{span}(\beta)$ by Proposition 1.44. Thus $S \subseteq \text{span}(\beta)$. □

A consequence of the above lemma is that every finite generating set can be reduced to a basis for V by removing the "redundant" vectors. Vector spaces that are not generated by a finite subset still have a basis (assuming the *Axiom of Choice*), but this is more complicated and we will not discuss this in this class since our main focus is on finitely generated vector spaces.

The following useful and technical result says that a linearly independent set L is always as most as big as a generating set G, and that L can be "completed" to a generating set in an efficient way using vectors from G.

Replacement Lemma 1.53. *Let V be a vector space generated by a subset $G \subseteq V$ with $|G| = n$, and let $L \subseteq V$ be a linearly independent set such that $|L| = m$. Then:*

(1) $m \leq n$, and
(2) there exists $H \subseteq G$ with $|H| = n - m$ such that $L \cup H$ generates V.

Proof. We will prove this by induction on $m \in \mathbb{N} \cup \{0\}$. For the base case $m = 0$, necessarily $L = \emptyset$, $m = 0 \leq n$, and $H := G$ has the property $|H| = n - m = n - 0$ and $L \cup H = G$ generates V.

Next, suppose that the result is true for a certain $m \geq 0$. We will prove it for $m + 1$. Let $L = \{v_1, \ldots, v_{m+1}\} \subseteq V$ be linearly independent, so $|L| = m + 1$. By Lemma 1.43, the set $\{v_1, \ldots, v_m\}$ is also linearly independent. By the inductive hypothesis, we have $m \leq n$ and there is a subset $\{u_1, \ldots, u_{n-m}\} \subseteq G$ such that $\{v_1, \ldots, v_m\} \cup \{u_1, \ldots, u_{n-m}\}$ generates V. Since $v_{m+1} \in V$, this means there are scalars $a_1, \ldots, a_m, b_1, \ldots, b_{n-m} \in F$ such that

$$(*) \qquad a_1v_1 + \cdots + a_mv_m + b_1u_1 + \cdots + b_{n-m}u_{n-m} \;=\; v_{m+1}$$

Note that in the above expression, we have $n - m > 0$ (so $n \geq m + 1$), and also $b_i \neq 0$ for some $i \in \{1, \ldots, n - m\}$, for otherwise we would be able to write v_{m+1} as a linear combination of

$v_1, \ldots, v_m$, contradicting the assumption that $v_1, \ldots, v_{m+1}$ is linearly independent. By rearranging the b_i's and u_i's, we may assume that $b_1 \neq 0$. Thus we can solve for u_1 in $(*)$:

$$(**) \quad u_1 = \left(-\frac{a_1}{b_1}\right) v_1 + \cdots + \left(-\frac{a_m}{b_1}\right) v_m + \left(\frac{1}{b_1}\right) v_{m+1} + \left(-\frac{b_2}{b_1}\right) u_2 + \cdots + \left(-\frac{b_{n-m}}{b_1}\right) u_{n-m}$$

Now let $H = \{u_2, \ldots, u_{n-m}\}$, so $|H| = (n-m) - 2 + 1 = n - (m+1)$. By $(**)$ we have $u_1 \in \text{span}(L \cup H)$, so $\{v_1, \ldots, v_m, u_1, \ldots, u_{n-m}\} \subseteq \text{span}(L \cup H)$. As $\{v_1, \ldots, v_m, u_1, \ldots, u_{n-m}\}$ generates V, $\text{span}(L \cup H) = V$ by Lemma 1.30 (because $\text{span}(L \cup H)$ is a subspace which contains $\{v_1, \ldots, v_m, u_1, \ldots, u_{n-m}\}$, so it must contain $\text{span}(v_1, \ldots, v_m, u_1, \ldots, u_{n-m}) = V$). Thus the lemma is true for $m+1$. □

For vector spaces with a finite basis, every basis has the same number of elements:

Corollary 1.54. *Let V be a vector space with a finite basis. Then there is a number $n \in \mathbb{N} \cup \{0\}$ such that for every basis β of V, $|\beta| = n$.*

Proof. By assumption, there is a finite basis γ of V. Define $n := |\gamma|$. Suppose β is an arbitrary basis of V. Recall that a basis of V is both linearly independent and generates V. Then since β is linearly independent and γ generates V, we have $|\beta| \leq |\gamma| = n$ by the Replacement Lemma 1.53(1). Likewise, since γ is linearly independent and β generates V, we also have $n = |\gamma| \leq |\beta|$. We conclude that $|\beta| = n$. □

Definition 1.55. Given a vector space V, if V has a finite basis then we say that V is **finite-dimensional** and we define $\dim(V)$ to be the unique number n from Corollary 1.54, i.e.,

$$\dim(V) = \text{size of any basis of } V.$$

If V does not have a finite basis, then we say that V is **infinite-dimensional**, and we write $\dim(V) = \infty$.

Example 1.56. (1) $\dim(\{0\}) = 0$, $\emptyset$ is a basis, and $|\emptyset| = 0$,
(2) $\dim(F^n) = n$, the standard basis $\{e_1, \ldots, e_n\}$ contains n vectors,
(3) $\dim(M_{m\times n}) = mn$, the basis $\{E^{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$ has mn vectors,
(4) $\dim(P_n(F)) = n+1$, the set $\{1, x, \ldots, x^n\}$ is a basis and has $n+1$ vectors in it.

Example 1.57. The vector space $P(F)$ has an infinite basis $\{1, x, x^2, \ldots\}$. It follows from the Replacement Lemma 1.53 that $P(F)$ cannot also have a finite basis. Indeed, assume towards a contradiction that β is a finite basis of $P(F)$, say $|\beta| = n$. Then the set $L = \{1, x, \ldots, x^n\}$ is linearly independent and $|L| = n+1$. The Replacement Lemma then implies that $|L| = n+1 \leq |\beta| = n$ (since β generates $P(F)$), a contradiction.

We conclude that $P(F)$ is infinite-dimensional.

Corollary 1.58. *Let V be a vector space of dimension n. Then*
(a) Every linearly independent subset of V with n elements is a basis for V.
(b) Every linearly independent subset of V can be extended to a basis for V.

Proof. Let β be a basis for V. In particular, $|\beta| = n$. Then for (a) we suppose $L \subseteq V$ is linearly independent with $|L| = n$. Then by the Replacement Lemma 1.53, there is $H \subseteq \beta$ such that $|H| = n - n - 0$ and $L \cup H = L$ generates V. Thus $H = \emptyset$ and L generates V, so L is a basis for V.

For (b), suppose L is a linearly independent subset of V such that $|L| = m$. By the Replacement Theorem there is $H \subseteq \beta$ such that $|H| = n - m$ and $L \cup H$ generates V. Note that $|L \cup H| = n$, and by Basis Existence 1.52, there is $S \subseteq L \cup H$ which is a basis of V. Since $|S| = n$ by Corollary 1.54, we must have $S = L \cup H$, i.e., $L \cup H$ is a basis for V. □

Dimension Monotonicity 1.59. *Let W be a subspace of a vector space V with $\dim(V) < \infty$. Then $\dim(W) \leq \dim(V)$. Moreover, if $\dim(W) = \dim(V)$, then $V = W$.*

Proof. Let β be a basis for W. Then since β is a linearly independent subset of W, hence also V, it can be extended to a basis γ of V by Corollary 1.58. In particular, $\dim(W) = |\beta| \leq \dim(V) = |\gamma|$. In the case where $\dim(W) = \dim(V)$, then necessarily $\beta = \gamma$, and so β is a basis for V as well. Thus $W = \text{span}(\beta) = V$, since β is a basis for both V and W. □

Corollary 1.60. *Suppose W is a subspace of a vector space V, with $\dim(V) < \infty$. Then any basis for W can be extended to a basis for V.*

Proof. If $S \subseteq W$ is a basis for W, it is a linearly independent subset of W and V, hence can be extended to a basis for V. □

2. Linear transformations

The primary objects of study in linear algebra are *linear transformations*. Here is the definition:

Definition 2.1. Let V and W be vector spaces over F. A function $T : V \to W$ is a **linear transformation from V to W** if, for every $x, y \in V$ and $c \in F$:

(1) $T(x + y) = T(x) + T(y)$
(2) $T(cx) = cT(y)$

2.1. Basic properties of linear transformations.

Lemma 2.2. *Let $T : V \to W$ be a linear transformation. Then*

(1) $T(0) = 0$,
(2) $T(cx + y) = cT(x) + T(y)$ for all $x, y \in V$ and $c \in F$ (in fact, this holds if and only if T is linear),
(3) $T(x - y) = T(x) - T(y)$,
(4) $T(\sum_{i=1}^n a_i x_i) = \sum_{i=1}^n a_i T(x_i)$, for all $x_1, \ldots, x_n \in V$ and $a_1, \ldots, a_n \in F$.

Example 2.3. Define $T : M_{m\times n}(F) \to M_{n\times m}(F)$ by $T(A) = A^t$, where A^t is the transpose of A. Then T is a linear transformation.

Example 2.4. For $n \geq 1$, let $T : P_n(\mathbb{R}) \to P_{n-1}(\mathbb{R})$ be defined by $T(f(x)) = f'(x)$, where $f'(x)$ denotes the derivative of $f(x)$. We will show that T is linear: let $g(x), h(x) \in P_n(\mathbb{R})$ and $a \in \mathbb{R}$ be arbitrary. Then

$$T(ag(x) + h(x)) \;=\; (ag(x) + h(x))' \;=\; ag'(x) + h'(x) \;=\; a \cdot T(g(x)) + T(h(x)).$$

Example 2.5. Let $V = C(\mathbb{R})$, the vector space of continuous real-valued functions on $\mathbb{R}$. Let $a, b \in \mathbb{R}$ be two fixed real numbers such that $a < b$. Define $T : V \to \mathbb{R}$ (where $\mathbb{R}$ is the vector space $\mathbb{R}^1$ over $\mathbb{R}$) by:

$$T(f(x)) \;:=\; \int_a^b f(t)dt$$

for all functions $f \in V$ (this definition uses the fact that continuous functions are integrable, something which is proved in math131a). Then T is linear:

$$T(ag(t) + h(t)) \;=\; \int_a^b (ag(t) + h(t))dt \;=\; a\int_a^b g(t)dt + \int_a^b h(t)dt \;=\; a \cdot T(g) + T(h).$$

2.2. Null space and range.

Definition 2.6. Let V and W be vector spaces over F, and $T : V \to W$ be linear.

(1) Let $N(T) := \{x \in V : T(x) = 0\}$, the **null space** (or **kernel**) of T,
(2) Let $R(T) := \{T(x) : x \in V\}$, the **range** (or **image**) of T.

Example 2.7. Let V and W be vector spaces over F. Then

(1) We define $I : V \to V$ by $I(x) = x$ for all $x \in V$. We call I the **identity transformation**. I is a linear transformation, $N(I) = \{0\}$ and $R(I) = V$.
(2) We define $T_0 : V \to W$ by $T_0(x) = 0$ for all $x \in V$. We call T_0 the **zero transformation**. Then T_0 is linear, $N(T_0) = V$ and $R(T_0) = \{0\}$.

Proposition 2.8. *Let V, W be vector spaces over F and suppose $T : V \to W$ is a linear transformation. Then*

(1) $N(T)$ is a subspace of V, and
(2) $R(T)$ is a subspace of W.

Proof. (1) We will show that $N(T) \subseteq V$ is a subspace of V. First, note that $T(0) = 0$, and so $0 \in N(T)$. Next, let $x, y \in N(T)$. Then $T(x+y) = T(x) + T(y) = 0 + 0 = 0$. Thus $x + y \in N(T)$. Finally, suppose $x, \in N(T)$ and $a \in F$. Then $T(ax) = a \cdot T(x) = a \cdot 0 = 0$. Thus $ax \in N(T)$. We conclude that $N(T)$ is a subspace of V.

(2) We will show that $R(T) \subseteq W$ is a subspace of W. First, note that $T(0) \in R(T)$ and $T(0) = 0$. Thus $0 \in R(T)$. Next, suppose $w_1, w_2 \in R(T)$. Then there exists $v_1, v_2 \in V$ such that $T(v_1) = w_1$ and $T(v_2) = w_2$. Then $T(v_1 + v_2) = T(v_1) + T(v_2) = w_1 + w_2$. Thus $w_1 + w_2 \in R(T)$. Finally, suppose $c \in F$ and $w \in R(T)$. Then there is $v \in V$ such that $T(v) = w$. Then $T(cv) = cT(v) = cw$, and so $cw \in R(T)$. We conclude that $R(T)$ is a subspace of W. □

Proposition 2.9. *Let V, W be vector spaces over F and assume $T : V \to W$ is a linear transformation. If $\beta = \{v_1, \ldots, v_n\}$ is a basis for V, then*

$$R(T) \;=\; \text{span}\{T(\beta)\} \;=\; \text{span}\big(\{T(v_1), \ldots, T(v_n)\}\big)$$

Proof. Clearly $T(v_i) \in R(T)$ for each $i = 1, \ldots, n$. As $R(T)$ is a subspace of W,

$$\text{span}\big(\{T(v_1), \ldots, T(v_n)\}\big) \subseteq R(T)$$

by Lemma 1.30(2). Now, suppose $w \in R(T)$, then $w = T(v)$ for some $v \in V$. Since β is a basis for V, we have scalars $a_1, \ldots, a_n \in F$ such that $v = \sum_{i=1}^{n} a_i v_i$. Since T is linear,

$$w \;=\; T(v) \;=\; \sum_{i=1}^{n} a_i T(v_i) \;\in\; \text{span}\big(\{T(v_1), \ldots, T(v_n)\}\big) \;=\; \text{span}(T(\beta)).$$

Thus $R(T) \subseteq \text{span}(T(\beta))$. □

Definition 2.10. Let V, W be vector spaces over F and $T : V \to W$ a linear transformation. If $N(T)$ and $R(T)$ are finite-dimensional then we define

$$\begin{aligned} \text{nullity}(T) &:= \dim\big(N(T)\big) \\ \text{rank}(T) &:= \dim\big(R(T)\big). \end{aligned}$$

Intuitively, if $N(T)$ is very large (i.e., T sends many vectors from V to 0), then $R(T)$ should be small (not too many vectors in W are obtained by being sent to V by T), and vice-versa. The following important theorem makes this precise:

Dimension Theorem 2.11. *Let $T : V \to W$ be a linear transformation. If $\dim(V) < \infty$, then*

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

Proof. Suppose that $\dim(V) = n$, $\dim(N(T)) = k$, and $\{v_1, \ldots, v_k\}$ is a basis for $N(T)$. By Corollary 1.60, we can extend $\{v_1, \ldots, v_k\}$ to a basis $\beta = \{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$ of V.

Claim. *$S := \{T(v_{k+1}), \ldots, T(v_n)\}$ is a basis for $R(T)$.*

Proof. First, we will show that S generates $R(T)$. As $T(v_i) = 0$ for $i = 1, \ldots, k$, by Proposition 2.9, we have

$$R(T) \;=\; \text{span}\big(\{\underbrace{T(v_1), \ldots, T(v_k)}_{=0}, T(v_{k+1}), \ldots, T(v_n)\}\big) \;=\; \text{span}\big(T(v_{k+1}), \ldots, T(v_n)\big).$$

Second, we will show that S is linearly indpendent. Suppose $\sum_{i=k+1}^{n} b_i T(v_i) = 0$ for some scalars $b_{k+1}, \ldots, b_n \in F$. Since T is linearly, we have $T(\sum_{i=k+1}^{n} b_i v_i) = 0$, so $\sum_{i=k+1}^{n} b_i v_i \in N(T)$. Thus,

there exists $c_1, \ldots, c_k \in F$ such that $\sum_{i=1}^{k} c_i v_i = \sum_{i=k+1}^{n} b_i v_i$, or rather,

$$\sum_{i=1}^{n} (-c_i) v_i + \sum_{i=k+1}^{n} b_i v_i \;=\; 0.$$

Since β is a basis for V, this implies that every $c_i = 0$ and every $b_i = 0$. This implies that S is linearly independent. □

Now counting the size of the relevant spaces gives

$$\begin{aligned} \dim(V) &= \#\{v_1, \ldots, v_n\} = n \\ \dim N(T) &= \#\{v_1, \ldots, v_k\} = k \\ \dim R(T) &= \#\{T(v_{k+1}), \ldots, T(v_n)\} = n - (k+1) + 1 = n - k. \end{aligned}$$

Thus the formula in the statement of the theorem is true. □

Example 2.12. (1) Let $T : F^n \to F^{n-1}$ be defined by $T((a_0, \ldots, a_n)) := (a_1, \ldots, a_{n-1})$ (so T "forgets" the nth component). Then T is a linear transformation,

$$N(T) = \big\{(\underbrace{0, \ldots, 0}_{n-1}, a_n) : a_n \in F\big\} \quad \text{and} \quad R(T) = F^{n-1}.$$

In this situation, $\dim(F^n) = n$, $\dim(N(T)) = 1$, and $\dim(R(T)) = \dim(F^{n-1}) = n - 1$. Note that these dimensions add up correctly as predicted by the Dimension Theorem 2.11.

(2) Let $T : P_n(\mathbb{R}) \to P_{n-1}(\mathbb{R})$ be the differentiation linear transformation, i.e., $T(p(x)) = p'(x)$ for every polynomial $p(x) \in P_n(\mathbb{R})$. Then $T(p(x)) = 0$ iff $p'(x) = 0$ iff $p(x)$ is a constant polynomial (i.e., has degree 0 or -1). Thus

$$N(T) \;=\; \{\text{constant polynomials in } P(\mathbb{R})\}.$$

Recall that $\{1, x, \ldots, x^{n-1}\}$ is a basis for $P_{n-1}(\mathbb{R})$. Since $1 = T(x), x = \frac{1}{2}T(x^2), \ldots, x^{n-1} = \frac{1}{n}T(x^n)$, it follows that $P_{n-1}(\mathbb{R}) = \operatorname{span}(\{T(x), \ldots, T(x^n)\}) = R(T)$. Thus $\dim(P_n(\mathbb{R})) = n + 1$, $\dim(R(T)) = n$ and $\dim(N(T)) = 1$, and we recognize that the dimensions add up correctly in this example as well.

Definition 2.13. Let $T : V \to W$ be a linear transformation. We say that

(1) T is **one-to-one** (or **injective**) if for all $u, v \in V$, if $T(u) = T(v)$, then $u = v$,
(2) T is **onto** (or **surjective**) if for every $w \in W$ there exists some $v \in V$ such that $T(v) = w$, and
(3) T is **bijective** if T is both one-to-one and onto.

The null space allows us to detect the "*one-to-one*ness" of a linear transformation:

One-to-one Criterion 2.14. *Let $T : V \to W$ be a linear transformation. Then T is one-to-one if and only if $N(T) = \{0\}$.*

Proof. ($\Rightarrow$) Suppose T is one-to-one, and let $x \in N(T)$. Then $T(x) = 0 = T(0)$. By the definition of one-to-one, this implies that $x = 0 \in \{0\}$. Thus $N(T) = \{0\}$.

($\Leftarrow$) Suppose $N(T) = \{0\}$. We want to show that T is one-to-one. Let $u, v \in V$ be such that $T(u) = T(v) = 0$. Consider the vector $u - v \in V$. Applying T and using linearity we get $T(u - v) = T(u) - T(v) = 0 - 0 = 0$, and thus $u - v \in N(T)$. By assumption $N(T) = \{0\}$ and thus $u - v = 0$, i.e., $u = v$. We conclude that T is one-to-one. □

Proposition 2.15. *Let $T : V \to W$ be a linear transformation, and suppose $\dim(V) = \dim(W) < \infty$. Then the following are equivalent:*

(1) T is one-to-one,

(2) T is onto,
(3) T is bijective (i.e., one-to-one and onto),
(4) $\dim(R(T)) = \dim(V)$.

Proof. Recall that by the Dimension Theorem 2.11, we have $\dim(N(T)) + \dim(R(T)) = \dim(V)$. Now note that

$$\begin{aligned} T \text{ is one-to-one} &\iff N(T) = 0, \quad \text{by One-to-one Criterion 2.14} \\ &\iff \dim(N(T)) = 0 \\ &\iff \dim(R(T)) = \dim(V) \quad \text{by the Dimension Theorem 2.11} \\ &\iff \dim(R(T)) = \dim(W) \quad \text{by assumption that } \dim(V) = \dim(W) \\ &\iff R(T) = W \quad \text{by Dimension Monotonicity 1.59} \\ &\iff T \text{ is onto.} \end{aligned}$$

□

Example 2.16. (1) Define $T : F^2 \to F^2$ by $T\big((a_1, a_2)\big) = (a_1 + a_2, a_1)$. Then $N(T) = \big\{(0, 0)\big\}$, so T is one-to-one. By Proposition 2.15, T is also onto.

(2) Define $T : P_n(\mathbb{R}) \to \mathbb{R}^{n+1}$ by $T(a_0 + a_1x + \cdots + a_nx^n) = (a_0, a_1, \ldots, a_n)$. Then T is a linear transformation and is one-to-one. Thus T is onto, since $\dim(P_n(\mathbb{R})) = \dim(\mathbb{R}^{n+1})$.

The following shows that a linear transformation is completely determined by the values it takes on a basis, and conversely, we can define a linear transformation by prescribing values that it takes on a basis (and we have complete freedom when doing so).

Linear Transformation Prescription 2.17. *Let V, W be vector spaces over a field F, and let $\{v_1, \ldots, v_n\}$ be a basis for V. Then for any $w_1, \ldots, w_n \in W$, there exists exactly one linear transformation $T : V \to W$ such that $T(v_i) = w_i$ for $i = 1, \ldots, n$.*

Proof. Let $x \in V$. Then $x = \sum_{i=1}^{n} a_i v_i$ for unique scalars $a_1, \ldots, a_n \in F$ (by Proposition 1.51, since $\{v_1, \ldots, v_n\}$ is a basis). Then we define a function $T : V \to W$ by

$$T(x) \;:=\; \sum_{i=1}^{n} a_i w_i$$

[T is well-defined because the scalars a_i are unique] We have a few things to show:

(a) T is a linear transformation. To show this, let $u, v \in V$ and $d \in F$. Then we can write

$$u \;=\; \sum_{i=1}^{n} b_i v_i \quad \text{and} \quad v \;=\; \sum_{i=1}^{n} c_i v_i$$

for some scalars $b_1, \ldots, b_n, c_1, \ldots, c_n \in F$. Then

$$du + v \;=\; \sum_{i=1}^{n} (db_i + c_i) v_i$$

and in particular, the scalars $db_i + c_i$ are the unique scalars used to represent $du + v$ in terms of the basis $\{v_1, \ldots, v_n\}$. Applying T to $du + v$ gives

$$T(du + v) \;=\; \sum_{i=1}^{n} (db_i + c_i) w_i \;=\; d \sum_{i=1}^{n} b_i w_i + \sum_{i=1}^{n} c_i w_i \;=\; dT(u) + T(v).$$

(b) It is clear that $T(v_i) = w_i$ for $i = 1, \ldots, n$, since $v_j = \sum_{i=1}^{n} a i v_i$ where $a_j = 1$ and $a_i = 0$ if $i \neq j$.

(c) T is unique. To prove this, assume that $U : V \to W$ is a linear transformation with the property that $U(v_i) = w_i$ for $i = 1, \ldots, n$. Then, for $x \in V$ with $x = \sum_{i=1}^{n} a_i v_i$ we have (using that U is linear)

$$U(x) \;=\; \sum_{i=1}^{n} a_i U(v_i) \;=\; \sum_{i=1}^{n} a_i w_i \;=\; T(x).$$

Thus $U = T$. □

As suggested by the proof, this gives us a way of checking if two linear transformations are actually the same:

Corollary 2.18. *Let V, W be vector spaces over F, and suppose V has a finite basis $\{v_1, \ldots, v_n\}$. If $U, T : V \to W$ are linear transformations and $U(v_i) = T(v_i)$ for $i = 1, \ldots, n$, then $U = T$ (i.e., $U(x) = T(x)$ for every $x \in V$).*

Example 2.19. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by $T(a_1, a_2) = (2a_2 - a_1, 3a_1)$. Suppose that $U : \mathbb{R}^2 \to \mathbb{R}^2$ is *any* linear transformation. If we know that $U(1, 2) = (3, 3)$ and $U(1, 1) = (1, 3)$, then $U = T$. This follows from the above corollary because $\{(1, 2), (1, 1)\}$ is a basis for $\mathbb{R}^2$.

2.3. **The matrix representation of a linear transformation.**

Definition 2.20. Let V be a finite dimension vector space over F. A **ordered basis** for V is a basis for V equipped with a specific order.

Example 2.21.
(1) In F^3, $\beta = \{e_1, e_2, e_3\}$ and $\gamma = \{e_3, e_1, e_2\}$ are both ordered bases for F^3. Note that β and γ describe the same (unordered) basis for F^3, but they are different as ordered bases because the first basis vector of β is e_1 whereas the first basis vector of γ is e_3.
(2) For the vectors space F^n over F, we call $\{e_1, e_2, \ldots, e_n\}$ the **standard ordered basis** for F^n.
(3) For the vector space $P_n(F)$ over F, we call $\{1, x, \ldots, x^n\}$ the **standard ordered basis** for $P_n(F)$.

We will not stress the precise set-theoretic difference between an ordered basis and an (unordered) basis. Usually it will be clear when the situation demands that we work with an ordered basis, for instance, when representing a linear transformation as a matrix as we'll see below.

Definition 2.22. Let $\beta = \{u_1, \ldots, u_n\}$ be an ordered basis for a finite-dimensional vector space V. For $x \in V$, let $a_1, \ldots, a_n \in F$ be the *unique* scalars (see Proposition 1.51) such that

$$x \;=\; \sum_{i=1}^{n} a_i u_i.$$

We define the **coordinate vector of x relative to β**, denoted $[x]_\beta$, by

$$[x]_\beta \;:=\; \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \;\in\; F^n.$$

Note that for each $i = 1, \ldots, n$, we have $[u_i]_\beta = e_i$. Furthermore, the correspondence $x \mapsto [x]_\beta : V \to F^n$ is actually a bijective linear transformation (exercise).

Example 2.23. Let $V = P_2(\mathbb{R})$, and let $\beta = \{1, x, x^2\}$ be the standard ordered basis for V. Let $f(x) = 4 + 6x - 7x^2 \in V$. Then

$$\big[f(x)\big]_\beta \;=\; \begin{pmatrix} 4 \\ 6 \\ 7 \end{pmatrix}.$$

Definition 2.24. Suppose V, W are finite-dimensional vector spaces, with ordered bases $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_m\}$, respectively. Suppose $T : V \to W$ is a linear transformation. Then for each $j = 1, \ldots, n$, there exists unique scalars (again, by Proposition 1.51) $a_{1j}, a_{2j}, \ldots, a_{mj} \in F$ such that

$$T(v_j) \;=\; \sum_{i=1}^{m} a_{ij} w_i \quad \text{for } j = 1, \ldots, n.$$

We call the $m \times n$ matrix A defined by $A_{ij} = a_{ij}$ the **matrix representation of T in the ordered bases β and γ**, and we write $A = [T]_\beta^\gamma$. In other words:

$$[T]_\beta^\gamma \;:=\; A \;=\; \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \vdots & a_{mn} \end{bmatrix}$$

In the special case where $V = W$ and $\beta = \gamma$, then we write just $A = [T]_\beta$. Note that

(1) for $j = 1, \ldots, n$, then jth column of A is $[T(v_j)]_\gamma$ (the coordinate vector of $T(v_j)$ relative to γ)
(2) If $U : V \to W$ is a linear transformation such that $[U]_\beta^\gamma = [T]_\beta^\gamma$, then $U = T$ by Corollary 2.18.
(3) In practice, $[T]_\beta^\gamma$ gives an explicit way to describe T which is very useful for computations.

Example 2.25. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be the linear transformation given by $T(a_1, a_2) := (a_1 + 3a_2, 0, 2a_1 - 4a_2)$. Let $\beta = \{e_1, e_2\}$ and $\gamma = \{e_1, e_2, e_3\}$ be the standard ordered basis for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Then we compute:

$$\begin{aligned} T(1,0) \;&=\; (1,0,2) \;&=\; 1e_1 + 0e_2 + 2e_3 \\ T(0,1) \;&=\; (3,0,-4) \;&=\; 3e_1 + 0e_2 - 4e_3 \end{aligned}$$

and so

$$[T]_\beta^\gamma \;=\; \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{pmatrix}$$

If instead we consider the ordered basis $\gamma' = \{e_3, e_2, e_1\}$, then

$$[T]_\beta^{\gamma'} \;=\; \begin{pmatrix} 2 & -4 \\ 0 & 0 \\ 1 & 3 \end{pmatrix}.$$

We have now given a method of associating to a linear transformation $T : V \to W$ a certain matrix $[T]_\beta^\gamma$. We will now show that this process of association faithfully preserves all of the "linear algebra" going on with $T : V \to W$. We will make this more precise.

Definition 2.26. Let $T, U : V \to W$ be functions, where V, W are vector spaces over F, and let $a \in F$. We define two new functions:

- $T + U : V \to W$ by $(T + U)(x) := T(x) + U(x)$ for all $x \in V$, and

- $aT : V \to W$ by $(aT)(x) := a \cdot T(x)$, for all $x \in V$.

Proposition 2.27. *Let V, W be vector spaces over F, and suppose $T, U : V \to W$ are linear transformations. Then*

(1) for every $a \in F$, $aT + U : V \to W$ is a linear transformation.
(2) the set of all linear transformations from V to W is a vector space over F (with the operations given in Definition 2.26).

Proof. (1) We need to show that $aT + U$ is a linear transformation. To do this, let $x, y \in V$ and $c \in F$ be arbitrary. Then note that

$$\begin{aligned}(aT + U)(cx + y) &= (aT)(cx + y) + U(cx + y) = a\big(T(cx + y)\big) + (cU)(x) + U(y)\\ &= a(cT(x) + T(y)) + cU(x) + U(y) = acT(x) + cU(x) + aT(y) + U(y)\\ &= c(aT + U)(x) + (aT + U)(y),\end{aligned}$$

and so the map $aT + U$ is linear.

(2) The zero vector is the linear transformation $T_0 : V \to W$ defined by $T_0(x) = 0$ for all $x \in V$. Verifying the axioms (VS1)-(VS8) is routine. □

Definition 2.28. For vector spaces V, W over F, we denote

$$\mathcal{L}(V, W) := \{T : T \text{ is a linear transformation from } V \text{ to } W\},$$

which is a vector space over F. In the special case where $V = W$, then we write $\mathcal{L}(V)$ instead of $\mathcal{L}(V, W)$.

2.4. **Algebraic description of the operations in $\mathcal{L}(V, W)$.** We have shown that every linear transformation $V \to W$ can be represented by a matrix. We will now show that the operations of pointwise addition and scalar multiplication on $\mathcal{L}(V, W)$ correspond to matrix addition and scalar multiplication of their matrix representations:

Proposition 2.29. *Let V and W be finite-dimensional vector spaces with ordered bases β and γ, respectively. Let $T, U : V \to W$ be linear transformations. Then:*

(1) $[T + U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$
(2) $[aT]_\beta^\gamma = a[T]_\beta^\gamma$, for all scalars $a \in F$.

Note, the operations on the right side of these equations are operations on matrices (adding to matrices and scalar multiplying a matrix by $a \in F$).

Proof. (1) Let $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_m\}$. Then there exist unique scalars a_{ij} and b_{ij}, for $1 \le i \le m, 1 \le j \le n$ such that

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i \quad \text{and} \quad U(v_j) = \sum_{i=1}^{m} b_{ij} w_i, \quad \text{for } 1 \le j \le n.$$

Thus

$$(T + U)(v_j) = \sum_{i=1}^{m} (a_{ij} + b_{ij}) w_i.$$

Thus, for the matrix $[T + U]_\beta^\gamma$ we have

$$\big([T + U]_\beta^\gamma\big)_{ij} = a_{ij} + b_{ij} = \big([T]_\beta^\gamma + [U]_\beta^\gamma\big).$$

(2) Similar. □

Example 2.30. Let $T, U : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by

$$\begin{aligned} T(a_1, a_2) &:= (a_1 + 3a_2, 0, 2a_1 - 4a_2) \\ U(a_1, a_2) &:= (a_1 - a_2, 2a_1, 3a_1 + 2a_2). \end{aligned}$$

Let β, γ be the standard ordered bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Then

$$[T]_\beta^\gamma = \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{pmatrix} \quad \text{and} \quad [U]_\beta^\gamma = \begin{pmatrix} 1 & -1 \\ 2 & 0 \\ 3 & 2 \end{pmatrix}$$

Applying the definition, we also have

$$(T+U)(a_1, a_2) = (a_1 + 3a_2, 0, 2a_1 - 4a_2) + (a_1 - a_2, 2a_1, 3a_1 + 2a_2) = (2a_1 + 2a_2, 2a_1, 5a_1 - 2a_2).$$

Thus

$$[T+U]_\beta^\gamma = \begin{pmatrix} 2 & 2 \\ 2 & 0 \\ 5 & -2 \end{pmatrix} = [T]_\beta^\gamma + [U]_\beta^\gamma,$$

as predicted by the above proposition.

2.5. **Composition of linear transformations and matrix multiplication.**

Definition 2.31. Let $T : V \to W$ and $U : W \to Z$ be two linear transformations of vector spaces. The **composition of T and U**, denoted $UT : V \to Z$, is a function from V to Z defined by $(UT)(x) := U(T(x))$ for every $x \in V$.

The composition of two linear transformations is also a linear transformation:

Lemma 2.32. *Suppose $T : V \to W$ and $U : W \to Z$ are linear transformations. Then $UT : V \to Z$ is also a linear transformation.*

Proof. Let $x, y \in V$ and $a \in F$ be arbitrary. Then

$$\begin{aligned} UT(ax + y) &= U(T(ax + y)) \quad \text{def. of composition} \\ &= U(aT(x) + T(y)) \quad \text{because } T \text{ is linear} \\ &= aU(T(x)) + U(T(y)) \quad \text{because } U \text{ is linear} \\ &= a \cdot UT(x) + UT(y) \quad \text{def. of composition.} \end{aligned}$$

□

Here are some properties of linear transformations which we state without proof:

Proposition 2.33. *Let V be a vector space over F and suppose $T, U_1, U_2 \in \mathcal{L}(V)$. Then*

(1) $T(U_1 + U_2) = TU_1 + TU_2$ and $(U_1 + U_2)T = U_1T + U_2T$,
(2) $T(U_1U_2) = (TU_1)U_2$,
(3) $TI = IT = T$, where $I : V \to V$ is the identity transformation, and
(4) $a(U_1U_2) = (aU_1)U_2 = U_1(aU_2)$ for every $a \in F$.

Now assume that V, W, Z are vector spaces over F, and let $\alpha = \{v_1, \ldots, v_p\}$, $\beta = \{w_1, \ldots, w_n\}$, $\gamma = \{z_1, \ldots, z_m\}$ be ordered bases for V, W, and Z, respectively. Furthermore, let $T : V \to W$ and $U : W \to Z$ be linear transformations. Then as before we can represent T and U as matrices with respect to these bases. Let:

$$A := [U]_\beta^\gamma \quad \text{and} \quad B := [T]_\alpha^\beta$$

be their matrix representations.

We also have the linear transformation $UT : V \to Z$. We are free to compute $[UT]_\alpha^\gamma$. For $1 \le j \le n$, we have

$$(UT)(v_j) = U\big(T(v_j)\big) = U\big(\textstyle\sum_{k=1}^n B_{kj} w_k\big) = \sum_{k=1}^n B_{kj} U(w_k)$$

$$= \sum_{k=1}^{n} B_{kj}\big(\sum_{i=1}^{m} A_{ik}z_i\big) = \sum_{i=1}^{m}\big(\sum_{k=1}^{n} A_{ik}B_{kj}\big)z_i = \sum_{i=1}^{m} C_{ij}z_i$$

where $C_{ij} := \sum_{k=1}^{n} A_{ik}B_{kj}$. Thus $[UT]_\alpha^\gamma = C = (C_{ij})_{1\le i\le p, 1\le j\le n}$.

This computation motivates the usual definition of matrix multiplication.

Definition 2.34. Let A be an $m \times n$ matrix and B an $n \times p$ matrix (both with coefficients in F). We define the **product of A and B**, denoted AB, to be the $m \times p$ matrix AB with the property that for $1 \le i \le m$ and $1 \le j \le p$, the ij-entry is

$$(AB)_{ij} := \sum_{k=1}^{n} A_{ik}B_{kj}.$$

Example 2.35. (1) We have the following multiplication (of matrices with coefficients in $\mathbb{R}$):

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 4 & -1 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 1\cdot 4 + 2\cdot 2 + 1\cdot 5 \\ 0\cdot 4 + 4\cdot 2 + (-1)\cdot 5 \end{pmatrix} = \begin{pmatrix} 13 \\ 3 \end{pmatrix}.$$

(2) In general, matrix multiplication is **NOT** commutative:

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad \text{but} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}.$$

Thus, in general you should expect that $AB \ne BA$ for square matrices A and B. (Occasionally we have $AB = BA$, but this is a rather special situation).

(3) Recall the definition of the **transpose** of a matrix $A \in M_{m\times n}(F)$: $A^t \in M_{n\times m}(F)$ is the matrix given by $(A^t)_{ij} = A_{ji}$ for $1 \le i \le n$, $1 \le j \le m$. We will show that $(AB)^t = B^tA^t$ for matrices $A \in M_{m\times n}(F)$ and $B \in M_{n\times p}(F)$. Indeed, note that for $1 \le i \le p$ and $1 \le j \le m$ we have

$$(AB)^t_{ij} = (AB)_{ji} = \sum_{k=1}^{n} A_{jk}B_{ki} = \sum_{k=1}^{n} B_{ki}A_{jk} = \sum_{k=1}^{n} (B^t)_{ik}(A^t)_{kj} = (B^tA^t)_{ij}$$

We can now relate composition of linear transformations with matrix multiplication:

Proposition 2.36. *Let V, W and Z be finite dimensional vector spaces with ordered bases α, β, and γ respectively (as above), and suppose $T : V \to W$ and $U : W \to Z$ are linear transformations. Then*

$$[UT]_\alpha^\gamma = [U]_\beta^\gamma [T]_\alpha^\beta.$$

[Note that the left side is the matrix representation of the linear transformation UT, whereas the right side is the matrix multiplication of two matrices.]

Proof. Note that no proof is actually needed since we conveniently defined the ij-entry of the right side to be the ij-entry of the left side. See our above motivating calculation to see what the ij-entry of the left side is. □

In the case where all vector spaces are the same and all ordered bases are the same this simplifies to:

Corollary 2.37. *Let V be a finite-dimensional vector space with an ordered basis β. Let $T, U \in \mathcal{L}(V)$. Then*

$$[UT]_\beta = [U]_\beta [T]_\beta.$$

Example 2.38. Let $U : P_3(\mathbb{R}) \to P_2(\mathbb{R})$ and $T : P_2(\mathbb{R}) \to P_3(\mathbb{R})$ be the linear transformations defined by $U(f(x)) := f'(x)$ and $T(f(x)) := \int_0^x f(t)dt$. Let $\alpha = \{1, x, x^2, x^3\}$ and $\beta = \{1, x, x^2\}$ be the standard ordered bases of $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$ respectively. Then we have

$$\begin{aligned} U(1) &= 0 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ U(x) &= 1 = 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ U(x^2) &= 2x = 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 \\ U(x^3) &= 3x^2 = 0 \cdot 1 + 0 \cdot x + 3 \cdot x^2 \end{aligned}$$

and so

$$[U]_\alpha^\beta = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

We also have

$$\begin{aligned} T(1) &= x = 0 \cdot 1 + 1 \cdot x + 0 \cdot x^2 + 0 \cdot x^3 \\ T(x) &= \frac{1}{2}x^2 = 0 \cdot 1 + 0 \cdot x + \frac{1}{2} \cdot x^2 + 0 \cdot x^3 \\ T(x^2) &= \frac{1}{3}x^3 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 + \frac{1}{3} \cdot x^3 \end{aligned}$$

and so

$$[T]_\beta^\alpha = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}$$

Thus

$$[UT]_\beta = [U]_\alpha^\beta [T]_\beta^\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = [I]_\beta,$$

where $I : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ is the identity transformation. This agrees with the fundamental theorem of calculus (that differentiation is the inverse operation to integration).

Definition 2.39. The $n \times n$ **identity matrix** I_n (over a field F) is defined by

$$(I_n)_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

so for instance,

$$I_1 = (1), \quad I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Here are some basic properties of matrix multiplication which we state without proof:

Proposition 2.40. *Let $A \in M_{m\times n}(F)$, $B, C \in M_{n\times p}(F)$, and $D, E \in M_{q\times m}(F)$. Then*

(1) $A(B + C) = AB + AC$ and $(D + E)A = DA + EA$,
(2) $a(AB) = (aA)B = A(aB)$ for every scalar $a \in F$,
(3) $I_m A = A = AI_n$, and
(4) if $\dim(V) = n$ and $I : V \to V$ is the identity transformation, then for every ordered basis β of V, $[I]_\beta = I_n$.

2.6. **Calculating the value of a linear transformation using its matrix representation.** The following shows that by using appropriate representations, we can reduce the action of applying a linear transformation to that of matrix multiplication. In some sense, this shows that much of finite-dimensional linear algebra can be reduced to the study of matrices (like in math33a).

Proposition 2.41. *Let $T : V \to W$ be a linear transformation such that V and W are finite-dimensional vector spaces with ordered bases β and γ, respectively. Then, for each $v \in V$ we have*

$$\big[T(v)\big]_\gamma \;=\; [T]^\gamma_\beta [v]_\beta.$$

Proof. Suppose $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_m\}$ are our ordered bases for V and W. Let $v \in V$ be arbitrary. Then $v = a_1 v_1 + \cdots + a_n v_n$ for a unique choice of scalars $a_1, \ldots, a_n \in F$. Thus

$$[v]_\beta \;=\; \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Now let $B = [T]^\gamma_\beta$. Then

$$\begin{aligned} T(v) \;&=\; a_1 T(v_1) + \cdots + a_n T(v_n) \;= \sum_{j=1}^{n} a_j T(v_j) \quad \text{(because } T \text{ is linear)} \\ &=\; \sum_{j=1}^{n} a_j \left(\sum_{i=1}^{m} B_{ij} w_i \right) \quad \text{(definition of } B\text{)} \\ &=\; \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_j B_{ij} \right) w_i \quad \text{(rearranging summation)} \end{aligned}$$

Thus

$$\big[T(v)\big]_\gamma \;=\; \begin{pmatrix} \sum_{j=1}^n a_j B_{1j} \\ \vdots \\ \sum_{j=1}^n a_j B_{mj} \end{pmatrix} \;=\; B \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix},$$

as desired. □

Example 2.42. Let $T : P_3(\mathbb{R}) \to P_2(\mathbb{R})$ be given by $T\big(f(x)\big) := f'(x)$. Then with β, γ the standard ordered bases of $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$, we have shown previously that

$$[T]^\gamma_\beta \;=\; \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

Suppose $p(x) = 2 - 4x + x^2 + 3x^3$. Then we compute directly that $T(p(x)) = p'(x) = -4 + 2x + 9x^2$. Representing this as a vector, we get

$$\big[T(p(x))\big] \;=\; \big[p'(x)\big] \;=\; \begin{pmatrix} -4 \\ 2 \\ 9 \end{pmatrix}.$$

We can also arrive at this via Proposition 2.41 by multiplying matrices:

$$[T]^\gamma_\beta \big[p(x)\big]_\beta \;=\; \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ -4 \\ 1 \\ 3 \end{pmatrix} \;=\; \begin{pmatrix} -4 \\ 2 \\ 9 \end{pmatrix}.$$

2.7. **Associating a linear transformation to a matrix.** We have just seen that a linear transformation can be represented by a matrix. We will now go in the reverse direction and associate to a matrix a linear transformation. Remember, *a priori* a matrix is just a rectangular array of scalars (a static object), it doesn't *do* anything. We'll show there is a natural way to associate a linear transformation (a dynamic object) to a matrix.

Definition 2.43. Let $A \in M_{m\times n}(F)$. We denote by L_A the mapping $L_A : F^n \to F^m$ defined by $L_A(x) = Ax$. Here we regard vectors from F^n and F^m now as **column** vectors, and the expression Ax denotes the multiplication of the $m \times n$ matrix A with the $m \times 1$ column vector x. We call L_A a **left-multiplication transformation**.

Example 2.44. Let $A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \in M_{2\times 3}(\mathbb{R})$. This gives rise to the linear transformation $L_A : \mathbb{R}^3 \to \mathbb{R}^2$. For example, If $x = \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix} \in \mathbb{R}^3$, then

$$L_A(x) \;=\; Ax \;=\; \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix} \;=\; \begin{pmatrix} 6 \\ 1 \end{pmatrix}$$

Here are some basic properties of L_A:

Proposition 2.45. *Let $A \in M_{m\times n}(F)$. Then the function $L_A : F^n \to F^m$ is a linear transformation. Furthermore, if $B \in M_{m\times n}(F)$ and β, γ are the standard ordered bases for F^n and F^m, respectively, then*

(1) $[L_A]^\gamma_\beta = A$,
(2) $L_A = L_B$ if and only if $A = B$,
(3) $L_{A+B} = L_A + L_B$, and $L_{aA} = a \cdot L_A$ for every $a \in F$,
(4) if $T : F^n \to F^m$ is a linear transformation, then there is a unique $C \in M_{m\times n}(F)$ such that $T = L_C$. In fact, $C = [T]^\gamma_\beta$,
(5) If $E \in M_{n\times p}(F)$, then $L_{AE} = L_A L_E$, and
(6) if $m = n$, then $L_{I_n} = I_{F^n}$, where $I_{F^n} : F^n \to F^n$ is the identity linear transformation.

Proof. The fact that L_A is linear is clear from Proposition 2.40 (using that we can view vectors in F^n as the same thing as matrices in $M_{n\times 1}(F)$).

(1) Note that the jth column of $[L_A]^\gamma_\beta$ is $L_A(e_j) = Ae_j$ (by definition of the matrix representation), which is also the jth column of the matrix A. Thus $[L_A]^\gamma_\beta = A$.
(2) ($\Leftarrow$) is clear. ($\Rightarrow$) Suppose $L_A = L_B$. Then from (1), $A = [L_A]^\gamma_\beta = [L_B]^\gamma_\beta = B$.
(3) Exercise.
(4) Suppose $T : F^n \to F^m$ is a linear transformation and set $C := [T]^\gamma_\beta$. Then by Proposition 2.41, $[T(x)]_\gamma = [T]^\gamma_\beta [x]_\beta$ for all $x \in V$, so $T(x) = Cx = L_C(x)$ for all $x \in F^n$. Thus $T = L_C$. The uniqueness follows from (2).
(5) We first claim that $(AE)e_j = A(Ee_j)$, for $j = 1, \ldots, p$. Note that these are both $m \times 1$ matrices. for $k = 1, \ldots, m$ we have

$$\big((AE)e_j\big)_{k1} \;=\; \sum_{l=1}^{p} (AE)_{kl}(e_j)_{l1} \;=\; \sum_{l=1}^{p} \left(\sum_{i=1}^{n} A_{ki}E_{il} \right) (e_j)_{l1} \;=\; \sum_{l=1}^{p} \sum_{i=1}^{n} A_{ki}E_{il}(e_j)_{l1}$$

$$= \sum_{i=1}^{n} \sum_{l=1}^{p} A_{ki}E_{il}(e_j)_{l1} \;=\; \sum_{i=1}^{n} A_{ki} \left(\sum_{l=1}^{p} E_{il}(e_j)_{l1} \right) \;=\; \sum_{i=1}^{n} A_{ki}(Ee_j)_{i1} \;=\; \big(A(Ee_j)\big)_{k1}.$$

Thus

$$L_{AE}(e_j) \;=\; (AE)e_j \;=\; A(Ee_j) \;=\; L_A(Ee_j) \;=\; L_A(L_E(e_j)) \;=\; L_AL_E(e_j).$$

Thus $L_{AE} = L_AL_E$ since they take the same values on a basis, by Corollary 2.18 to Linear Transformation Prescription 2.17.

(6) Exercise □

The next proposition shows that matrix multiplication is associative:

Proposition 2.46. *Let $A \in M_{m\times n}(F)$, $B \in M_{n\times p}(F)$, and $C \in M_{p\times r}(F)$. Then*

$$A(BC) \;=\; (AB)C.$$

Proof. By Proposition 2.45(5) and associativity of composition of functions (Proposition 6.8), we have

$$L_{A(BC)} \;=\; L_AL_{BC} \;=\; L_A(L_BL_C) \;=\; (L_AL_B)L_C \;=\; L_{AB}L_C \;=\; L_{(AB)C}.$$

Thus $A(BC) = (AB)C$ by Proposition 2.45(2).

Note: this can also be proved directly by a computation similar to the one done in the proof of Proposition 2.45(5). □

2.8. **Invertibility.** In this subsection, we generalize the notion of invertible square matrix to arbitrary linear transformations. We assume the reader is familiar with basic facts about invertibility of functions.

Definition 2.47. Let V and W be vector spaces, and let $T : V \to W$ be a linear transformation. We say T is **invertible**, if it is invertible as a function (see appendix).

If T is invertible, then it has an inverse *function*. The next Proposition tells us that this inverse function is automatically also a linear transformation:

Proposition 2.48. *Let $T : V \to W$ be an invertible linear transformation. Then the function $T^{-1} : V \to W$ is also a linear transformation.*

Proof. Let $w_1, w_2 \in W$ and $c \in F$ be arbitrary. Since T is one-to-one and onto (because it is invertible), there are *unique* $v_1, v_2 \in V$ such that $T(v_1) = w_1$ and $T(v_2) = w_2$. Thus $v_1 = T^{-1}(w_1)$ and $v_2 = T^{-1}(w_2)$, and so

$$\begin{aligned} T^{-1}(cw_1 + w_2) \;&=\; T^{-1}(cT(v_1) + T(v_2)) \;=\; T^{-1}T(cv_1 + v_2) \\ &=\; I_V(cv_1 + v_2) \;=\; cv_1 + v_2 \;= cT^{-1}(w_1) + T^{-1}(w_2) \end{aligned}$$

□

Example 2.49. Let $T : P_1(\mathbb{R}) \to \mathbb{R}^2$ be the linear transformation defined by $T(a+bx) := (a, a+b)$. Then $T^{-1} : \mathbb{R}^2 \to P_1(\mathbb{R})$ is defined by $T^{-1}(c, d) = c+(d-c)x$, which is also a linear transformation.

We also have a notion of *invertibility* for matrices:

Definition 2.50. Let $A \in M_{n\times n}(F)$. We say that A is **invertible** if there is $B \in M_{n\times n}(F)$ such that $AB = BA = I_n$. We call such a matrix B the **inverse of** A, and we write $B = A^{-1}$. Just like for functions, inverses for matrices are unique (when they exist). The proof of this uniqueness is the same.

We will only talk about the invertibility of a matrix when it is a square matrix. All non-square matrices are automatically not invertible.

Example 2.51. The inverse of $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ is $\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$. Indeed,

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \;=\; \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \;=\; \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Lemma 2.52. *Let $T : V \to W$ be an invertible linear transformation, and suppose* $\dim V < \infty$. *Then* $\dim V = \dim W$.

Proof. Recall that since T is invertible, it is a bijection and so it is one-to-one and onto. Let $\beta = \{x_1, \ldots, x_n\}$ be a basis for V. By Proposition 2.9, $\text{span}(T(\beta)) = R(T) = W$ (since T is onto). Next, since T is one-to-one, we have that $N(T) = \{0\}$, so $\dim N(T) = 0$. By the Dimension Theorem, this implies that $\dim W = \dim R(T) = \dim V$. □

Invertibility of linear transformations corresponds to invertibility of matrices:

Proposition 2.53. *Let V, W be finite-dimensional vector spaces with ordered bases β and γ, respectively. Suppose $T : V \to W$ is a linear transformation. Then T is invertible if and only if the matrix $[T]_\beta^\gamma$ is invertible. Furthermore, if either of these hold, then $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.*

Proof. ($\Rightarrow$) Suppose T is invertible. By Lemma 2.52, we have that both $[T]_\beta^\gamma, [T^{-1}]_\gamma^\beta \in M_{n\times n}(F)$. Note that

$$[T^{-1}]_\gamma^\beta [T]_\beta^\gamma \;=\; [T^{-1}T]_\beta \;=\; [I_V]_\beta \;=\; I_n,$$

using Proposition 2.36 and Proposition 2.45(4). Likewise, we have $[T]_\beta^\gamma [T^{-1}]_\gamma^\beta = I_n$. Thus $[T]_\beta^\gamma$ is an invertible matrix, and its inverse is $([T]_\beta^\gamma) = [T^{-1}]_\gamma^\beta$.

($\Leftarrow$) Now suppose $A = [T]_\beta^\gamma$ is invertible, say with inverse $B \in M_{n\times n}(F)$, so $AB = BA = I_n$. By Linear Transformation Prescription 2.17, there is $U \in \mathcal{L}(W, V)$ such that $U(w_j) = \sum_{i=1}^n B_{ij} v_i$ for $j = 1, \ldots, n$, where $\gamma = \{w_1, \ldots, w_m\}$ and $\beta = \{v_1, \ldots, v_n\}$. Then $[U]_\gamma^\beta = B$. To show that $U = T^{-1}$, note that

$$[UT]_\beta \;=\; [U]_\gamma^\beta [T]_\beta^\gamma \;=\; BA \;=\; I_n \;=\; [I_V]_\beta,$$

by Proposition 2.36. Thus $UT = I_V$. Similarly $TU = I_W$. □

Example 2.54. Let β and γ be the standard ordered bases of $P_1(\mathbb{R})$ and $\mathbb{R}^2$ respectively. For T given by $T(a + bx) = (a, a + b)$ from the previous example, we have

$$[T]_\beta^\gamma \;=\; \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad [T^{-1}]_\gamma^\beta \;=\; \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

which we already know are matrix inverses of each other.

Corollary 2.55. *Let $A \in M_{n\times n}(F)$. Then A is invertible if and only if L_A is invertible. If either of these hold, then $(L_A)^{-1} = L_{A^{-1}}$.*

2.9. **Isomorphisms.** You may have noticed, for instance, that the vector spaces $P_3(\mathbb{R})$ and $\mathbb{R}^4$ are essentially "the same", at least from linear algebra's point of view. The concept of *isomorphism* makes this precise.

Definition 2.56. Let V, W be vector spaces over F. We say that V **is isomorphic to** W if there exists an invertible linear transformation $T : V \to W$. Such a linear transformation T is called an **isomorphism** from V to W.

Remark 2.57. The following observations follow immediately from the definition of isomorphism:

(1) V is isomorphic to V (using I_V)
(2) V is isomorphic to W if and only if W is isomorphic to V.
(3) If V is isomorphic to W and W is isomorphic to Z, then V is isomorphic to Z.

The above three properties, taken together, say that isomorphism is a so-called *equivalence relation.*

Example 2.58. Let $T : F^2 \to P_1(F)$ be given by $T(a_1, a_2) := a_1 + a_2 x$. Then T is an isomorphism, so F^2 is isomorphic to $P_1(F)$.

By Lemma 2.52, isomorphic vector spaces have the same dimension. Remarkably, the converse also holds for finite-dimensional spaces:

Proposition 2.59. *Let V, W be finite-dimensional vector spaces over F. Then V is isomorphic to W if and only if $\dim(V) = \dim(W)$.*

Proof. ($\Rightarrow$) Suppose $T : V \to W$ is an isomorphism. Then $\dim V = \dim W$ by Lemma 2.52.

($\Leftarrow$) Suppose $\dim V = \dim W$, and let $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_n\}$ be bases for V and W, respectively. By Linear Transformation Prescription 2.17, there is a linear transformation $T : V \to W$ such that $T(v_i) = w_i$ for each $i = 1, \ldots, n$. By Proposition 2.9, $R(T) = \text{span}(T(\beta)) = \text{span}(\gamma) = W$, so T is surjective. By Proposition 2.15, T is in fact a bijection (hence invertible). Thus T is an isomorphism. □

Corollary 2.60. *Let V be a vector space over F. Then V is isomorphic to F^n if and only if $\dim V = n$.*

The following shows that we can identify the spaces $\mathcal{L}(V, W)$ with $M_{m\times n}(F)$ (for V, W of appropriate finite-dimensions).

Proposition 2.61. *Let V, W be vector spaces over F with $\dim V = n$ and $\dim W = m$. Suppose β, γ are ordered bases for V and W, respectively. Then the linear transformation $\Phi : \mathcal{L}(V, W) \to M_{m\times n}(F)$ defined by $\Phi(T) := [T]^\gamma_\beta$ for all $T \in \mathcal{L}(V, W)$ is an isomorphism.*

Proof. Φ is a linear transformation by Proposition 2.29, so it remains to show that Φ is a bijection. This means we need to show that for every $A \in M_{m\times n}(F)$, there is a *unique* linear transformation $T : V \to W$ such that $\Phi(T) = A$.

Suppose $\beta = \{v_1, \ldots, v_n\}, \gamma = \{w_1, \ldots, w_m\}$ and $A \in M_{m\times n}(F)$ are given. By Linear Transformation Prescription 2.17, there is a *unique* linear transformation $T : V \to W$ such that $T(v_j) = \sum_{i=1}^m A_{ij} w_i$ for each $j = 1, \ldots, n$. Then $[T]^\gamma_\beta = A$, so $\Phi(T) = A$. □

Corollary 2.62. *If $\dim V = n$ and $\dim W = m$, then $\dim \mathcal{L}(V, W) = mn = \dim M_{m\times n}(F)$.*

Here is an example of an isomorphism we have been working with all along.

Example 2.63. Suppose V is an n-dimensional vector space over F, with ordered basis β. Then the linear transformation $\phi_\beta : V \to F^n$ given by $\phi_\beta(v) := [v]_\beta$ is an isomorphism.

In light of this, another way to state Proposition 2.41 would be to say that the following compositions are equal:

$$L_A \phi_\beta \;=\; \phi_\gamma T$$

where $T : V \to W$ is a linear transformation between finite-dimensional vector spaces V and W, β, γ are ordered bases for V, W, and $A = [T]^\gamma_\beta$. In diagram form we would say that the following diagram "commutes":

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ {\scriptstyle \phi_\beta}\downarrow & & \downarrow{\scriptstyle \phi_\gamma} \\ F^n & \xrightarrow[L_A]{} & F^m \end{array}$$

In other words, if a vector $v \in V$ begins its journey in the upper-left space V, then it has two possible paths to get to the bottom-right space: one way is to go over to $T(v) \in W$ and then down to $\phi_\gamma T(v) \in F^m$. The other path is to go down first to $\phi_\beta(v) \in F^n$ and then over to $L_A \phi_\beta(v) \in F^m$. Saying that this diagram "commutes" means that these two different paths arrive at the same place, $\phi_\gamma T(v) = L_A \phi_\beta(v)$.

2.10. **Change of coordinate matrix.** Everything in this subsection is an extremely useful special case of everything we have done so far. The following is almost immediate:

Proposition 2.64. *Let β and β' be ordered bases for a finite-dimensional vector space V, and let $Q = [I_V]^{\beta}_{\beta'}$. Then*

(1) Q is invertible (with $Q^{-1} = [I_V]^{\beta'}_{\beta}$),
(2) For every $v \in V$, $[v]_\beta = Q[v]_{\beta'}$.

(1) Since I_V is invertible, Proposition 2.53 implies $Q = [I_V]^{\beta}_{\beta'}$ is invertible, with inverse $Q^{-1} = [I_V^{-1}]^{\beta'}_{\beta} = [I_V]^{\beta'}_{\beta}$.
(2) Let $v \in V$, then
$$[v]_\beta \;=\; [I_V(v)]_\beta \;=\; [I_V]^{\beta}_{\beta'}[v]_{\beta'} \;=\; Q[v]_{\beta'},$$
by Proposition 2.41.

The above matrix shows that multiplying by Q changes the β'-coordinates of a vector v to the β-coordinates of v. This motivates the following definition:

Definition 2.65. Let β' and β be ordered bases for a finite-dimensional vector space V. We define the **change of coordinate matrix from β' to β** to be $Q = [I_V]^{\beta}_{\beta'}$. Sometimes this is also called the **change of basis matrix**.

Example 2.66. Suppose $V = \mathbb{R}^2$, $u = \begin{pmatrix}1\\1\end{pmatrix}$. Consider the ordered bases $\beta' = \{(-1,0),(0,-1)\}$ and $\beta = \{(1,0),(0,1)\}$. Then we can compute directly
$$[u]_{\beta'} \;=\; \begin{pmatrix}-1\\-1\end{pmatrix} \quad \text{and} \quad [u]_\beta \;=\; \begin{pmatrix}1\\1\end{pmatrix}.$$
However, we could separately compute
$$[I_V]^{\beta}_{\beta'} \;=\; \begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix} \quad \text{and then conclude} \quad [u]_\beta \;=\; \begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}-1\\-1\end{pmatrix} \;=\; \begin{pmatrix}1\\1\end{pmatrix}.$$

Definition 2.67. A linear transformation $T : V \to V$ from a vector space to itself is called a **linear operator on V**.

Suppose we have ordered bases β', β for V. Then how do you compute $[T]_{\beta'}$ from $[T]_\beta$?

Proposition 2.68. *Let T be a linear operator on a finite-dimensional vector space V and suppose β', β are ordered bases for V. Let $Q = [I_V]^{\beta}_{\beta'}$ be the change of coordinate matrix (from β'-coordinates to β-coordinates). Then*
$$[T]_{\beta'} \;=\; Q^{-1}[T]_\beta Q.$$

Proof. First, recall that
$$T \;=\; I_V T = T I_V.$$
Thus
$$Q[T]_{\beta'} \;=\; [I_V]^{\beta}_{\beta'}[T]^{\beta'}_{\beta'} \;=^* \; [I_V T]^{\beta}_{\beta'} \;=\; [T I_V]^{\beta}_{\beta'} \;=^* \; [T]^{\beta}_{\beta}[I_V]^{\beta}_{\beta'} \;=\; [T]_\beta Q$$
using Proposition 2.36 at each step *. Multiplying the first and last on the left by Q^{-1} then yields
$$[T]_{\beta'} \;=\; Q^{-1}[T]_\beta Q,$$
as desired. □

Example 2.69. Consider the linear operator T on $\mathbb{R}^2$ defined by $T(x, y) = (x + y, x - y)$. Let $\beta = \{(1, 0), (0, 1)\}$ and $\beta' = \{(-1, 0), (0, -1)\}$ be ordered bases. By the previous example:

$$Q \;=\; [I_V]^{\beta}_{\beta'} \;=\; \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Coincidentally, we also have

$$Q^{-1} \;=\; [I_V]^{\beta'}_{\beta} \;=\; \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We easily see that

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

Thus

$$[T]_{\beta'} \;=\; Q^{-1}[T]_{\beta}Q \;=\; \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

3. Determinants

This section we briefly review determinants of matrices. Determinants are these magical and mysterious functions defined on square matrices. We won't study them or their secrets in this class, but instead use them as mathematical tools for use in later sections. Another way to say this is that we will *black-box*[1] the theory of determinants.

Let F be a field. Then for each $n \geq 1$ there is a function

$$\det : M_{n\times n}(F) \to F$$

called the *determinant*. We will not officially define the determinant, instead we will pretend that it is already given to us and we will just say what its properties are. If you like, you can take the formulas in the "computing the determinant" subsection below to be the definition of determinant, although the real story is much more elaborate and elegant (and unfortunately outside the scope of the class if we really want to do the theory of determinants justice).

3.1. Computing the determinant.

For $n = 1$, computing the determinant is easy:

$$\text{Given } A = (A_{11}) \in M_{1\times 1}(F), \text{ we have } \det A = A_{11}.$$

For $n = 2$, there is also a fairly simple formula for computing the determinant:

$$\text{Given } A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \in M_{2\times 2}(F), \text{ we have } \det A = A_{11}A_{22} - A_{21}A_{12}.$$

Now suppose $n \geq 2$, and let $A \in M_{n\times n}(F)$. Then for any $i, j \in \{1, \ldots, n\}$ we define the **ij-cofactor matrix of A** to be the matrix $\widetilde{A}_{ij} \in M_{(n-1)\times(n-1)}(F)$ obtained from A by deleting the ith row and the jth column. Then we have can compute the determinant of A by *cofactor expansion*:

$$\det(A) \;=\; \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\widetilde{A}_{ij}) \quad \text{for } \textit{any } 1 \leq i \leq n,$$

i.e., we can use cofactor expansion along any row, not just the top row $i = 1$. Similarly, we can use cofactor expansion along any column to compute the determinant of A:

$$\det(A) \;=\; \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\widetilde{A}_{ij}) \quad \text{for } \textit{any } 1 \leq j \leq n.$$

Note that the cofactor expansion formulas reduce the computation of the determinant of an $n \times n$ matrix down to the computation of several $(n-1) \times (n-1)$ sized determinants. Applying cofactor expansion recursively, eventually the computation will reduce to 2×2 or 1×1-sized determinants, which we know how to compute directly from above.

Example 3.1. Consider the 3×3 matrix

$$A \;=\; \begin{pmatrix} 1 & 3 & -3 \\ -3 & -5 & 2 \\ -4 & 4 & -6 \end{pmatrix} \;\in\; M_{3\times 3}(\mathbb{R}).$$

[1] See `https://en.wikipedia.org/wiki/Black_box`

We will calculate the determinant using cofactor expansion along the 1st row ($i = 1$):

$$\begin{aligned}\det\begin{pmatrix}1 & 3 & -3\\ -3 & -5 & 2\\ -4 & 4 & -6\end{pmatrix} &= (-1)^{1+1}A_{11}\det(\widetilde{A}_{11}) + (-1)^{1+2}A_{12}\det(\widetilde{A}_{12}) + (-1)^{1+3}A_{13}\det(\widetilde{A}_{13})\\ &= \det\begin{pmatrix}-5 & 2\\ 4 & -6\end{pmatrix} - 3\det\begin{pmatrix}-3 & 2\\ -4 & -6\end{pmatrix} - 3\det\begin{pmatrix}-3 & -5\\ -4 & 4\end{pmatrix}\\ &= \big[(-5)(-6) - 2\cdot 4\big] - 3\big[(-3)(-6) - 2(-4)\big] - 3\big[(-3)4 - (-5)(-4)\big]\\ &= 22 - 3\cdot 26 - 3(-32) \;=\; 40.\end{aligned}$$

In general, when using cofactor expansion to compute determinants, it helps to judiciously pick a row or a column that has many zeros, if there is one.

3.2. **Properties of the determinant.** For the sake of computation, we also record here how the determinant changes when you apply row or column operations to a matrix. Suppose $A \in M_{n\times n}(F)$. If B is a matrix obtained from A by...

(1) switching two rows (or two columns), then

$$\det B \;=\; -\det A,$$

(2) multiplying a row (or a column) of A by a scalar $c \in F$, then

$$\det(B) = c\cdot\det A,$$

(3) for $i \neq j$, adding a multiple of row i to row j (or a multiple of column i to column j), then

$$\det B \;=\; \det A.$$

Using these properties and the following fact allows you to easily compute the determinant of a matrix using the usual row-reducing algorithm.

Fact 3.2. *If $A \in M_{n\times n}(F)$ is **upper triangular**, i.e., if $A_{ij} = 0$ for all $i > j$ (entries below the diagonal are $= 0$), then* $\det A = A_{11}\cdot A_{22}\cdots A_{nn}$.

Example 3.3. Let $B = \begin{pmatrix}0 & 1 & 3\\ -2 & -3 & -5\\ 4 & -4 & 4\end{pmatrix}$. We will use row operations to take B to an upper-triangular matrix:

$$B \;=\; \begin{pmatrix}0 & 1 & 3\\ -2 & -3 & -5\\ 4 & -4 & 4\end{pmatrix} \xrightarrow{R_1\leftrightarrow R_2} \begin{pmatrix}-2 & -3 & -5\\ 0 & 1 & 3\\ 4 & -4 & 4\end{pmatrix} \xrightarrow{R_3\to R_3+2R_1} \begin{pmatrix}-2 & -3 & -5\\ 0 & 1 & 3\\ 0 & -10 & -6\end{pmatrix}$$

$$\xrightarrow{R_3\to R_3+10R_3} \begin{pmatrix}-2 & -3 & -5\\ 0 & 1 & 3\\ 0 & 0 & 24\end{pmatrix}$$

This shows us that $\det B = (-1)\cdot(-2)\cdot 1\cdot 24 = 48$, since the determinant of the final matrix is $(-2)\cdot 1\cdot 24 = -48$, and then we have to multiply by an additional (-1) since we did a row exchange in the first step (the other steps leave the determinant unchanged).

Going forward, the following facts about the determinant are the most important:

Fact 3.4. *Let $A, B \in M_{n\times n}(F)$. Then*

(1) $\det(AB) = \det(A)\cdot\det(B)$.

(2) If $I_n \in M_{n\times n}(F)$ is the identity matrix, then $\det(I_n) = 1$.

(3) A is invertible if and only if $\det(A) \neq 0$. *In this case,*

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

(4) $\det(A) = \det(A^t)$.

Finally, we say that A and B are ***similar*** *if there is an invertible matrix $Q \in M_{n\times n}(F)$ such that $B = Q^{-1}AQ$. Then we have that*

(5) if A and B are similar, then $\det A = \det B$.

Proof. (1) We will just take this for granted.

(2) Since I_n is a particular upper-triangular matrix, we have $\det I_n = \underbrace{1 \cdots 1}_{n \text{ times}} = 1$.

(3) Suppose A is invertible. Then there is B such that $AB = I_n$. Applying (1) and (2) gives $\det(AB) = \det(A) \cdot \det(B) = \det(I_n) = 1$. Thus $\det(A) \neq 0$ and $\det(B) = 1/\det(A)$.

Suppose A is not invertible. Then the columns are not linearly-independent, so we can apply column operations to transform A into a matrix that has a column of all zeros. This last matrix will have determinant 0, so A has determinant 0.

(4) We will take this for granted. To convince yourself, note that if you compute $\det(A)$ completely by always doing cofactor expansion along the top-most row, and if you compute $\det(A^t)$ completely by always doing cofactor expansion along the left-most column, then your two computations will be identical (mirror-images of each other about the main diagonal), so you will get the same number.

(5) Follows from (1) and (3). □

3.3. **The determinant of a linear operator (New!)** We saw previously that linear operators $T : V \to V$ on finite-dimensional vector spaces are very analogous to square matrices. This analogy also applies to determinants:

Lemma 3.5. *Suppose V is a finite-dimensional vector space over a field F, and $T : V \to V$. Then there is a scalar $d \in F$ such that for every ordered basis β of V, we have $d = \det[T]_\beta$.*

Proof. First, let γ be an ordered basis of V, and set $d := \det[T]_\gamma$. Next, let β be an arbitrary ordered basis of V. Consider the (invertible) change of coordinates matrix $Q := [I_V]^\beta_\gamma$. Then we have $[T]_\gamma = Q^{-1}[T]_\beta Q$, i.e., $[T]_\gamma$ and $[T]_\beta$ are similar matrices. Thus $\det[T]_\beta = \det[T]_\gamma = d$. □

Definition 3.6. For a linear operator $T : V \to V$ on a finite-dimensional vector space, we define its **determinant**, $\det T$, as follows: choose any ordered basis β of V and define $\det T = \det[T]_\beta$. By the previous lemma, the choice of β does not matter.

Example 3.7. Define the operator $T : M_{2\times 2}(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ as follows:

$$T\begin{pmatrix} a & b \\ c & d \end{pmatrix} := \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We want to compute $\det T$. Consider the standard ordered basis $\beta = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ of $M_{2\times 2}(\mathbb{R})$. We first compute $[T]_\beta$:

$$\begin{aligned}
T\begin{pmatrix}1&0\\0&0\end{pmatrix} &= \begin{pmatrix}2&1\\0&3\end{pmatrix}\begin{pmatrix}1&0\\0&0\end{pmatrix} = \begin{pmatrix}2&0\\0&0\end{pmatrix} &&= 2E_{11}+0E_{12}+0E_{21}+0E_{22}\\
T\begin{pmatrix}0&1\\0&0\end{pmatrix} &= \begin{pmatrix}2&1\\0&3\end{pmatrix}\begin{pmatrix}0&1\\0&0\end{pmatrix} = \begin{pmatrix}0&2\\0&0\end{pmatrix} &&= 0E_{11}+2E_{12}+0E_{21}+0E_{22}\\
T\begin{pmatrix}0&0\\1&0\end{pmatrix} &= \begin{pmatrix}2&1\\0&3\end{pmatrix}\begin{pmatrix}0&0\\1&0\end{pmatrix} = \begin{pmatrix}1&0\\3&0\end{pmatrix} &&= 1E_{11}+0E_{12}+3E_{21}+0E_{22}\\
T\begin{pmatrix}0&0\\0&1\end{pmatrix} &= \begin{pmatrix}2&1\\0&3\end{pmatrix}\begin{pmatrix}0&0\\0&1\end{pmatrix} = \begin{pmatrix}0&1\\0&3\end{pmatrix} &&= 0E_{11}+1E_{12}+0E_{21}+3E_{22}
\end{aligned}$$

and thus

$$[T]_\beta = \begin{pmatrix}2&0&1&0\\0&2&0&1\\0&0&3&0\\0&0&0&3\end{pmatrix}.$$

We now compute

$$\det T = \det[T]_\beta = \det\begin{pmatrix}2&0&1&0\\0&2&0&1\\0&0&3&0\\0&0&0&3\end{pmatrix} = 2\cdot 2\cdot 3\cdot 3 = 36.$$

by cofactor expansion along the bottom row twice.

Proposition 3.8. *Suppose $T : V \to V$ is a linear operator. Then*

(1) T is bijective if and only if $\det T \neq 0$,
(2) if T is bijective, then $\det(T^{-1}) = (\det T)^{-1}$,
(3) if $U : V \to V$ is another linear operator on V, then $\det(TU) = \det(T)\det(U)$.

Proof. (1) T is bijective if and only if $[T]_\beta$ is invertible, where β is some (any) ordered basis of V (by Proposition 2.53). The matrix $[T]_\beta$ is invertible if and only if $\det[T]_\beta \neq 0$, by Fact 3.4(3).

(2) Suppose T is bijective (hence injective), and β is an ordered basis. Then

$$\begin{aligned}
1 = \det I_n = \det[I_V]_\beta = \det[TT^{-1}]_\beta &= \det\big([T]_\beta[T^{-1}]_\beta\big)\\
&= (\det[T]_\beta)(\det[T^{-1}]_\beta) = (\det T)(\det T^{-1}).
\end{aligned}$$

Thus $(\det T)^{-1} = \det(T^{-1})$.

(3) Again, let β be an ordered basis of V. Then we have

$$\det(TU) = \det[TU]_\beta = \det\big([T]_\beta[U]_\beta\big) = \big(\det[T]_\beta\big)\big(\det[U]_\beta\big) = (\det T)(\det U). \qquad \square$$

4. Eigenvalues and eigenvectors

In this section, V is a finite-dimensional vector space over a field F.
Since diagonal matrices are very nice to work with, our goal in this section is to study, given a linear operator $T : V \to V$ whether or not there is an ordered basis β of V such that $[T]_\beta$ is a diagonal matrix. To make this precise:

Definition 4.1. (1) A matrix $B \in M_{n\times n}(F)$ is a **diagonal matrix** if for all $i, j \in \{1, \dots, n\}$, if $i \neq j$, then $B_{ij} = 0$ (i.e., the non-diagonal entries of B are all 0). So a diagonal matrix looks like

$$B = \begin{pmatrix} B_{11} & 0 & \cdots & 0 \\ 0 & B_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_{nn} \end{pmatrix}$$

where $B_{11}, \dots, B_{nn} \in F$ (possibly also zero).
(2) A linear operator $T : V \to V$ is called **diagonalizable** if there is an ordered basis β of V such that $[T]_\beta$ is a diagonal matrix.
(3) A matrix $A \in M_{n\times n}(F)$ is **diagonalizable** if A is similar to a diagonal matrix, i.e., if there is an invertible matrix $Q \in M_{n\times n}(F)$ such that $B = Q^{-1}AQ$ is a diagonal matrix.

Diagonalizability for linear operators and for matrices essentially amounts to the same thing:

Proposition 4.2. *Let $T : V \to V$ be a linear operator and suppose β is an ordered basis for V. Then T is diagonalizable if and only if $[T]_\beta$ is a diagonalizable matrix.*

Proof. ($\Rightarrow$) Suppose T is diagonalizable. Then there is an ordered basis γ for V such that $D = [T]_\gamma$ is a diagonal matrix. Let $Q = [I_V]^\beta_\gamma$ be the change of coordinate matrix from γ-coordinates to β-coordinates. Then $D = [T]_\gamma = Q^{-1}[T]_\beta Q$, $[T]_\beta$ is similar to a diagonal matrix D, hence $[T]_\beta$ is diagonalizable.

($\Leftarrow$) Suppose $[T]_\beta$ is a diagonalizable matrix. Then there is an invertible $Q \in M_{n\times n}(F)$ such that $Q^{-1}[T]_\beta Q$ is a diagonal matrix. We want to show that the operator T is diagonalizable, i.e., we want to find an ordered basis γ of V such that $[T]_\gamma$ is a diagonal matrix. How do we find this basis γ? The idea is to use the entries of Q to construct γ using β.

Specifically, suppose $\beta = \{v_1, \dots, v_n\}$. For $j = 1, \dots, n$, define $w_j := \sum_{i=1}^n Q_{ij}v_i$. Define $\gamma = \{w_1, \dots, w_n\}$. First, we need to show that γ is actually a basis of V. To do this, define the auxiliary operator $U : V \to V$ by $U(v_j) := w_j$ for all $j = 1, \dots, n$ (defining in this way is possible by Linear Transformation Prescription and the fact that β is a basis). Then $[U]_\beta = Q$ is invertible, so U is invertible. Thus $R(U) = V = \text{span}(\gamma)$, and so γ is a basis (it has n vectors and it spans an n-dimensional space). Since γ is an ordered basis, we can ask what is the matrix $[I_V]^\beta_\gamma$? By definition of γ, we actually have $[I_V]^\beta_\gamma = Q$. Finally, note that $Q^{-1}[T]_\beta Q = [I_V]^\gamma_\beta [T]_\beta [I_V]^\beta_\gamma = [T]_\gamma$ is a diagonal matrix. Thus T is diagonal. □

Corollary 4.3. *$A \in M_{n\times n}(F)$ is diagonalizable if and only if L_A is diagonalizable.*

In this section we will answer the following question:

Question 4.4. *When is a matrix $A \in M_{n\times n}(F)$ (equivalently, a linear operator $T : V \to V$) diagonalizable?*

Already, we can provide an answer (not necessarily the definitive answer, since its basically just a restatement of the definition):

Proposition 4.5. *Suppose $T : V \to V$ is a linear operator. Then T is diagonalizable if and only if there is an ordered basis $\beta = \{v_1, \ldots, v_n\}$ for V and scalars $\lambda_1, \ldots, \lambda_n \in F$ such that*

$$T(v_j) \;=\; \lambda_j v_j \quad \text{for } 1 \le j \le n.$$

Proof. ($\Rightarrow$) Suppose T is diagonalizable, i.e., suppose there is an ordered basis $\beta = \{v_1, \ldots, v_n\}$ such that $D = [T]_\beta$ is a diagonal matrix. Then for each $v_j \in \beta$, we have $T(v_j) = \sum_{i=1}^n D_{ij} v_j = D_{jj} v_j = \lambda_j v_J$, for $\lambda_j = D_{jj}$.

($\Leftarrow$) Suppose there is an ordered basis $\beta = \{v_1, \ldots, v_n\}$ and scalars $\lambda_1, \ldots, \lambda_n \in F$ such that $T(v_j) = \lambda_j v_j$ for each $1 \le j \le n$. Then

$$[T]_\beta \;=\; \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

is a diagonal matrix, so T is diagonalizable. □

The above Proposition suggests the following definition:

Definition 4.6. (1) A *non-zero* vector $v \in V$ is an **eigenvector** of T if $T(v) = \lambda v$ for some $\lambda \in F$. We call λ the **eigenvalue** of T corresponding to the eigenvector v.

(2) Let $A \in M_{n\times n}(F)$. A non-zero $v \in F^n$ is an **eigenvector** of A if $Av = \lambda v$ for some $\lambda \in F$, and λ is called the **eigenvalue** of A corresponding to the eigenvector v.

(3) The vectors v_j in the basis β in Proposition 4.5 are eigenvectors of T with corresponding eigenvalues λ_j.

Eigenvalue Criterion 4.7. *Suppose $T : V \to V$ is a linear operator and $A \in M_{n\times n}(F)$ is a matrix. Let $\lambda \in F$.*

(1) The scalar λ is an eigenvalue of T if and only if $\det(T - \lambda I_V) = 0$.

(2) The scalar λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.

Proof. We'll show (1). We have

$$\begin{aligned} \lambda \text{ is an eigenvalue of } T \quad &\Leftrightarrow \quad T(v) = \lambda v \text{ for some } v \neq 0 \text{ in } V \\ &\Leftrightarrow \quad (T - \lambda I_n)(v) = 0 \text{ for some } v \neq 0 \text{ in } V \\ &\Leftrightarrow \quad N(T - \lambda I_V) \neq \{0\} \\ &\Leftrightarrow \quad T - \lambda I_V \text{ is not bijective, One-to-one Criterion 2.14, Prop 2.15} \\ &\Leftrightarrow \quad \det(T - \lambda I_V) = 0. \end{aligned}$$

The proof for (2) is analogous. □

Example 4.8. Let $A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} \in M_{2\times 2}(\mathbb{R})$. Then

$$\det(A - \lambda I_2) \;=\; \det \begin{pmatrix} 1-\lambda & 1 \\ 4 & 1-\lambda \end{pmatrix} \;=\; (1-\lambda)^2 - 4 \;=\; (\lambda - 3)(\lambda + 1).$$

Thus the eigenvalues of A are the solutions to the equation $(\lambda - 3)(\lambda + 1) = 0$, which are $\lambda = 3, -1$.

Definition 4.9. (1) The polynomial $f(t) := \det(A - tI_n)$ in the variable t is called the **characteristic polynomial** of A.

(2) Given a linear operator $T : V \to V$, we define the **characteristic polynomial** of T to be $f(t) = \det(T - tI_V)$. Note that this is the same thing as $f(t) = \det(A - tI_n)$, where $A = [T]_\beta$ and β is any ordered basis for V. Indeed, let β be an ordered basis for V, then

$$f(t) \;=\; \det(T - tI_V) \;=\; \det([T - t_I V]_\beta) \;=\; \det([T]_\beta - t[I_V]_\beta) \;=\; \det([T]_\beta - tI_n).$$

Here are some easy consequences of the definitions so far:

Lemma 4.10. *Let $A \in M_{n\times n}(F)$ be given, and let $f(t)$ be its characteristic polynomial*

(1) $f(t)$ is a polynomial of degree n, with leading coefficient $(-1)^n$:

$$f(t) \;=\; (-1)^n t^n + c_{n-1}t^{n-1} + \cdots + c_0 \quad \text{for some } c_0, \ldots, c_n \in F.$$

(2) A scalar $\lambda \in F$ is an eigenvalue of A if and only if $f(\lambda) = 0$.

(3) A has at most n distinct eigenvalues (since $f(t)$ has at most n distinct roots).

(4) If $\lambda \in F$ is an eigenvalue of A, then a vector $x \in F^n$ is an eigenvector of A corresponding to λ if and only if $x \neq 0$ and $x \in N(L_A - \lambda I_{F^n})$.

Example 4.11. We'll consider $A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$ again and find all eigenvectors corresponding to each of its eigenvalues.

(1) We say the eigenvalues of A are $\lambda_1 = 3$ and $\lambda_2 = -1$.

(2) Let $B_1 = A - \lambda_1 I_2 = \begin{pmatrix} -2 & 1 \\ 4 & -2 \end{pmatrix}$. Then $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ is an eigenvector of A corresponding to $\lambda_1 = 3$ if and only if $x \neq 0$ and $x \in N(L_B)$, if and only if $x \neq 0$ and $\begin{pmatrix} -2 & 1 \\ 4 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, if and only if

$$\begin{aligned} -2x_1 + x_2 &= 0 \\ 4x_1 - 2x_2 &= 0. \end{aligned}$$

The set of all solutions to this system of equations is

$$\left\{ t \begin{pmatrix} 1 \\ 2 \end{pmatrix} : t \in \mathbb{R} \right\}$$

Thus $x \in \mathbb{R}^2$ is an eigenvector of A corresponding to $\lambda_1 = 3$ if and only if $x = t\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ for some $t \neq 0$.

(3) Now let $B_2 := A - \lambda_2 I_2 = \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}$. Hence $x \in \mathbb{R}^2$ is an eigenvector of A corresponding to λ_2 if and only if $x \neq 0$ and $x \in N(L_{B_2})$, if and only if $B_2 \cdot x = 0$, if and only if $\begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, if and only if

$$\begin{aligned} 2x_1 + x_2 &= 0 \\ 4x_1 + 2x_2 &= 0. \end{aligned}$$

Thus

$$N(L_{B_2}) \;=\; \left\{ t \begin{pmatrix} 1 \\ -2 \end{pmatrix} : t \in \mathbb{R} \right\}.$$

This means x is an eigenvector of A corresponding to $\lambda_2 = -1$ if and only if $x = t\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ for some $t \neq 0$.

Note that $\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\}$ is a basis for $\mathbb{R}^2$ consisting of eigenvectors of A. Thus L_A, and hence A, is diagonalizable.

4.1. **Determining eigenvectors and eigenvalues of a linear operator.** Suppose $\dim(V) = n$ and let β be some ordered basis for V. Let $T \in \mathcal{L}(V)$ be a linear operator on V. Summarizing the results of the previous section, we describe how to determine all eigenvalues and corresponding eigenvectors of T.

(1) First, determine the matrix representation $[T]_\beta$ of T.
(2) Next, determine the eigenvalues of T. Use that $\lambda \in F$ is an eigenvalue of T if and only if λ is a root of the characteristic polynomial of T. That is, we need to find the solutions $x \in F$ of $\det([T]_\beta - xI_n) = 0$. There are at most n distinct solutions $\lambda_1, \ldots, \lambda_n$ (but possibly fewer).
(3) For each eigenvalue λ of T, we can determine the corresponding eigenvectors. We have $T(v) = \lambda v$ if and only if $(T - \lambda I_V)(v) = 0$ if and only if $[T - \lambda I_V]_\beta [v]_\beta = 0$. Thus, eigenvectors v corresponding to λ are the nonzero solutions of this system of linear equations (more precisely, solving this system we find the β-coordinate vector $[v]_\beta$, which then determines v).

Distinctness of eigenvalues goes a long way:

Proposition 4.12. *Let $T \in \mathcal{L}(V)$ be a linear operator on V, and let $\lambda_1, \ldots, \lambda_k$ be* distinct *eigenvalues of T. If $v_1, \ldots, v_k$ are eigenvectors of T such that v_i corresponds to λ_i for $i = 1, \ldots, k$, then $\{v_1, \ldots, v_k\}$ is linearly independent.*

Proof. We prove this by induction on k.

For the base case $k = 1$, suppose v_1 is an eigenvector corresponding to λ_1. Then $v_1 \neq 0$, by definition of eigenvector. Thus the set $\{v_1\}$ is linearly independent.

Let $k \geq 2$, and suppose we know the theorem is true for $k-1$ many eigenvalues and eigenvectors. We have the vectors $v_1, \ldots, v_k$, eigenvectors corresponding to the distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Suppose $a_1, \ldots, a_k \in F$ are arbitrary such that

$$a_1 v_1 + \cdots + a_k v_k \;=\; 0.$$

Apply the linear transformation $T - \lambda_k I_V$ to both sides and use linearity to get

$$(T - \lambda_k I_V)(0) \;=\; T(0) - \lambda_k I_V(0) \;=\; 0 - 0 \;=\; 0$$

for the right side, and for the left side:

$$\begin{aligned}
(T - \lambda_k I_V)(a_1 v_1 + \cdots + a_k v_k) &= (aT(v_1) + \cdots + a_k T(v_k)) - \lambda_k (a_1 v_1 + \cdots + a_k v_k) \\
&= a\lambda_1 v_1 + \cdots + a_k \lambda_k v_k - \lambda_k a_1 v_1 - \cdots - \lambda_k a_k v_k \quad \text{as } T(v_i) = \lambda_i v_i \\
&= a_1(\lambda_1 - \lambda_k) v_1 + \cdots + a_{k-1}(\lambda_{k-1} - \lambda_k) v_{k-1} \\
&= 0 \quad \text{because left side must equal right side.}
\end{aligned}$$

By the inductive hypothesis, $\{v_1, \ldots, v_{k-1}\}$ are linearly independent, so

$$a_1(\lambda_1 - \lambda_k) \;=\; \cdots \;=\; a_{k-1}(\lambda_{k-1} - \lambda_k) \;=\; 0.$$

Since $\lambda_1, \ldots, \lambda_k$ are distinct by assumption, $\lambda_i - \lambda_k \neq 0$ for $i = 1, \ldots, k-1$. Thus $a_1 = \cdots = a_{k-1} = 0$. Thus $a_k v_k = 0$, which implies $a_k = 0$ since $v_k \neq 0$ (as v_k is an eigenvector). We conclude that $\{v_1, \ldots, v_k\}$ are linearly independent. □

Corollary 4.13. *Let $T \in \mathcal{L}(V)$ and $\dim(V) = n$. If T has n distinct eigenvalues, then T is diagonalizable.*

Proof. Let $\lambda_1, \ldots, \lambda_n$ be n distinct eigenvalues of T. For each i, let v_i be an eigenvector corresponding to λ_i. By Proposition 4.12, $\{v_1, \ldots, v_n\}$ is linearly independent. Since $\dim(V) = n$, this set is a basis for V. Thus V has a basis consisting of eigenvectors for T, so T is diagonalizable. □

Example 4.14. The converse of Corollary 4.13 is false. For instance, the identity operator I_V has only one eigenvalue $\lambda = 1$, however it is diagonalizable.

So far in this class, which field F we are working over hasn't really mattered. Now it matters.

Definition 4.15. A polynomial $f(t) \in P(F)$ **splits over** F if there are scalars $c, a_1, \ldots, a_n \in F$ (not necessarily distinct) such that

$$f(t) \;=\; c(t-a_1)(t-a_2)\cdots(t-a_n)$$

Example 4.16. (1) $t^2 - 1 \in P_2(\mathbb{R})$ splits over $\mathbb{R}$, namely $t^2 - 1 = (t-1)(t+1)$.

(2) $t^2 + 1 \in P_2(\mathbb{R})$ does not split over $\mathbb{R}$. However, viewed as a polynomial $t^2 + 1 \in P_2(\mathbb{C})$, it does split over $\mathbb{C}$, namely $t^2 + 1 = (t+i)(t-i)$.

Lemma 4.17. *Suppose $T : V \to V$ is a diagonalizable linear operator. Then the characteristic polynomial of T splits.*

Proof. Let $n = \dim(V)$ and suppose $T \in \mathcal{L}(V)$ is diagonalizable. Then there is an ordered basis β of V such that $[T]_\beta = D$ is a diagonal matrix:

$$D \;=\; \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

If $f(t)$ is the characteristic polynomial of T, then

$$f(t) \;=\; \det(T - tI_V) \;=\; \det(D - tI_n) \;=\; (\lambda_1 - t)\cdots(\lambda_n - t) \;=\; (-1)^n(t-\lambda_1)\cdots(t-\lambda_n). \quad \square$$

Definition 4.18. Let λ be an eigenvalue of a linear operator or matrix with characteristic polynomial $f(t)$. The **(algebraic) multiplicity of** λ is the largest positive integer k for which $(t-\lambda)^k$ is a factor of $f(t)$ (i.e., $f(t)$ can be written as $f(t) = (t-\lambda)^k g(t)$ for some polynomial $g(t)$).

Example 4.19. Let $A = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 4 \\ 0 & 0 & 4 \end{pmatrix}$. Then the characteristic polynomial is $f(t) = -(t-3)^2(t-4)$. Hence $\lambda = 3$ is an eigenvalue of A with multiplicity 2, and $\lambda = 4$ is an eigenvalue of A with multiplicity 1.

Definition 4.20. Let $T \in \mathcal{L}(V)$, λ an eigenvalue of T. We define E_λ, the **eigenspace of** T **corresponding to** λ, as

$$E_\lambda \;=\; \{x \in V : T(x) = \lambda x\} \;=\; N(T - \lambda I_V) \quad \text{(and similarly for a matrix)}.$$

Note that this is a subspace of V, consisting of 0 and the eigenvectors of T corresponding of λ.

Sometimes we refer to $\dim E_\lambda$ as the **geometric multiplicity of** λ. The next proposition says that the geometric multiplicity of a particular eigenvalue is always at most the algebraic multiplicity:

Proposition 4.21. *Let $T \in \mathcal{L}(V)$, $\dim(V) < \infty$, λ an eigenvalue of T with multiplicity m. Then $1 \le \dim(E_\lambda) \le m$.*

Proof. Since λ is an eigenvalue, there is at least one nonzero $v \in E_\lambda$, and thus $1 \le \dim E_\lambda$.

Next, choose an ordered basis $\{v_1, \ldots, v_p\}$ for E_λ. By Corollary 1.58 to the Replacement Lemma, we can extend this to an ordered basis $\beta = \{v_1, \ldots, v_p, v_{p+1}, \ldots, v_n\}$ for V. Let $A = [T]_\beta$. Since $v_1, \ldots, v_p$ are eigenvectors of T corresponding to λ, we have

$$A \;=\; \begin{pmatrix} \lambda I_p & B \\ 0 & C \end{pmatrix}$$

where I_p is the $p\times p$ identity matrix, $B \in M_{p\times(n-p)}(F)$, $C \in M_{(n-p)\times(n-p)}(F)$, and 0 is the $(n-p)\times p$ zero matrix. Then

$$\begin{aligned} f(t) &= \det(A - tI_n) \\ &= \begin{pmatrix} (\lambda - t)I_p & B \\ 0 & C - tI_{n-p} \end{pmatrix} \\ &= \det((\lambda - t)I_p)\det(C - tI_{n-p}) \quad \text{(exercise)} \\ &= (\lambda - t)^p g(t), \end{aligned}$$

where $g(t)$ is a polynomial. Thus $(\lambda - t)^p$ is a factor of $f(t)$, hence the (algebraic) multiplicity of λ is at least p. Since $\dim(E_\lambda) = p$, this shows $\dim E_\lambda \leq m$. □

Lemma 4.22. *Let $T \in \mathcal{L}(V)$, $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues of T. Let $v_i \in E_{\lambda_i}$ for each $i = 1, \ldots, k$. If $v_1 + \cdots + v_k = 0$, then $v_i = 0$ for all i.*

Proof. Assume towards a contradiction that (after possibly rearranging the order), we have $v_i \neq 0$ for $1 \leq i \leq m$, and $v_i = 0$ for $i > m$, for some $1 \leq m \leq k$. Then for each $i \leq m$, v_i is an eigenvector of T corresponding to λ_i (since $v_i \neq 0$), and $v_1 + \cdots + v_m = 0$. This contradicts Proposition 4.12 since $v_1, \ldots, v_m$ must be linearly independent. □

The following shows there is no extra linear dependence happening between different eigenspaces:

Proposition 4.23. *Let $T \in \mathcal{L}(V)$, let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues of T. For each $i = 1, \ldots, k$, let S_i be a finite linearly independent subset of E_{λ_i}. Then $S = S_1 \cup \cdots \cup S_k$ is also a linearly independent subset of V.*

Proof. Suppose that $S_i = \{v_{i,1}, \ldots, v_{i,n_i}\}$ for each $i = 1, \ldots, k$, and some integer $n_i \geq 0$. Then $S = \{v_{i,j} : 1 \leq i \leq k, 1 \leq j \leq n_i\}$. Let $\{a_{i,j}\}$ be a collection of scalars in F such that

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} a_{i,j}v_{i,j} = 0.$$

For each i, let $w_i := \sum_{j=1}^{n_i} a_{i,j}v_{i,j}$. Then $w_i \in E_{\lambda_i}$ and $w_1 + \cdots + w_k = 0$. By the above lemma, $w_i = 0$ for each $i = 1, \ldots, k$. But since each S_i is linearly independent, it follows that $a_{i,j} = 0$ for all j. Thus S is linearly independent. □

Theorem 4.24. *Let $T \in \mathcal{L}(V)$, $\dim(V)$, and assume that the characteristic polynomial of T splits. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T. Then*

(1) T is diagonalizable if and only if the multiplicity of λ_i is equal to $\dim(E_{\lambda_i})$ for each i.
(2) If T is diagonalizable and β_i is an ordered basis for E_{λ_i} for each i, then $\beta = \beta_1 \cup \cdots \cup \beta_k$ is an ordered basis for V consisting of eigenvectors of T (and hence $[T]_\beta$ is a diagonal matrix).

Proof. For each $i = 1, \ldots, k$, let m_i denote the (algebraic) multiplicity of λ_i, and $d_i = \dim E_{\lambda_i}$, and suppose $n = \dim V$.

($\Rightarrow$) Suppose T is diagonalizable. Then there is β, an ordered basis of eigenvectors of T. For each i, we let $\beta_i = \beta \cap E_{\lambda_i}$, and set $n_i := |\beta_i|$. Then we know

- $n_i \leq d_i$ for each i, because β_i is a linearly independent subset of E_{λ_i}, and $d_i = \dim E_{\lambda_i}$,
- $d_i \leq m_i$ by Proposition 4.21
- $\sum_{i=1}^{k} n_i = n$, because β contains n vectors in total,
- $\sum_{i=1}^{k} m_i = n$, because the degree of the characteristic polynomial of T is equal to the sum of the algebraic multiplicities of the eigenvalues (since it splits), and also equal to $\dim(V) = n$.

Combining these yields:

$$n \;=\; \sum_{i=1}^{k} n_i \;\le\; \sum_{i=1}^{k} d_i \;\le\; \sum_{i=1}^{k} m_i \;=\; n.$$

Thus $\sum_{i=1}^{k}(m_i - d_i) = 0$. This implies $m_i = d_i$ for all i.

($\Leftarrow$) Conversely, suppose that $m_i = d_i$ for all i. For each i, let β_i be an ordered basis for E_{λ_i}, and set $\beta = \beta_1 \cup \cdots \cup \beta_k$. By Proposition 4.23, β is linearly independent. Furthermore, since $d_i = m_i$ for all i, β has $\sum_{i=1}^{k} d_i = \sum_{i=1}^{k} m_i = n$ many vectors in it. Thus β is a basis for all of V consisting of eigenvectors of T. Thus T is diagonalizable. □

We now summarize what we know so far:

Test for diagonalization 4.25. *Suppose $T \in \mathcal{L}(V)$, where $\dim V = n$. Then T is **diagonalizable** if and only if* both *of the following conditions hold:*

(1) The characteristic polynomial of T splits, and
(2) For each eigenvalue λ of T, the (algebraic) multiplicity of λ in the characteristic polynomial equals $\dim E_\lambda = \dim N(T - \lambda I_V) = n - \operatorname{rank}(T - \lambda I_V)$.

An analogous statement holds for square matrices $A \in M_{n\times n}(F)$.

Example 4.26. Consider $A = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix} \in M_{3\times 3}(\mathbb{R})$. We'll test A's diagonalizability. The characteristic polynomial is

$$f(t) \;=\; \det(A - tI_3) \;=\; \det\begin{pmatrix} 3-t & 1 & 0 \\ 0 & 3-t & 0 \\ 0 & 0 & 4-t \end{pmatrix} \;=\; -(t-4)(t-3)^2.$$

This shows that $f(t)$ splits, so condition (1) for diagonalizability holds. The eigenvalues are $\lambda_1 = 4$ (with multiplicity 1) and $\lambda_2 = 3$ (with multiplicity 2). Condition (2) is automatically satisfied for λ_1 (Since Proposition 4.21 says that $1 \le \dim E_{\lambda_1} \le \operatorname{mult} \lambda_1 = 1$). Thus we need only to check condition (2) for λ_2. We see that the matrix

$$A - \lambda_2 I_3 \;=\; \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

has rank 2 (by which we mean $L_{A-\lambda_2 I_3}$ has rank 2). Thus $\dim E_{\lambda_2} = 3 - \operatorname{rank}(A - \lambda_2 I_3) = 3 - 2 = 1 \neq 2$ (the multiplicity of λ_2). We conclude that A is not diagonalizable.

Example 4.27. Now consider $A = \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}$. Then $f(t) = \det(A - tI_2) = (t-1)(t-2)$. Thus $\lambda_1 = 1$, $\lambda_2 = 2$ are the eigenvalues of A, both with multiplicity 1, so A is diagonalizable since this forces conditions (1) and (2) to be satisfied. Furthermore, calculations show that

$$E_{\lambda_1} \;=\; N(L_A - 1 \cdot I_{\mathbb{R}^2}) \;=\; \operatorname{span}\{\begin{pmatrix} -2 \\ 1 \end{pmatrix}\} \quad \text{and} \quad E_{\lambda_2} \;=\; \operatorname{span}\{\begin{pmatrix} -1 \\ 1 \end{pmatrix}\}$$

Thus $\beta = \{\begin{pmatrix} -2 \\ 1 \end{pmatrix}\}$ is a basis for E_{λ_1} and $\beta_2 = \{\begin{pmatrix} -1 \\ 1 \end{pmatrix}\}$ is a basis for E_{λ_2}. By Theorem 4.24, $\beta = \beta_1 \cup \beta_2 = \{\begin{pmatrix} -2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\}$ is a basis for $V = \mathbb{R}^2$ consisting of eigenvectors of A. Thus $[L_A]_\beta$ is a diagonal matrix. In particular, if we let $Q = \begin{pmatrix} -2 & -1 \\ 1 & 1 \end{pmatrix}$, then $Q^{-1}AQ = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, which shows that A is similar to a diagonal matrix, hence diagonalizable.

Above, we used the following consequence of our change-of-basis material:

Remark 4.28. Let $A \in M_{n\times n}(F)$, and suppose $\gamma = \{v_1, \ldots, v_n\}$ is an ordered basis for F^n. Then $[L_A]_\gamma = Q^{-1}AQ$, where $Q = (v_1\ v_2\ \cdots\ v_n)$. This is because for Q defined this way, we have $Q = [I_V]^\beta_\gamma$, where β is the standard ordered basis of F^n.

5. Inner product spaces

We will now look at a new topic, *inner product spaces*. These are an abstract version of $\mathbb{R}^n$ equipped with the dot product from multivariable calculus. Inner products give rise to a notion of *norm* (i.e., length). Consequently, we need to work over a field of scalars where the scalars themselves have a certain "magnitude":

Convention 5.1. *In this section, $F = \mathbb{R}$ or $F = \mathbb{C}$. We are no longer assuming that our vector spaces are always finite-dimensional.*

Recall that the field of complex numbers $\mathbb{C}$ comes equipped with the operation of *complex conjugation*: given $a + bi \in \mathbb{C}$, where $a, b \in \mathbb{R}$, we define $\overline{a + bi} := a - bi$ to be the complex conjugate of $a + bi$. Here are some basic arithmetic properties of complex conjugation:

- for $\alpha \in \mathbb{C}$, we have $\overline{\alpha} = \alpha$ iff $\alpha \in \mathbb{R}$
- for $\alpha, \beta \in \mathbb{C}$, we have $\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}$ and $\overline{\alpha\beta} = \overline{\alpha}\,\overline{\beta}$ and $\overline{(\alpha/\beta)} = \overline{\alpha}/\overline{\beta}$ if $\beta \neq 0$
- for $\alpha \in \mathbb{C}$, we have $\overline{\overline{\alpha}} = \alpha$ (conjugating twice does nothing)
- for $\alpha = a + bi \in \mathbb{C}$ with $a, b \in \mathbb{R}$, we define the **real part** of α as $\operatorname{Re}\alpha := a$ and the **imaginary part** of α as $\operatorname{Im}\alpha := b$. We can calculate these parts using complex conjugates:

$$\operatorname{Re}\alpha \;=\; (\alpha + \overline{\alpha})/2 \quad \text{and} \quad \operatorname{Im}\alpha \;=\; (\alpha - \overline{\alpha})/2i$$

We also have the **absolute value** (or **modulus** or **norm**) of a complex number: for $\alpha = a + bi \in \mathbb{C}$ with $a, b \in \mathbb{R}$, the absolute value is $|\alpha| := \sqrt{a^2 + b^2} \in \mathbb{R}$. Here are the basic properties:

- for $\alpha, \beta \in \mathbb{C}$, $|\alpha\beta| = |\alpha| \cdot |\beta|$ and $|\alpha/\beta| = |\alpha|/|\beta|$ for $\beta \neq 0$
- for $\alpha, \beta \in \mathbb{C}$, $|\alpha + \beta| \leq |\alpha| + |\beta|$ (Triangle inequality)
- for $\alpha \in \mathbb{C}$, $\alpha\overline{\alpha} = |\alpha|^2 \in \mathbb{R}$ (this is the key reason we care about complex conjugates in this section).

5.1. Inner products and norms.

We start by defining what an inner product is:

Definition 5.2. Let V be a vector space over F. An **inner product** on V is a function

$$\langle \cdot, \cdot \rangle : V \times V \to F$$

which assigns to each pair $(x, y) \in V \times V$ a scalar $\langle x, y \rangle \in F$ such that for all $x, y, z \in V$ and $c \in F$:

(1) $\langle cx + y, z \rangle = c\langle x, z \rangle + \langle y, z \rangle$ (linear in the first variable)
(2) $\overline{\langle x, y \rangle} = \langle y, x \rangle$ (conjugate symmetry)
(3) if $x \neq 0$, then $\langle x, x \rangle > 0$ (positivity)

Remark 5.3. Suppose $\langle \cdot, \cdot \rangle$ is an inner product on a vector space V over F.

(1) If $F = \mathbb{R}$, then conjugate symmetry is just $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in V$, since $\overline{c} = c$ for all $c \in \mathbb{R}$.
(2) As usual for linearity, we actually have for all $a_1, \ldots, a_n \in F$ and $x_1, \ldots, x_n, y \in V$ that

$$\left\langle \sum_{i=1}^{n} a_i x_i, y \right\rangle \;=\; \sum_{i=1}^{n} a_i \langle x_i, y \rangle.$$

Example 5.4. We define the **standard inner product** on F^n as follows: for $x = (a_1, \ldots, a_n), y = (b_1, \ldots, b_n) \in F^n$, we define:

$$\langle x, y \rangle \;:=\; \sum_{i=1}^{n} a_i \overline{b_i}.$$

It is straightforward to verify the condition (1), (2), and (3) for this inner product. For example, if $z = (c_1, \ldots, c_n)$ and $c \in F$, then

$$\langle cx + z, y\rangle \;=\; \sum_{i=1}^{n}(ca_i + c_i)\overline{b_i} \;=\; c\sum_{i=1}^{n} a_i\overline{b_i} + \sum_{i=1}^{n} c_i\overline{b_i} \;=\; c\langle x, y\rangle + \langle z, y\rangle.$$

Note: when $F = \mathbb{R}$, then the complex conjugations play no role and $\langle x, y\rangle$ is just the usual dot product from 33A.

The next example shows that if we have an inner product, we can define many other inner products:

Example 5.5. If $\langle\cdot,\cdot\rangle$ is an inner product on V, then given $r > 0$ from $\mathbb{R}$, we can define another inner product $\langle\cdot,\cdot\rangle'$ on V by defining $\langle x, y\rangle' := r\langle x, y\rangle$ for all $x, y \in V$. (If $r \leq 0$, this would not give an inner product).

The next example shows up all the time in math, although it looks nothing like the usual dot product:

Example 5.6. Let $V = C([1, 0])$, the $\mathbb{R}$-vector space of all continuous functions $f : [0, 1] \to \mathbb{R}$. For $f, g \in V$ we define:

$$\langle f, g\rangle \;:=\; \int_0^1 f(t)g(t)dt.$$

This defines an inner product on $C([0, 1])$. For example, given $f, g, h \in V$ and $c \in \mathbb{R}$, we have

$$\langle cf + g, h\rangle \;=\; \int_0^1 (cf(t) + g(t))h(t)dt \;=\; c\int_0^1 f(t)h(t)dt + \int_0^1 g(t)h(t)dt \;=\; c\langle f, h\rangle + \langle g, h\rangle.$$

Conjugate symmetry is clear. For positivity, if $f \neq 0 \in C([0, 1])$, then $(f(t))^2 > 0$ for some $t \in [0, 1]$, so $\int_0^1 (f(t))^2 dt > 0$ since f is continuous (you will prove this type of thing in math131a).

Example 5.7. Let $A \in M_{m\times n}(F)$. We define the **conjugate transpose** of A as the $n \times m$ matrix A^* such that $(A^*)_{ij} := \overline{A_{ji}}$. Note that when $F = \mathbb{R}$, $A^* = A^t$. For example:

$$A \;=\; \begin{pmatrix} i & 1+2i \\ 2 & 3+4i \end{pmatrix} \quad \text{and} \quad A^* \;=\; \begin{pmatrix} -i & 2 \\ 1-2i & 3-4i \end{pmatrix}.$$

Now consider $V = M_{n\times n}(F)$, and define $\langle A, B\rangle := \mathrm{tr}(B^*A)$ for $A, B \in V$. This defines an inner product on V, the so-called **Frobenius inner product**. We'll check the positivity condition (the rest are obvious): Note that

$$\langle A, A\rangle \;=\; \mathrm{tr}(A^*A) \;=\sum_{i=1}^{n}(A^*A)_{ii} \;=\; \sum_{i=1}^{n}\sum_{k=1}^{n}(A^*)_{ik}A_{ki} \;=\; \sum_{i=1}^{n}\sum_{k=1}^{n}\overline{A_{ki}}A_{ki} \;=\; \sum_{i=1}^{n}\sum_{k=1}^{n}|A_{ki}|^2,$$

which is the sum of squares of the magnitudes of each entry, so if $A \neq 0$, then $\langle A, A\rangle > 0$.

Vector spaces themselves do not come with an inner product operation. If we choose to equip a vector space with a particular inner product (which may or may not be possible), then the vector space upgrades itself to a so-called *inner product space*:

Definition 5.8. A vector space V over F equipped with an inner product $\langle\cdot,\cdot\rangle$ is called an **inner product space**. If $F = \mathbb{R}$, V is called a **real inner product space**, and if $F = \mathbb{C}$, then V is called a **complex inner product space**.

Note that if V is an inner product space with inner product $\langle\cdot,\cdot\rangle$ and $W \subseteq V$ is a subspace of V, then we may naturally also consider W as an inner product space using the restriction of $\langle\cdot,\cdot\rangle$ to W.

Here are some basic properties of inner product spaces:

Lemma 5.9. *Let V be an inner product space, let $x, y, z \in V$ and $c \in F$. Then*

(1) $\langle x, cy + z\rangle = \overline{c}\langle x, y\rangle + \langle x, z\rangle$ ("conjugate linear" in second variable)
(2) $\langle x, 0\rangle = \langle 0, x\rangle = 0$, and if $\langle x, x\rangle = 0$, then $x = 0$ (properties of zero)
(3) If $\langle x, z\rangle = \langle y, z\rangle$ for all $z \in V$, then $x = y$ (fancy way of showing two vectors are equal)

Proof. (1) Note that

$$\langle x, cy + z\rangle \;=\; \overline{\langle cy + z, x\rangle} \;= \overline{c\langle y, x\rangle + \langle z, x\rangle} \;=\; \overline{c}\overline{\langle y, x\rangle} + \overline{\langle z, x\rangle} \;=\; \overline{c}\langle x, y\rangle + \langle x, z\rangle.$$

(2) Note that $\langle x, 0\rangle = \langle x, 0x\rangle = \overline{0}\langle x, x\rangle = 0\langle x, x\rangle = 0$. Similarly $\langle 0, x\rangle = 0$. Also, if $x \neq 0$, then $\langle x, x\rangle > 0$.

(3) Suppose $\langle x, z\rangle = \langle y, z\rangle$ for all $z \in V$. We want to show that $x = y$. It is sufficient to show that $x - y = 0$. Note that

$$\langle x - y, x - y\rangle \;= \langle x, x - y\rangle - \langle y, x - y\rangle \;=\; 0$$

since $\langle x, x - y\rangle = \langle y, x - y\rangle$ by assumption (using $z := x - y$). Thus $x - y = 0$ by (2) above. □

Recall that the usual dot product in $\mathbb{R}^3$ gave us a round-about way of defining the length of a vector: the length of $x = (a, b, c) \in \mathbb{R}^3$ is the distance between (a, b, c) and $(0, 0, 0)$ which is $\sqrt{a^2 + b^2 + c^2}$. In other words, the length of x is $\sqrt{x \cdot x}$. We use this idea as the definition of the norm (i.e., length) of a vector in an abstract inner product space:

Definition 5.10. Suppose V is an inner product space. Then for every $x \in V$, we define the **norm** or **length** of x as $\|x\| := \sqrt{\langle x, x\rangle}$.

Example 5.11. Let $V = F^n$, equipped with the standard inner product. Given $x = (a_1, \ldots, a_n) \in V$, then

$$\|x\| \;=\; \|(a_1, \ldots, a_n)\| \;=\; \left(\sum_{i=1}^{n} |a_i|^2\right)^{1/2}$$

is the usual Euclidean definition of length.

Many familiar properties of length hold in this generality:

Lemma 5.12. *Suppose V is an inner product space over F. Then for all $x, y \in V$ and $c \in F$ we have*

(1) $\|cx\| = |c| \cdot \|x\|$.
(2) $\|x\| \geq 0$, and $\|x\| = 0$ iff $x = 0$,
(3) $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$ (Cauchy-Schwarz Inequality)
(4) $\|x + y\| \leq \|x\| + \|y\|$ (Triangle Inequality)

Proof. (1) and (2) are routine.

(3) If $y = 0$, then $\langle x, y\rangle = 0$ and $\|x\| \cdot \|y\| = 0$, so the result holds. Now assume $y \neq 0$, so $\langle y, y\rangle > 0$. We will employ a sneaky trick. Let $c \in F$ be arbitrary. Note that

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

The above inequality holds for any $c \in F$. Thus it holds for $c = \langle x, y\rangle / \langle y, y\rangle$. Plugging in this c in gives us:

$$\begin{aligned} 0 &\leq \langle x, x\rangle - \frac{\overline{\langle x, y\rangle}}{\langle y, y\rangle}\langle x, y\rangle - \frac{\langle x, y\rangle}{\langle y, y\rangle}\langle y, x\rangle + \frac{\langle x, y\rangle}{\langle y, y\rangle}\frac{\overline{\langle x, y\rangle}}{\langle y, y\rangle}\langle y, y\rangle \\ &= \langle x, x\rangle - \frac{|\langle x, y\rangle|^2}{\langle y, y\rangle} - \frac{|\langle x, y\rangle|^2}{\langle y, y\rangle} + \frac{|\langle x, y\rangle|^2}{\langle y, y\rangle} \quad \text{using } \alpha\overline{\alpha} = |\alpha|^2 \text{ for any } \alpha \in F \\ &= \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y\|^2}. \end{aligned}$$

Rearranging this inequality yields

$$|\langle x, y\rangle|^2 \leq \|x\|^2\|y\|^2$$

and taking square roots gives the desired inequality.

(4) We will prove instead that $\|x + y\|^2 \leq (\|x\| + \|y\|)^2$. Note that

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y\rangle \\ &= \langle x, x\rangle + \langle y, x\rangle + \langle x, y\rangle + \langle y, y\rangle \\ &= \|x\|^2 + 2\,\mathrm{Re}(\langle x, y\rangle) + \|y\|^2 \quad \text{using } \alpha + \overline{\alpha} = 2\,\mathrm{Re}(\alpha) \text{ for } \alpha \in F \\ &\leq \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 \quad \text{by Cauchy-Schwarz} \\ &= (\|x\| + \|y\|)^2. \end{aligned}$$

□

5.2. **Orthogonality.** You may recall in earlier courses in $\mathbb{R}^2$ and $\mathbb{R}^3$, the following formula for the dot product:

$$\vec{x} \cdot \vec{y} = \|\vec{x}\| \cdot \|\vec{y}\| \cos\theta$$

where $\theta \in [0, \pi]$ is the angle between $\vec{x}$ and $\vec{y}$. The most important case of this is when $\theta = \pi/2$ (the angle is 90°, a right-angle), i.e., when $\vec{x}$ and $\vec{y}$ are *orthogonal* or *perpendicular*. This happens precisely when the dot product $\vec{x} \cdot \vec{y} = 0$ equals zero. We will generalize this to arbitrary inner product spaces now.

Definition 5.13. Let V be an inner product space.

(1) Vectors $x, y \in V$ are **orthogonal** (think "perpendicular") if $\langle x, y\rangle 0$.
(2) A subset $S \subseteq V$ is **orthogonal** if for every $x, y \in V$, if $x \neq y$, then $\langle x, y\rangle$.
(3) A vector $x \in V$ is a **unit vector** if $\|x\| = 1$.
(4) A subset $S \subseteq V$ is **orthonormal** if S is orthogonal and consists entirely of unit vectors.

Remark 5.14. In general, it is favorable to work with unit vectors. We can often replace vectors with unit vectors. This process is called *normalization*. More specifically:

(1) A set of vectors $S = \{v_1, v_2, \ldots\}$ is orthonormal iff

$$\begin{cases} \langle v_i, v_j\rangle = 0 & \text{for all } i \neq j \\ \langle v_i, v_i\rangle = 1 & \text{for all } i. \end{cases}$$

(2) If $S = \{v_1, v_2, \ldots\}$ is orthogonal, and $a_1, a_2, a_3, \ldots \in F$ are non-zero scalars, then the set $\{a_1v_1, a_2v_2, \ldots\}$ is also orthogonal.
(3) If $x \in V$ is such that $x \neq 0$, then $y := \left(\frac{1}{\|x\|}\right) x$ is a unit vector. We say that y is obtained from x by **normalizing**.

(4) By (2) and (3), given a set of nonzero orthogonal vectors, we can obtain an orthonormal set by normalizing every vector in it.

Example 5.15. In $\mathbb{R}^3$, $\{(1,1,0),(1,-1,1),(-1,1,2)\}$ is an orthogonal set of nonzero vectors, but it is not orthonormal. By normalizing each of the vectors, we obtain an orthonormal set:

$$\Big\{\frac{1}{\sqrt{2}}(1,1,0),\frac{1}{\sqrt{3}}(1,-1,1),\frac{1}{\sqrt{6}}(-1,1,2)\Big\}.$$

5.3. **Orthonormal bases and Gram-Schmidt orthogonalization.**

Definition 5.16. Let V be an inner product space. A subset S of V is an **orthonormal basis** for V is S is an ordered basis for V and S is orthonormal.

The same way a bases are the building blocks of a vector space, orthonormal bases are the building blocks of inner product spaces.

Example 5.17. The standard ordered basis for F^n is an orthonormal basis for the inner product space F^n (equipped with the standard inner product).

Example 5.18. The set

$$\left\{\left(\frac{1}{\sqrt{5}},\frac{2}{\sqrt{5}}\right),\left(\frac{2}{\sqrt{5}},\frac{-1}{\sqrt{5}}\right)\right\}$$

is an orthonormal basis for $\mathbb{R}^2$.

The following illustrates the utility of orthonormal/orthogonal bases/sets; it makes finding the coefficients in linear combinations very easy.

Proposition 5.19. *Let V be an inner product space and $S=\{v_1,\dots,v_k\}$ be an orthogonal subset of V consisting of distinct nonzero vectors. If $y\in\operatorname{span}(S)$, then*

$$y \;=\; \sum_{i=1}^{k}\frac{\langle y,v_i\rangle}{\|v_i\|^2}v_i.$$

*In addition, if S is **orthonormal**, then*

$$y \;=\; \sum_{i=1}^{k}\langle y,v_i\rangle v_i.$$

Proof. Note first that since $\{v_1,\dots,v_k\}$ is orthogonal, if $i\neq j$, then $\langle v_i,v_j\rangle=0$ (we'll use this below). Write $y=\sum_{i=1}^k a_iv_i$, where $a_1,\dots,a_k\in F$. Then for $j=1,\dots,k$ we can "apply $\langle\cdot,v_j\rangle$" to "$y=\sum_{i=1}^k a_iv_i$" to get

$$\begin{aligned}
\langle y,v_j\rangle &= \left\langle\sum_{i=1}^{k}a_iv_i,v_j\right\rangle \quad \text{because } y=\textstyle\sum_{i=1}^k a_iv_i\\
&= \sum_{i=1}^{k}a_i\langle v_i,v_j\rangle \quad \text{because } \langle\cdot,v_j\rangle \text{ is linear in first variable}\\
&= a_j\langle v_j,v_j\rangle \quad \text{because } \langle v_i,v_j\rangle=0 \text{ if } i\neq j\\
&= a_j\|v_j\|^2
\end{aligned}$$

Since $v_j\neq 0$ by assumption, we can solve for a_j to get $a_j=\langle v,v_j\rangle/\|v_j\|^2$.

We are now done. However, to belabor the point, we have shown for each $j=1,\dots,n$, that $a_j=\langle v,v_j\rangle/\|v_j\|^2$. Since j is just a dummy index, this means that for each $i=1,\dots,n$, we

have $a_i = \langle v, v_i\rangle / \|v_i\|^2$. Plugging this expression for a_i back into our original linear combination $y = \sum_{i=1}^k a_i v_i$ yields the desired formula

$$y \;=\; \sum_{i=1}^{k} \frac{\langle y, v_i\rangle}{\|v_i\|^2} v_i.$$

In case S is orthonormal, then each v_i is unit length, i.e., $\|v_i\| = 1$ for each i. Thus the above expression simplifies to

$$y \;=\; \sum_{i=1}^{k} \langle y, v_i\rangle v_i$$

in this case. □

The proof of Proposition 5.19 also implies that orthogonal sets of nonzero vectors are linearly independent:

Corollary 5.20. *Let V be an inner product space and let S be an orthogonal subset of S consisting of nonzero vectors. Then S is linearly independent.*

Proof. Suppose $v_1, \ldots, v_k \in S$ and $\sum_{i=1}^k a_i v_i = 0$. As in the proof of Proposition 5.19 with $y = 0$, if we apply $\langle \cdot, v_j\rangle$ to this linear combination, we deduce that $a_j = \langle 0, v_j\rangle / \|v_j\|^2 = 0$ for all j. Thus S is linearly independent. □

Example 5.21. By the previous corollary, the orthonormal set

$$\beta \;=\; \left\{ \frac{1}{\sqrt{2}}(1,1,0), \frac{1}{\sqrt{3}}(1,-1,1), \frac{1}{\sqrt{6}}(-1,1,2) \right\}.$$

from a previous example is an orthonormal basis for $\mathbb{R}^3$ (since it is in fact linearly independent). Let $x = (2,1,3)$. Using Proposition 5.19 we can compute:

$$a_1 \;=\; \langle x, v_1\rangle \;=\; 2 \cdot \frac{1}{\sqrt{2}} + 1 \cdot \frac{1}{\sqrt{2}} + 3 \cdot 0 \;=\; \frac{3}{\sqrt{2}}$$

$$a_2 \;=\; \langle x, v_2\rangle \;=\; 2 \cdot \frac{1}{\sqrt{3}} - 1 \cdot \frac{1}{\sqrt{3}} + 3 \cdot \frac{1}{\sqrt{3}} \;=\; \frac{4}{\sqrt{3}}$$

$$a_3 \;=\; \langle x, v_3\rangle \;=\; 2 \cdot \left(-\frac{1}{\sqrt{6}}\right) + 1 \cdot \frac{1}{\sqrt{6}} + 3 \cdot \frac{2}{\sqrt{6}} \;=\; \frac{5}{\sqrt{6}}$$

and thus

$$x \;=\; \frac{3}{\sqrt{2}} v_1 + \frac{4}{\sqrt{3}} v_2 + \frac{5}{\sqrt{6}} v_3.$$

This example shows that expressing an arbitrary vector in terms of an orthonormal basis is as easy as taking inner products. Now the questions is, how do we obtain an orthonormal basis?

We will look at a special case first (see Figure 6.1 on Page 344 of Friedberg). Suppose $\{w_1, w_2\}$ is a linearly independent set of vectors in an inner product space V. Let $W := \text{Span}(w_1, w_2)$

- We want to replace $\{w_1, w_2\}$ with an orthogonal set $\{v_1, v_2\}$ which spans the same subspace.
- We keep w_1 as is, i.e., set $v_1 := w_1$.
- We want to subtract some multiple of w_1 from w_2 to create "a version" of w_2 which is orthogonal to w_1 (we'll call it v_2).
- I.e., we want to find $c \in F$ such that w_1 and $v_2 := w_2 - cw_1$ are orthogonal.

- Let's solve for $c \in F$:

$$0 \;=\; \langle v_2, w_1\rangle \;=\; \langle w_2 - cw_1, w_1\rangle \;=\; \langle w_2, w_1\rangle - c\langle w_1, w_1\rangle$$

and thus

$$c \;=\; \frac{\langle w_2, w_1\rangle}{\|w_1\|^2}, \quad \text{and} \quad v_2 \;=\; w_2 - \frac{\langle w_2, w_1\rangle}{\|w_1\|^2} w_1.$$

The following gives the general case:

Gram-Schmidt Process 5.22. *Let V be an inner product space, and $S = \{w_1, \ldots, w_n\}$ a linearly independent set of vectors. Define $S' = \{v_1, \ldots, v_n\}$, where $v_1 := w_1$, and*

$$v_k \;:=\; w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j\rangle}{\|v_j\|^2} v_j \quad \textit{for } 2 \le k \le n$$

Then S' is an orthogonal set of non-zero vectors such that $\mathrm{Span}(S') = \mathrm{Span}(S)$.

Proof. We will proceed by induction on n, the number of vectors in S.

Base Case: If $n = 1$, then we are done with $S'_1 = S_1$ and $v_1 = w_1 \neq 0$.

Induction step: Suppose $n > 1$. Assume that we know the statement of "Gram-Schmidt Process" is true for any linearly independent set of $n-1$ vectors. Then we can apply the statement to the set $S_{n-1} := \{v_1, \ldots, v_{n-1}\}$ to obtain a set $S'_{n-1} = \{v_1, \ldots, v_{n-1}\}$ with the desired properties, i.e., that

$$v_k \;:=\; w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j\rangle}{\|v_j\|^2} v_j \quad \text{for } 2 \le k \le n-1,$$

and S'_{n-1} is an orthogonal set of nonzero vectors such that $\mathrm{Span}(S'_{n-1}) = \mathrm{Span}(S_{n-1})$.

We will show that $S' := S'_{n-1} \cup \{v_n\} = \{v_1, \ldots, v_{n-1}, v_n\}$ also has the desired properties, where

$$(*) \qquad v_n \;:=\; w_n - \sum_{j=1}^{n-1} \frac{\langle w_n, v_j\rangle}{\|v_j\|^2} v_j$$

If $v_n = 0$, then $(*)$ implies that $w_n \in \mathrm{Span}(S'_{n-1}) = \mathrm{Span}(S_{n-1})$, contradicting the assumption that S is linearly independent, thus $v_n \neq 0$. Now note that for $i = 1, \ldots, n-1$, we have

$$\langle v_n, v_i\rangle \;=\; \langle w_n, v_i\rangle - \sum_{j=1}^{n-1} \frac{\langle w_n, v_j\rangle}{\|v_j\|^2} \underbrace{\langle v_j, v_i\rangle}_{=0 \text{ if } i \neq j} \;=\; \langle w_n, v_i\rangle - \frac{\langle w_n, v_i\rangle}{\|v_i\|^2}\|v_i\|^2 \;=\; 0$$

Thus S' is an orthogonal set of nonzero vectors (we already know $\langle v_i, v_j\rangle$ behaves as aspected for $i, j \in \{1, \ldots, n-1\}$ since S'_{n-1} is an orthogonal set of nonzero vectors by assumption). It follows that $\mathrm{Span}(S') \subseteq \mathrm{Span}(S)$. However, S' is linearly independent by Corollary 5.20 and has n vectors, thus $\mathrm{Span}(S') = \mathrm{Span}(S)$. □

Example 5.23. Use Gram-Schmidt to orthogonalize the following subset of $V = \mathbb{R}^4$ (with the standard inner product): $\{w_1, w_2, w_3\}$ where $w_1 = (1, 0, 1, 0)$, $w_2 = (1, 1, 1, 1)$, and $w_3 = (0, 1, 2, 1)$. First, we define $v_1 := w_1 = (1, 0, 1, 0)$. Next we have

$$v_2 \;:=\; w_2 - \frac{\langle w_2, v_1\rangle}{\|v_1\|^2} \;=\; (1, 1, 1, 1) - \frac{2}{2}(1, 0, 1, 0) \;=\; (0, 1, 0, 1).$$

Furthermore,

$$v_3 \;:=\; w_3 - \frac{\langle w_3, v_1\rangle}{\|v_1\|^2} - \frac{\langle w_3, v_2\rangle}{\|v_2\|^2} \;=\; (0, 1, 2, 1) - \frac{2}{2}(1, 0, 1, 0) - \frac{2}{2}(0, 1, 0, 1) \;=\; (-1, 0, 1, 0).$$

Now we have an orthogonal set of vectors $\{v_1, v_2, v_3\}$ which have the same span as $\{w_1, w_2, w_3\}$. Suppose we want to go one step further and get an orthonormal basis for this subspace. Then we just normalize our orthogonal set:

$$\begin{aligned} u_1 &:= \frac{1}{\|v_1\|}v_1 = \frac{1}{\sqrt{2}}(1,0,1,0) \\ u_2 &:= \frac{1}{\|v_2\|}v_2 = \frac{1}{\sqrt{2}}(0,1,0,1) \\ u_3 &:= \frac{1}{\|v_3\|}v_3 = \frac{1}{\sqrt{2}}(-1,0,1,0) \end{aligned}$$

Thus $\{u_1, u_2, u_3\}$ is an orthonormal basis for the subspace $\mathrm{Span}(w_1, w_2, w_3)$.

Corollary 5.24. *Let V be a finite-dimensional inner product space. Then V has an orthonormal basis β. Furthermore, if $\beta = \{v_1, \ldots, v_n\}$ and $x \in V$, then $x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i$.*

Proof. Let β_0 be an ordered basis for V (not necessarily orthogonal or orthonormal). Applying Gram-Schmidt 5.22, we obtain an orthogonal set β' of nonzero vectors with $\mathrm{Span}(\beta_0) = \mathrm{Span}(\beta') = V$. Normalizing each vector in β', we obtain an *orthonormal* set β such that $\mathrm{Span}(\beta) = \mathrm{Span}(\beta') = V$. β is linearly independent by Corollary 5.20, hence is an orthonormal basis. The rest follows from Proposition 5.19. □

Corollary 5.25. *Let V be a finite-dimensional inner product space with an orthonormal basis $\beta = \{v_1, \ldots, v_n\}$. Let T be a linear operator on V, and let $A = [T]_\beta$. Then for any i, j, $A_{ij} = \langle T(v_j), v_i \rangle$.*

5.4. **Orthogonal complement.** In this subsection, we generalize the relationship between a plane in $\mathbb{R}^3$ which passes through the origin, and a line through the origin which is normal to that plane.

Definition 5.26. Let V be an inner product and let $S \subseteq V$ be a non-empty set of vectors. We define the **orthogonal complement of** S to be the set $S^\perp$ (pronounced "S perp") to be the set of all vectors in V that are orthgonal to *all* vectors in S, i.e.,

$$S^\perp := \big\{x \in V : \langle x, y \rangle = 0 \text{ for all } y \in S\big\}.$$

Note that $S^\perp \subseteq V$ is a subspace of V, even if S was not a subspace.

Example 5.27.
(1) For V any inner product space, we always have $\{0\}^\perp = V$ and $V^\perp = \{0\}$.
(2) For $V = \mathbb{R}^3$ (with standard inner product), and $S = \{e_3\}$, then $S^\perp$ is the xy-plane, i.e., $S^\perp = \mathrm{Span}(e_1, e_2)$.

Proposition 5.28. *Suppose V is an inner product space, $W \subseteq V$ is a subspace such that $\dim(W) < \infty$. Let $y \in V$. Then there are unique vectors $u \in W$ and $z \in W^\perp$ such that $y = u + z$. Furthermore, if $\{v_1, \ldots, v_k\}$ is an orthonormal basis for W, then $u = \sum_{i=1}^{k} \langle y, v_i \rangle v_i$.*

Proof. Let $\{v_1, \ldots, v_k\}$ be an orthonormal basis for W and let $u := \sum_{i=1}^{k} \langle y, v_i \rangle v_i$. Define $z := y - u$. Then $u \in W$ and $y = u + z$. To show that $z \in W^\perp$, it suffices to show that z is orthogonal to each v_j (by conjugate linearity, it follows that z will be orthogonal to $\mathrm{Span}(v_1, \ldots, v_k) = W$). Then for

$j = 1, \ldots, k$ we have

$$\begin{aligned}\langle z, v_j\rangle &= \left\langle y - \sum_{i=1}^{k}\langle y, v_i\rangle v_i, v_j \right\rangle \\ &= \langle y, v_j\rangle - \sum_{i=1}^{k}\langle y, v_i\rangle\langle v_i, v_j\rangle \\ &= \langle y, v_j\rangle - \langle y, v_j\rangle \\ &= 0.\end{aligned}$$

For uniqueness of u and z, suppose that $y = u + z = u' + z'$ where $u' \in W$ and $z' \in W^\perp$. Then $u - u' = z' - z \in W \cap W^\perp = \{0\}$. Thus $u = u'$ and $z = z'$. □

Corollary 5.29. *Let V, W, $y = u + z$ be as in Proposition 5.28. The vector u is the unique vector in W that is "closest" to y in the following sense: given any $x \in W$, $\|y - x\| \geq \|y - u\|$, and this inequality is an equality if and only if $x = u$.*

Proof. Suppose $x \in W$. Then $u - x \in W$ is orthogonal to to $z \in W^\perp$, so we have

$$\begin{aligned}\|y - x\|^2 &= \|u + z - x\|^2 \\ &= \|(u - x) + z\|^2 \\ &= \|u - x\|^2 + \|z\|^2 \quad \text{because } u - x \perp z, \text{ see HW problem} \\ &\geq \|z\|^2 \\ &= \|y - u\|^2.\end{aligned}$$

Next, suppose $\|y-x\| = \|y-u\|$. Then the inequality above is equality, so we have $\|u-x\|^2+\|z\|^2 = \|z\|^2$. Thus $\|u - x\| = 0$, so $x = u$. □

The vector u in the above Corollary 5.29 and Proposition 5.28 is called the **orthogonal projection of y on W**.

Proposition 5.30. *Suppose V is an inner product space with $\dim(V) = n$. Let $S = \{v_1, \ldots, v_k\} \subseteq V$ be orthonormal. Then*

(1) S can be extended to an orthonormal basis $\{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$ for V.
(2) If $W = \mathrm{Span}(S)$, then $S_1 := \{v_{k+1}, \ldots, v_n\}$ is an orthonormal basis for $W^\perp$.
(3) If W is any subspace of V, then $\dim(V) = \dim(W) + \dim(W^\perp)$.

Proof. (1) By Corollary 1.60, S can be extended to an ordered basis $S' = \{v_1, \ldots, v_k, w_{k+1}, \ldots, w_n\}$ for V. Now we apply the Gram-Schmidt Process to S'. The first k vectors will stay the same since they are already orthogonal to each other. Thus we get a new orthogonal ordered basis of V, and then normalizing the last $n - k$ gives us an orthonormal basis.

(2) S_1 is linearly independent because it is contained in a basis. Since S_1 is a subset of $W^\perp$, we need to show that $\mathrm{Span}(S_1) \supseteq W^\perp$. For any $x \in V$, we have $x = \sum_{i=1}^{n}\langle x, v_i\rangle v_i$. If $x \in W^\perp$, then $\langle x, v_i\rangle = 0$ for $i = 1, \ldots, k$. Thus

$$x = \sum_{i=k+1}^{n}\langle x, v_i\rangle v_i \in \mathrm{Span}(S_1).$$

(3) Let W be a subspace of V. It is a finite dimensional inner product space since V is, so W has an orthonormal basis $\{v_1, \ldots, v_k\}$. By (1) and (2), it follows that $\dim(V) = n = k + (n - k) = \dim(W) + \dim(W^\perp)$. □

Example 5.31. Let $W = \text{Span}(\{e_1, e_2\})$ in F^3. Then $x = (a, b, c) \in W^\perp$ iff $0 = \langle x, e_1 \rangle = a$ and $0 = \langle x, e_1 \rangle = b$. Thus $x = (a, b, c) \in W^\perp$ iff $x = (0, 0, c)$ (so $a = b = 0$). so $W^\perp = \text{Span}(\{e_3\})$.

6. Appendix: non-linear algebra math

6.1. **Sets.** A **set** is a collection of mathematical objects. Mathematical objects can be almost anything: numbers, other sets, functions, vectors, etc. For instance:

$$\{2,5,7\}, \quad \{3,5,\{8,9\}\}, \quad \text{and} \quad \{1,3,5,7,\ldots\}$$

are all sets. A member of a set is called is called an **element** of the set. The membership relation is denoted with the symbol "$\in$", for instance, we write "$2 \in \{2,5,7\}$" (pronounced "2 is an element of the set $\{2,5,7\}$") to denote that the number 2 is a member of the set $\{2,5,7\}$. There are several ways to describe a set:

- by explicitly listing the elements in that set, i.e., the set $\{2,5,7\}$ is a set with three elements, the number 2, the number 5, and the number 7.
- by specifying a "membership requirement" that determines precisely which objects are in that set. For instance:

  $$\big\{n \in \mathbb{Z} : n \text{ is positive and odd}\big\}$$

  is the set of all odd positive integers. The above set is pronounced "the set of all integers n such that n is positive and odd". The colon ":" is usually pronounced "such that", and the condition to the right of the colon is the membership requirement. Defining a set in this way is sometimes referred to as using *set-builder notation* since you are describing how the set is built (in the above example, the set is built by taking all integers and keeping the ones that are positive and odd), instead of explicitly specifying which elements are in the set.

Here are some famous sets you should be familiar with:

(1) The **emptyset** is the set with no elements. It is denoted by $\emptyset$ or $\{\}$.
(2) The set of **natural numbers** is the set $\mathbb{N} = \{1,2,3,4,\ldots\}$ (in this class we do not consider 0 to be a natural number, in accordance with the textbook).
(3) The set of **integers** is the set $\mathbb{Z} = \{0,1,-1,2,-2,3,-3,\ldots\}$.
(4) The set of **rational numbers** is the set

$$\mathbb{Q} \;=\; \left\{\frac{k}{l} : k,l \in \mathbb{Z} \text{ and } l \neq 0\right\}$$

of all fractions of integers.
(5) The set of **real numbers**, which we denote by $\mathbb{R}$. This set of numbers contains all rational numbers but it also contains numbers like $\sqrt{2}$ and π. In this class we will not discuss exactly what the real numbers are (this is done in MATH131A), and we will assume familiarity with the basic properties of $\mathbb{R}$.
(6) The set of **complex numbers** are all numbers of the form

$$\{a+bi : a,b \in \mathbb{R}\}$$

where $i^2 = 1$.

Consider the following two sets:

$$\{2,5,7\} \quad \text{versus} \quad \{2,5,7,9\}$$

Notice how every element of the first set is contained in the second set. This relationship is denoted with the symbol $\subseteq$, pronounced "is a subset of". Specifically: given two sets A and B, we say that A is a **subset** of B (written: $A \subseteq B$) if for every $x \in A$, it follows that $x \in B$. Note that for every set A, it is always automatically true that $\emptyset \subseteq A$, because there are no elements $x \in \emptyset$ such that

$x \notin A$ (since there are no elements in $\emptyset$, period). We have the following subset relations among our famous sets from above:

$$\emptyset \subseteq \mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}.$$

Question 6.1. *Suppose you are in a situation where you are asked to prove $A \subseteq B$ for some sets A and B. What is the general strategy?*

Answer. This depends on what the sets A and B actually are, but in general, you take an arbitrary element $x \in A$ and by some argument, conclude that x also has to be an element of B, i.e., that $x \in B$. Here, "arbitrary" means that you are not allowed to assume anything specific about the element x *except* that it belongs to A. We give an example from linear algebra. □

Example 6.2. Prove that

$$\{(a,b,c) \in \mathbb{R}^3 : a = 0 \text{ and } b = 0\} \subseteq \{(a,b,c) \in \mathbb{R}^3 : a + b = 0\}$$

Proof. Call the first set A and the second set B. We want to prove that $A \subseteq B$. Let $(a,b,c) \in A$ be arbitrary. This means that $a = 0$ and $b = 0$. We want to show that (a,b,c) is an element of B. In order to be an element of B, it would have to be true that $a + b = 0$. However, since $a = 0$ and $b = 0$, then we have $a + b = 0 + 0 = 0$. Thus (a,b,c) satisfies the membership requirement for B so we can conclude that $(a,b,c) \in B$. Since (a,b,c) was an arbitrary element of A, we can conclude that $A \subseteq B$. □

INCORRECT proof. We want to show that $A \subseteq B$. Let (a,b,c) be an element of A, for instance, $(0,0,2)$. The vector $(0,0,2)$ is also in B since $a + b = 0 + 0 = 0$ for this vector. Thus $A \subseteq B$. [Here, the crime is that we showed that a single specific vector from A is also in B. This does not constitute a proof that *all* vectors from A are also elements of B.] □

Since sets are just collections of elements, we would like to say that

$$\{2,5,7\} \quad \text{is the same set as} \quad \{7,5,2\},$$

i.e., the only thing that determines a set completely is which elements are in it. In other words, we say that two sets A and B are **equal** (written $A = B$) if they have the same elements.

Question 6.3. *Suppose you are asked to prove that $A = B$ where A and B are sets (possibly with two different-seeming descriptions). How do you prove that $A = B$?*

Answer. This means you have to prove two separate things:

(1) prove $A \subseteq B$, and
(2) prove $B \subseteq A$.

So this breaks down to two different proofs, each one then reduces to answering Question 6.1 for those particular sets. □

6.2. **Set operations - making new sets from old.** Let A, B be sets. We define the **union** of A and B to be the set of all elements which are members of either A *or* B, written:

$$A \cup B := \{x : x \in A \text{ or } x \in B\}$$

We define the **intersection** of A and B to be the set of all elements which are members of both A *and* B, written:

$$A \cap B := \{x : x \in A \text{ and } x \in B\}$$

For example:

$$\{1,2,3\} \cup \{3,4,5\} = \{1,2,3,4,5\} \quad \text{and} \quad \{1,2,3\} \cap \{3,4,5\} = \{3\}$$

We define the **(cartesian) product** of A and B to be the set of all ordered pairs (a, b) such that $a \in A$ and $b \in B$, written:

$$A \times B \ := \ \big\{(a, b) : a \in A \text{ and } b \in B\big\}$$

For example,

$$\{1, 2\} \times \{3, 4\} \ = \ \big\{(1, 3), (1, 4), (2, 3), (2, 4)\big\}$$

We can also take the cartesian product of finitely many sets $A_1, \ldots, A_n$:

$$A_1 \times \cdots \times A_n \ := \ \big\{(a_1, a_2, \ldots, a_n) : a_i \in A_i \text{ for } i = 1, \ldots, n\big\}$$

Given a set A and $n \geq 1$, we define

$$A^n \ := \ \underbrace{A \times \cdots \times A}_{n \text{ times}}$$

For example, this is how we define the vector space F^n, where F is a field.

6.3. **Functions.** In this subsection we give a precise set-theoretic definition of a function.

Definition 6.4. A **function** is an ordered triple (f, A, B) such that

(1) A and B are sets, and
(2) $f \subseteq A \times B$ is a subset of the cartesian product $A \times B$ with the properties:
   (a) for every $a \in A$, there is $b \in B$ such that $(a, b) \in f$, and
   (b) for every $a \in A$, if $b, b' \in B$ are such that $(a, b), (a, b') \in f$, then $b = b'$.

Given a function (f, A, B), the set A is called the **domain**, and the set B is called the **codomain**. We express this fact by writing $f : A \to B$, which is read "f is a function from A to B". Sometimes we will just refer to a function f instead of (f, A, B) when the domain and codomain are clear from context (or don't matter).

In practice, we almost never think of a function as an ordered triple (f, A, B). Furthermore, instead of writing "$(a, b) \in f$", we write instead "$f(a) = b$". We think of a function as being a rule which assigns to each $a \in A$, a unique element $b \in B$.

Example 6.5. Let f be the following function from $A = \{0, 1\}$ to $B = \{2, 3, 4\}$:

$$f \ = \ \big\{(0, 3), (1, 2)\big\}.$$

Usually, we would describe f instead by specifying in this case: $f(0) = 3$ and $f(1) = 2$.

Question 6.6. *Suppose you have two functions $f : A \to B$ and $g : A \to B$ that have the same domain and codomain. How do you prove that $f = g$?*

Answer. To show that $f = g$, this means that f and g take the same values on every possible input. Thus in a proof, you let $x \in A$ be arbitrary, and then you somehow prove that $f(x) = g(x)$. □

Definition 6.7. Given functions $f : A \to B$ and $g : B \to C$, we define the **composition** of f and g as the function $g \circ f : A \to C$ defined by $(g \circ f)(a) := g(f(a))$ for every $a \in A$. In other words, the pair $(a, c) \in g \circ f$ if and only if there is $b \in B$ such that $f(a) = b$ and $g(b) = c$.

Perhaps the most important thing we can say about composition of functions at this level of generality is the following:

Proposition 6.8 (Associativity of function composition)**.** *Suppose $f : A \to B$, $g : B \to C$ and $h : C \to D$. Then we have to compositions $g \circ f : A \to C$ and $h \circ g : B \to D$, and*

$$h \circ (g \circ f) \ = \ (h \circ g) \circ f$$

as functions $A \to D$.

Proof. We want to show that two functions are equal. This means we need to show that they take the same values on all possible inputs Let $a \in A$ be arbitrary. Note that

$$\begin{aligned}\big(h \circ (g \circ f)\big)(a) &= h\big((g \circ f)(a)\big) \quad \text{(definition of } \circ) \\ &= h\big(g(f(a))\big) \quad \text{(definition of } \circ) \\ &= (h \circ g)\big(f(a)\big) \quad \text{(definition of } \circ) \\ &= \big((h \circ g) \circ f\big)(a) \quad \text{(definition of } \circ)\end{aligned}$$

Thus $h \circ (g \circ f) = (h \circ g) \circ f$ as functions. □

Definition 6.9. Let A be a set. Then we define the **identity function** $I_A : A \to A$ by $I_A(a) = a$ for every $a \in A$.

Suppose $f : A \to B$ is a function. We say a function $g : B \to A$ is an **inverse** of f if $fg = I_B$ and $gf = I_A$. If f has an inverse, then we say f is **invertible**.

Note that if f is invertible, then the inverse of f is unique and we denote it by f^{-1}. Indeed, suppose $g, g' : B \to A$ are inverses of f. Then

$$g \;=\; gI_B \;=\; g(fg') \;=\; (gf)g' \;=\; I_A g' \;=\; g'.$$

Here are some basic facts about invertibility of functions you should know:

Lemma 6.10. *Suppose $f : A \to B$ and $g : B \to C$ are functions. Then*

(1) If f and g are both invertible, then
  (a) gf is invertible, with $(gf)^{-1} = f^{-1}g^{-1}$,
  (b) f^{-1} is also invertible, with $(f^{-1})^{-1} = f$.
(2) f is invertible if and only if f is a bijection (one-to-one and onto).

Proof. (1) For (a), it is enough to show that $f^{-1}g^{-1}$ is an inverse to gf. Note that

$$(f^{-1}g^{-1})(gf) \;=\; f^{-1}(g^{-1}g)f \;=\; f^{-1}I_B f \;=\; f^{-1}f \;=\; I_A.$$

Likewise, we also have $(gf)(f^{-1}g^{-1}) = I_B$. This shows that gf is invertible and its (unique) inverse is $f^{-1}g^{-1}$.

For (b), it is enough to show that f is an inverse to f^{-1}. We know that $ff^{-1} = I_B$ and $f^{-1}f = I_A$. Thus f^{-1} is invertible and its inverse is f, i.e., $(f^{-1})^{-1} = f$.

(2) ($\Rightarrow$) For any $b \in B$, $ff^{-1}(b) = I_B(b) = b$, thus $f(f^{-1}(b)) = b$, with $f^{-1}(b) \in A$, so f is onto. Likewise, if $f(a_1) = f(a_2)$, then $f^{-1}f(a_1) = f^{-1}f(a_2)$, and so $I_A(a_1) = I_A(a_2)$, i.e., $a_1 = a_2$. Thus f is one-to-one.

We leave the ($\Leftarrow$) direction as an exercise. □

6.4. **Induction.** In this subsection we will review the proof principle of induction. Occasionally we will use induction to prove results in class. In this class[2], the **natural numbers** is the set

$$\mathbb{N} \;=\; \{1, 2, 3, 4, \ldots\}$$

of positive integers. We will not attempt to construct the natural numbers axiomatically, instead we assume that they are already given and that we are familiar with their basic properties, for instance, how the operations $+, \cdot$ and the ordering $\leq$ work with $\mathbb{N}$. Here is an important basic property about $\mathbb{N}$ which we can take for granted:

[2]In other textbooks, sometimes the natural numbers include zero, i.e., $\mathbb{N} = \{0, 1, 2, 3, 4, \ldots\}$. Also sometimes people might write $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$. We won't do this here, but it's good to know about it. The important thing is that when you are communicating about the natural numbers with someone else, you are always on the same page about whether you are including 0 or not.

Well-ordering Principle 6.11. *Suppose $S \subseteq \mathbb{N}$ is such that $S \neq \emptyset$. Then S has a least element, i.e., there is some $a \in S$ such that for all $b \in S$, $a \leq b$.*

The Well-Ordering Principle of $\mathbb{N}$ gives us the following important *proof principle* about natural numbers:

Principle of Induction 6.12. *Suppose $P(n)$ is a property that a natural number n may or may not have. Suppose that*

(1) $P(1)$ holds (this is called the "base case for the induction"), and

(2) for every $n \in \mathbb{N}$, if $P(n)$ holds, then $P(n+1)$ holds (this is called the "inductive step").

Then $P(n)$ holds for every *natural number $n \in \mathbb{N}$.*

Proof. Define the set:

$$S \;:=\; \big\{n \in \mathbb{N} : P(n) \text{ is false}\big\} \;\subseteq\; \mathbb{N}.$$

Assume towards a contradiction that $P(n)$ does not hold for every natural number $n \in \mathbb{N}$. Thus $S \neq \emptyset$. By the Well-Ordering Principle, the set S has a least element a. Since $P(1)$ holds by assumption, we know that $1 < a$ (so $a - 1 \in \mathbb{N}$). Since a is the least element of S, then the natural number $a - 1 \notin S$, so $P(a-1)$ holds. By assumption (2), this implies $P(a)$ holds, a contradiction. □

Warning 6.13. In part (2) of the Principle of Induction, it does not say you have to prove $P(n+1)$ is true. It says you have to prove that the following implication holds:

$$\big(P(n) \text{ is true}\big) \implies \big(P(n+1) \text{ is true}\big)$$

Here is the standard first example of a proof by induction:

Example 6.14. The equality

$$1 + 2 + \cdots + n \;=\; \frac{n(n+1)}{2}$$

holds for all $n \in \mathbb{N}$.

Proof. Let $P(n)$ be the assertion:

$$P(n): \quad \text{“}1 + 2 + \cdots + n = \frac{1}{2}n(n+1) \text{ is true.”}$$

We will show that $P(n)$ holds for all $n \in \mathbb{N}$ by induction on n.

First, we show that $P(1)$ holds outright. This is easy because $P(1)$ says “$1 = \frac{1}{2} \cdot 1 \cdot 2$”, which is obviously true.

Next, we will show that $P(n)$ implies $P(n+1)$. Suppose $P(n)$ holds, i.e.,

$$1 + 2 + \cdots + n \;=\; \frac{1}{2}n(n+1).$$

We must now show that $P(n+1)$ also holds. To see this, add $n+1$ to both sides of the above equality:

$$\begin{aligned} 1 + 2 + \cdots + n + (n+1) \;&=\; \frac{1}{2}n(n+1) + (n+1) \\ &=\; (n/2 + 1)(n+1) \\ &=\; \frac{1}{2}(n+2)(n+1) \\ &=\; \frac{1}{2}(n+1)\big((n+1)+1\big). \end{aligned}$$

Thus $P(n+1)$ holds as well. □

We also have the following variant of the Principle of Induction, which starts at some integer other than 1 (for instance, if we want to start at 0):

Corollary 6.15 (Principle of Induction starting at N)**.** *Let $N \in \mathbb{Z}$ and suppose $P(n)$ is a property that an $n \geq N$ may or may not have. Suppose that*

(1) $P(N)$ holds.

(2) for every $n \geq N$, if $P(n)$ holds, then $P(n+1)$ holds.

Then $P(n)$ holds for every natural number $n \geq N$.

Proof. We will prove this by reducing it to the original Induction Principle by shifting. Let $Q(n)$ be the statement:

$$Q(n): \quad \text{“}P(n+N-1) \text{ holds.”}$$

Then (1) implies that $Q(1)$ holds. Also, (2) implies that for every $n \geq 1$, $Q(n) \Rightarrow Q(n+1)$. Thus $Q(n)$ is true for all $n \geq 1$ by the Principle of Induction. In other words, $P(n)$ is true for all $n \geq N$. □

References

1. Stephen H. Friedberg, Arnold J. Insel, and Lawrence E. Spence, *Linear algebra*, custom edition for university of california, los angeles ed., Prentice Hall, Inc., Upper Saddle River, NJ, 2003.

E-mail address: allen@math.ucla.edu

Department of Mathematics, University of California, Los Angeles, Los Angeles, CA 90095